

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO



FACULTAD DE INGENIERÍA

TESIS

ANÁLISIS PREDICTIVO / DESCRIPTIVO ENFOCADO
AL PROGRAMA TUTORÍA "NUEVA ERA"

QUE PARA OBTENER EL TÍTULO DE
INGENIERO EN COMPUTACIÓN

PRESENTA:

RAFAEL GARCÍA FERNÁNDEZ

DIRECTOR DE TESIS:

ING. GABRIELA BETZABÉ LIZÁRRAGA RAMÍREZ

SINODALES:

M.I. MIGUEL EDUARDO GONZÁLEZ CÁRDENAS

ING. FRANCISCO JAVIER MONTOYA CERVANTES

ING. CRUZ SERGIO AGUILAR DÍAZ



CIUDAD UNIVERSITARIA

MÉXICO, D.F. 2011



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos

A mis padres, ya que sin su apoyo, confianza y paciencia no hubiera logrado la consecución de este trabajo. Todo lo logre ser en la vida será gracias a ustedes. En todo momento los llevo conmigo.

A la Universidad Nacional Autónoma de México por todos estos años en los que he aprendido mucho más de lo que indican los temarios, he aprendido a ser mejor alumno, mejor profesionista, mejor amigo, mejor hijo y sobre todo mejor persona. Muchas gracias.

A mi directora de tesis la Ing. Gabriela Betzabé Lizárraga Ramírez por sus sugerencias, confianza y sobre todo por su amistad.

A mi codirector de la tesis el Prof. José Enrique Larios Canale por todos sus consejos y por compartir conmigo sus conocimientos.

Al Lic. Pablo Medina Mora y al Prof. Miguel Eduardo González Cárdenas por la información que me brindaron y la orientación necesaria para realizar el presente trabajo.

A mi jefe el Ing. Sergio Noble Camargo por las facilidades laborales prestadas durante este último año.

Índice

1. El problema	5
1.1 Presentación del problema	5
1.1.1 Antecedentes de la tutoría en la Facultad de Ingeniería	7
1.1.2 Conceptos básicos y enfoque para abordar el problema	10
1.1.3 Delimitación y enunciado del problema	13
1.2 Objetivos	13
1.2.1 Generales	13
1.2.2 Específicos	14
1.3 Alcances y limitaciones	14
2. Marco teórico del problema	15
2.1 Conceptos principales	15
2.2 Definición de tareas de minería	24
2.3 Hipótesis	27
3. Fases de descubrimiento en bases de datos	28
3.1 Sistema de gestión de bases de datos	28
3.2 Integración, almacenamiento y limpieza	29
3.3 Descripción del almacén de datos	31
3.3.1 Modelo Entidad/Relación	31
3.3.2 Diccionario de datos corporativo	32
3.4 Herramientas y técnicas de análisis de datos y evaluación de las mismas	44
4. Análisis de datos y presentación de resultados	47
4.1 Métodos de minería de datos	47
4.1.1 Árboles de decisión (Clasificación y Regresión)	47
4.1.2 Clústeres (Agrupamiento)	50
4.1.3 Reglas de asociación y dependencia (Agrupamiento, Reglas de asociación)	52
4.1.4 Naive Bayes (Clasificación)	53
4.1.5 Redes Neuronales (Clasificación, Regresión, Agrupamiento)	53
4.1.6 Algoritmo de regresión lineal	56
4.2 Primer análisis: Los tutorados	57
4.2.1 Analizar el impacto de la tutoría	57
4.2.1.1 Definición del problema	57
4.2.1.2 Vista minable	58
4.2.1.3 Elección del algoritmo	59
4.2.1.4 Minería de datos	59
4.2.1.5 Patrones	62
4.2.1.6 Evaluación e interpretación del modelo	63
4.2.2 La asistencia a la tutoría	64
4.2.2.1 Definición del problema	64
4.2.2.2 Vista minable	65
4.2.2.3 Elección del algoritmo	67
4.2.2.4 Minería de datos - Clústeres	67
4.2.2.5 Patrones - Clústeres	69
4.2.2.6 Minería de datos – Árboles de decisión	72
4.2.2.7 Patrones – Árboles de decisión	73
4.2.2.8 Realizar minería de datos –Naive Bayes	74

4.2.2.9 Patrones – Naive Bayes	75
4.2.2.10 Evaluación e interpretación	76
4.2.3 El índice académico	77
4.2.3.1 Definición del problema	77
4.2.3.2 Vista minable	77
4.2.3.3 Elección del algoritmo	79
4.2.3.4 Minería de datos	80
4.2.3.5 Patrones	81
4.2.3.6 Evaluación e interpretación	82
4.2.4 Los que trabajan	83
4.2.4.1 Definición del problema – Los que trabajan	83
4.2.4.2 Vista minable	83
4.2.4.3 Elección del algoritmo	84
4.2.4.4 Minería de datos	85
4.2.4.5 Patrones	85
4.2.4.6 Evaluación e interpretación	86
4.2.4.7 Conocimiento	86
4.2.4.8 Toma de decisiones	87
4.3 Segundo análisis: Los tutores	88
4.3.1 Definición del problema	89
4.3.2 Vista minable	90
4.3.3 Elección del algoritmo	92
4.3.4 Minería de datos	92
4.3.5 Patrones	92
4.3.6 Evaluación e interpretación	93
4.3.7 Conocimiento	94
4.3.8 Decisiones	94
4.4 Tercer análisis: La herramienta	96
4.4.1 Definición del problema	96
4.4.2 Vista minable	96
4.4.3 Elección del algoritmo	100
4.4.4 Minería de datos	100
4.4.5 Patrones	103
4.4.6 Evaluación e interpretación	106
4.4.7 Conocimiento	106
4.4.8 Decisiones	107
4.5 Conclusiones Finales	108
APÉNDICE I Cuestionario de evaluación del programa Tutoría Nueva Era	114
APÉNDICE II Encuesta socioeconómica – Hábitos de estudio	117
APÉNDICE III Creando predicciones	123
Recursos bibliográficos y mesográficos	128

Capítulo 1. El problema

1.1 Presentación del problema

Entre los problemas más complejos y frecuentes que enfrentan las Instituciones de Educación Superior (IES) del país en el nivel de licenciatura, se encuentran la deserción, el rezago escolar y los bajos índices de eficiencia terminal. Tanto la deserción y el rezago son condiciones que afectan el logro de una alta eficiencia terminal en las instituciones. Lo anterior producido por múltiples factores entre los que destacan los problemas económicos y la falta de conocimientos antecedentes con que los estudiantes acceden al nivel superior, sin dejar a un lado las diversas problemáticas sociales y culturales que puedan existir en torno al estudiante. [1]

En cifras generales y como promedio nacional, de cada 100 alumnos que inician estudios de licenciatura, entre 50 y 60 concluyen las materias del plan de estudios al cumplirse cinco años de su ingreso de éstos, tan sólo 20 obtienen su título. De los que se titulan solamente un 10%, es decir 2 egresados, lo hacen a la edad considerada como deseable (24 ó 25 años); los demás entre los 27 y los 60 años. [1]

El rezago escolar a nivel licenciatura consiste en el avance irregular en la inscripción a las asignaturas subsecuentes del plan de estudios en que el alumno se encuentra inscrito. El alumno rezagado es por lo tanto aquel que se atrasa en las inscripciones que corresponden al trayecto escolar de su cohorte generacional, por tanto, en el egreso de la misma. El rezago afecta a aquellos alumnos que no acreditan las asignaturas y no pueden avanzar hasta que las aprueben. Estas, frecuentemente son asignaturas básicas, que por su grado de dificultad presentan mayores índices de reprobación y como consecuencia, de rezago. [1]

Para lograr índices de aprovechamiento y eficiencia terminal satisfactorios es imperativo para las IES incrementar la calidad del proceso formativo, aumentar el rendimiento de los estudiantes, reducir la reprobación y el abandono. Esto les permitirá cumplir con el objetivo de responder a las demandas sociales con más y mejores egresados que, al mismo tiempo, puedan lograr una incorporación exitosa al mercado de trabajo. [1]

Para alcanzar estos objetivos es indispensable consolidar una oferta educativa de calidad; en otras palabras, mejorar cualitativa y trascendentalmente el servicio que se ofrece a los estudiantes. [1]

Lo anterior refleja la necesidad de realizar estudios sobre las características y el comportamiento de los estudiantes, en relación con los factores internos y externos que influyen en su desempeño escolar, y de manera directa en sus índices de ingreso, permanencia, egreso y titulación. Un joven puede acceder a sus estudios de licenciatura con toda la disposición de ánimo para el estudio, aunado a todas las facilidades y oportunidades que le ofrecieron una niñez y adolescencia sin carencias y sin conflictos existenciales; en contraparte, puede llegar con una fuerte carga de subalimentación y subdesarrollo, aunada a inestabilidad emocional o crisis familiar. Así mismo a lo largo de los años se han diseñado estrategias con el único objetivo de ayudar a los jóvenes en su preparación tanto profesional como la del lado humano, de esta forma fue inevitable la creación de programas de tutoría que ayuden con todo lo antes mencionado, donde los programas de tutoría se entienden como aquellos que conectan a la gente que tiene habilidades específicas y conocimiento (tutores) con individuos (tutorados) quienes necesitan o desean las mismas habilidades y ventajas para superarse en su trabajo, vida cotidiana o ejecución escolar, donde se comparten sus valores y metas personales en forma de apoyo y respeto mutuo lo cual lleva a una vida más rica para ambos, ayudando con esto a romper barreras y crear oportunidades para el éxito. [1]

Sin embargo, llegados a este punto en el cual la tutoría es una realidad como estrategia para abatir los problemas de aprovechamiento en los alumnos, nos enfrentamos a un nuevo problema, que no es otro si no el de determinar el impacto que tiene la tutoría sobre dichos problemas, lo cual nos lleva a nuevos cuestionamientos:

- ¿Qué impacto tiene la tutoría en el desempeño de los alumnos?
- ¿Qué características tienen aquellos alumnos que asisten a la tutoría al menos una vez?
- ¿Cuáles son los principales hábitos de estudio de aquellos alumnos que tienen un mejor desempeño académico?
- ¿Cómo puedo lograr que más alumnos asistan a la tutoría?

A esto le podemos sumar más preguntas, que sin duda desencadenaran otras más, y para poder resolver todas estas cuestiones existen diversas formas de analizar la información que se tiene y con ello tomar decisiones para mejorar. De esta manera dichos estudios para la resolución de estas cuestiones pueden ser realizados con diversos métodos y técnicas, algunos proveerán mejores resultados que otros, permitiendo conocer más a fondo las características y comportamientos de los alumnos, así como la influencia de la tutoría en los mismos, sin embargo en la actualidad existen un conjunto de técnicas y tecnologías cuyo objetivo es el de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto, en este caso los alumnos y su desempeño, y con ello realizar la toma de decisiones, lo anterior es conocido como Minería de Datos y cuyos modelos son capaces de brindar apoyo para resolver las cuestiones antes planteadas.

Con todo lo anteriormente mencionado, podremos decir que este trabajo de investigación estará enfocado completamente en realizar minería de datos para:

- Mejores horarios
- Identificación de patrones
- Mejores formas de comunicación con el alumno

Estas son solo algunas formas en la que la minería puede ayudarnos, sin embargo cabe decir que la creación de un modelo de minería de datos es un proceso dinámico e iterativo y que una vez que se han explorado los datos, puede que se descubra que resultan insuficientes para crear los modelos de minería de datos adecuados y que, por tanto, se deben buscar más datos. De esta manera se puede generar varios modelos y descubrir que no responden al problema planteado cuando fue definido y que, por tanto, se deba volver a definir el problema. Por esto, es importante comprender que la creación de cualquier análisis de minería de datos es un proceso, y que cada paso del proceso puede repetirse tantas veces como sea necesario para crear un análisis válido. [1]



1.1.1 Antecedentes de la tutoría en la Facultad de Ingeniería

A lo largo de los años se han desarrollado diversas estrategias para atender estos problemas, aún sin haber agotado todas las posibilidades; se avanza en la formación de los profesores, en la actualización y adecuación de los planes de estudio, en el equipamiento de laboratorios, centros de cómputo, biblioteca, etcétera. Sin embargo, los asuntos referidos a las características de los estudiantes no han tenido la suficiente relevancia. En la mayoría de las instituciones se opera desde el supuesto de una cierta homogeneidad de los alumnos; por consiguiente, los programas académicos no toman en consideración la evidente heterogeneidad de sus características, tanto en lo que se refiere a sus habilidades básicas como a los conocimientos que dominan. Las capacidades y conocimientos indispensables para desarrollar una carrera universitaria se dan por hecho, a pesar de las evidencias existentes sobre la insuficiencia de la formación en los niveles elemental y medio.



Algunos de los estudiantes que ingresan a la educación superior no dominan las habilidades ni la información y conocimientos indispensables para utilizar, de la mejor manera posible, los recursos que la universidad pone a su disposición. [1]

La ANUIES¹, cuando habla del desarrollo integral de los alumnos, señala que "como componente clave para dar coherencia al conjunto, se requiere que las instituciones de educación superior pongan en marcha sistemas de tutoría, gracias a los cuales, los alumnos cuenten, a lo largo de toda su formación, con el consejo y apoyo de un profesor debidamente preparado". Y cuando se refiere a la consolidación de los cuerpos académicos señala que "en el rol que habitualmente han desempeñado los profesores de carrera, deberá contemplarse una importante transformación, de manera que ya no sean fundamentalmente conferencistas o expositores, sino además tutores; guías que conduzcan los esfuerzos individuales y grupales de auto aprendizaje por parte de los alumnos, personas que los induzcan a la investigación o a la práctica profesional; y ejemplos de compromiso con los valores académicos, humanistas y sociales que las instituciones promuevan". [1]

Uno de los objetivos principales de la Facultad de Ingeniería es el de formar el mayor número de ingenieros posibles en las distintas especialidades, con esto su labor es la de crear ingenieros capacitados para la práctica profesional, esto le obliga a concebir y desarrollar opciones de acción que conduzcan a un mejor cumplimiento de su objetivo. [2]

Programa de Tutoría para Estudiantes de Licenciatura

En el año de 1987, al analizar los bajos índices de eficiencia terminal y de titulación emitidos por la Secretaría de Servicios Escolares de la propia Facultad se obtuvieron los siguientes datos.

"La deserción acumulada en las generaciones 1980 a 1984 era del 30% en los 10 semestres de la carrera. Del primero al segundo semestre de la carrera, la deserción era del 12%. En cada semestres escolar subsecuente un promedio del 2.5% de la población de cada generación abandonaba la carrera." [2]

Así es como en 1987 surge el Primer programa de Tutoría en donde el único criterio de selección de tutor era que

¹ La Asociación Nacional de Universidades e Instituciones de Educación Superior es una Asociación no gubernamental, de carácter plural, que agremia a las principales instituciones de educación superior del país, cuyo común denominador es su voluntad para promover su mejoramiento integral en los campos de la docencia, la investigación y la extensión de la cultura y los servicios.

fueran profesores de la facultad de Ingeniería, de medio o tiempo completo, y que ocuparan cargos académico – administrativo. Además se contaría con el apoyo pedagógico del personal del Centro de Servicios Educativos de la Facultad de Ingeniería (CESEFI²).

El Programa de Tutoría para Estudiantes de Licenciatura tendría una duración de casi 4 años, finalizando en 1991. [2]

Programa Tutoría Para Todos

En el año 2000 la Coordinación de Evaluación Educativa pasa a ser COPADI³ y junto a esto surge el Programa de Tutoría para todos [3], cuyos elementos fundamentales serían:



- El aprendizaje para toda la vida, es decir, la plena disposición a prepararse de manera continua y ordenada para actualizar los conocimientos y evitar la obsolescencia. [4]
- El aprender a aprender, aprender a emprender y aprender a ser. Esto considera la sensacional posibilidad que tienen muchos seres humanos de descubrir en el estudio y su aplicación, así como en la formación académica y humana, un elemento indispensable para el mejoramiento de la calidad de sus vidas, la fórmula para realizarse plenamente y ser felices. [4]
- El reconocimiento de que el proceso educativo puede desarrollarse en diversos lugares formales e informales. Un profesional de la ingeniería debe estar siempre e invariablemente dispuesto a aprender y a enseñar, independientemente de las circunstancias que enmarquen su quehacer como investigador, docente o hacedor de ingeniería. [4]
- El diseño de nuevas modalidades educativas con el alumno y el profesor como actores centrales. Se trata pues, de que el docente y el estudiante sean actores y no simples espectadores en el proceso de enseñanza y aprendizaje, y que puedan escribir su historia como protagonistas de su presente y futuro. [4]

De igual manera los objetivos de este Programa de Tutoría para todos serían:

- Establecer y consolidar un sistema de tutoría en el que el tutor considere al estudiante en lo individual. [5]
- Lograr que la tutoría sea parte de la tarea propia de todo profesor de carrera y un derecho de todos los alumnos. [5]
- Proporcionar a los estudiantes orientaciones diversas en lo que se refiere a:
 - Orientación para reafirmar la elección de carrera.
 - Campo de trabajo de la carrera elegida.
 - Formación técnica.
 - Formación humanista.
 - Visitas a obras y empresas de ingeniería.
 - Crecimiento personal y problemática existencial.
 - Inserción en el mercado de trabajo.

² En 1972 se crea en Facultad de Ingeniería el Centro de Didáctica y Técnicas del Aprendizaje. Posteriormente en 1975 cambia de nombre a Centro de Servicios Educativos de la Facultad de Ingeniería (CESEFI)

³ En el año 2000 la CEE pasa a ser Coordinación de Programas de Atención Diferenciada para Alumnos

Sentido de pertenencia a la Facultad y a la UNAM.

- Lograr que los estudiantes sean creativos, reflexivos, polifuncionales y emprendedores, que asuman su calidad de sujeto activo, protagonista de su propio aprendizaje y gestor de su proyecto de vida, y dispuestos a estudiar toda la vida, aprender a aprender, aprender a emprender y aprender a ser. [5]

Las etapas que se consideraron para este Programa en la Facultad fueron:

En una primera etapa de este Programa se atendía a los alumnos durante su primer semestre curricular.

Y en una segunda etapa se buscaba que los alumnos, en sus semestres 9º y 10º contaran con un tutor que sea un profesional de la ingeniería, no necesariamente profesor de asignatura, a quien el estudiante pudiera visitar y convivir e interactuar con él en su lugar de trabajo. Esta modalidad resultaba para el estudiante que está en la etapa final de su licenciatura, un estímulo y la oportunidad de acercarse y conocer su futuro quehacer cotidiano.

Programa Tutoría Nueva Era

El programa Tutoría para Todos tuvo una duración poco mayor a 7 años, siendo en el año del 2008 que surge el Programa Tutoría Nueva Era, con la coordinación de la Secretaría de Apoyo a la Docencia, cuyos propósitos principales son:

- Integrar y sistematizar las tutorías. [6]
- Incrementar su cobertura. [6]
- Garantizar su efectividad. [6]

El Programa Tutoría Nueva Era considera como factor crítico para que la tutoría funcione la confianza que logre establecer el tutor con el alumno en las sesiones de tutoría iniciales, con esto se lograra asegurar las bases para establecer una relación tutorial firme durante el primer año de la carrera del alumno y aprovechar estas bases para emprender su seguimiento los siguientes años. [6]

El Programa Tutoría Nueva Era tiene como objetivo general:

- Apoyar de manera directa e individualizada el desarrollo académico de los estudiantes de la Facultad de Ingeniería. [6]

De igual forma sus objetivos particulares son:

- Brindar el apoyo requerido para asegurar la permanencia mejorando los niveles de desempeño académico y de desarrollo de habilidades académicas fortaleciendo la base de conocimientos indispensables para los estudios de la carrera. [6]
- Fortalecer y asegurar el desarrollo de habilidades y actitudes profesionales propias del perfil de egreso de la profesión. [6]

Así mismo el programa de tutoría está planeado en tres diferentes etapas: al inicio de la carrera, durante a mitad de la carrera y al finalizar la carrera, en todas ellas se cuentan con dos modalidades: la tutoría individual y la tutoría grupal. [6]

1.1.2 Conceptos básicos y enfoque para abordar el problema

En esta sección se pretende dar un significado claro y puntual de los conceptos básicos que se engloban en la tutoría, conceptos que aunque de uso relativamente común son vitales para abordar el problema de manera correcta. Lo anterior tiene el fin de ubicar en su respectivo sitio a cada uno de los factores que intervienen en la tutoría para su posterior uso.



Tutoría

Se conceptualiza como un proceso de acompañamiento al estudiante por profesionales de la ingeniería y la docencia, para orientarlo y apoyarlo en su formación integral, al inicio, durante y al final de su carrera profesional, y en el desempeño eficiente de funciones y responsabilidades de los diversos momentos y compromisos propios de su aprendizaje. [7]

Tutor

Para fines del programa "Nueva Era", un tutor es un profesor:

- Que tiene la voluntad y vocación para establecer un vínculo entre el estudiante y las diversas problemáticas académicas y existenciales que enfrenta durante su vida universitaria. [7]
- Que facilita la integración del estudiante con la Universidad. [7]
- Que demuestra al estudiante un genuino interés por su desempeño académico, así como por su crecimiento personal. [7]
- Al que le interesa que el estudiante sea creativo, innovador y con espíritu libre y crítico. [7]
- Que tiene empatía con el alumno, entendida como la capacidad de sentir y comprender las emociones del alumno como propias, mediante un proceso de plena y genuina identificación. [7]

Tutorado

Es aquel al que:

- Se apoya en el desarrollo de una metodología de estudio y de trabajo sujeta de ser utilizada a lo largo de su vida académica. [1]
- Se le ofrece apoyo y supervisión en los temas de mayor dificultad de sus asignaturas, ya sea de manera directa o consiguiéndole quien le ayude. [1]
- Se le ofrece la libertad y confianza para abrirse y así estar dispuesto a recibir ayuda. [1]
- Se le sugieren extracurriculares que favorezcan un desarrollo profesional integral. Entre éstas vale destacar aquellas que alientan la formación y realización plenas, como la lectura, la asistencia a diversas manifestaciones culturales, la recreación y el deporte, y la participación en acciones de beneficio social. [1]
- Se le proporciona información académico-administrativa. [1]

Pertinencia

Oportunidad, adecuación y conveniencia de una cosa.

Para fines del programa "Nueva Era", grado en que el programa satisface las necesidades por las que fue establecido, a saber, las necesidades educativas del alumnado: integración a la institución, confirmación vocacional, etc. [6]

Cobertura

Cantidad o porcentaje abarcado por una cosa o una actividad.

Para fines del programa "Nueva Era", número de alumnos que la tutoría alcanza en relación al número total de alumnos de la escuela. El alcance, desde luego, puede ser en distintos grados, desde una ausencia hasta un contacto plenamente satisfactorio. [6]

Utilidad

Provecho, conveniencia, interés o fruto que se saca de algo.

Para fines del programa "Nueva Era", impacto educativo del programa, en términos de la mejora en el rendimiento escolar y la formación integral del alumno, así como en el rendimiento de la institución. [6]

Deserción

Forma de abandono de los estudios superiores, adopta distintas modalidades en los estudiantes que afecta la continuidad de sus trayectorias escolares. [1]

Estas modalidades incluyen:

- Abandono o suspensión voluntaria y definitiva de los estudios y del sistema de educación superior por parte del alumno. [1]
- Salida de alumnos debido a deficiencias académicas y bajo rendimiento escolar. [1]
- Cambio de carrera (el alumno continúa en la misma institución pero se incorpora a otro cohorte generacional) o de institución. [1]
- Baja de los alumnos que alteran el orden y la disciplina institucional, y que generalmente obstaculiza su ingreso a otra escuela o facultad. [1]

Rezago

Consiste en el avance irregular en la inscripción a las asignaturas subsecuentes del plan de estudios. [1]

Eficiencia terminal

Proporción de alumnos que concluyen el programa en el plazo establecido en el plan de estudios. [1]

Aprendizaje

Proceso a través del cual se adquieren nuevas habilidades, destrezas, conocimientos, conductas o valores como resultado del estudio la experiencia, la instrucción y la observación. El aprendizaje humano está relacionado con la educación y el desarrollo personal. Debe estar orientado adecuadamente y es favorecido cuando el individuo está motivado.

Aprovechamiento del estudiante

Uso que el estudiante hace de los recursos del proceso de aprendizaje presentados en un salón de clases.

Disposición

Para el estudiante: capacidad y actitud que se muestra para el buen desempeño o ejercicio de las actividades escolares.

Para el tutor: capacidad y actitud que se muestra para atender a los alumnos

Interés

Para el estudiante: Inclinação del ánimo del estudiante hacia las actividades educativas que realiza dentro y fuera de la institución.

Para el tutor: Inclinação del ánimo del tutor hacia los problemas académicos y personales que afectan el rendimiento de los alumnos.

Comunicación entre el estudiante y el tutor

Forma en la que se establecen lazos entre el estudiante y el tutor; existen muchas formas de comunicación, gestual, a través de los signos, verbal, escrito, etc.

Confianza hacia el tutor

Esperanza firme que el estudiante tiene del tutor para transmitir sus logros, anhelos y problemas tanto dentro como fuera de la institución.

Orientación hacia el estudiante

Es un conjunto de actividades destinadas al alumno para atender las necesidades de orden personal y de salud que están fuera del ámbito de competencia del tutor.

Estimulación hacia el estudiante

Incitar, excitar con viveza a la ejecución de actividades destinadas a los **alumnos** para mejorar su desempeño educativo.

Seguimiento sistemático de las actividades en la tutoría

Modelo que implica la vigilancia y observación detallada de manera secuencial y organizada para alcanzar el cumplimiento de actividades y acuerdos establecidos entre el estudiante y su tutor.

Puntualidad

Cuidado y diligencia en hacer las cosas a su debido tiempo, así como de llegar a un lugar o partir de él a la hora convenida.

Improvisación

Realizar algo de pronto, sin estudio ni preparación para dicha tarea.

Evaluación

Estimar, apreciar, calcular el valor y rendimiento de algo o alguien.

Conciencia

Propiedad del espíritu humano de reconocerse en sus atributos esenciales y en todas las modificaciones que en sí mismo experimenta.

Relaciones humanas

Conjunto de interacciones que se da entre los individuos de una sociedad. Las relaciones humanas se basan principalmente en los vínculos existentes entre los miembros de la sociedad, gracias a la comunicación.

Planeación

Implica tener uno o varios objetivos a realizar junto con las acciones requeridas para concluirse exitosamente. Va de lo más simple a lo complejo, dependiendo el medio a aplicarse. La acción de planear en la gestión se refiere a planes y proyectos en sus diferentes, ámbito, niveles y actitudes.

Definición de objetivos

Establecer clara y puntualmente los propósitos que se desean alcanzar en un tiempo determinado con la inversión de ciertos recursos.

Participación escolar

Tomar parte en las actividades destinadas al mejoramiento de la educación.

1.1.3 Delimitación y enunciado del problema

Conforme se ha formalizado el programa de tutoría de la Facultad de Ingeniería de la Universidad Nacional Autónoma de México, se ha logrado recopilar la información de los cuestionarios de evaluación del programa Tutoría “Nueva era” de los alumnos de la generación del 2010-1, así mismo se cuenta con los resultados de la encuesta socioeconómica y los resultados tanto del semestre 2010-1 como del examen diagnóstico correspondientes a la misma generación. Se tiene con ello, la necesidad de aplicar minería de datos que ayude a tomar mejores decisiones con la finalidad de un mejor desempeño y aprovechamiento académico de los alumnos.

1.2 Objetivos

El presente trabajo tiene como principal objetivo optimizar y mejorar las futuras tomas de decisiones para enriquecer el programa de tutoría y con ello aumentar el aprovechamiento de los alumnos, reflejado no solo en el aumento de sus calificaciones sino también en el de sus asistencias a clases. De esta forma el rezago estudiantil habrá de disminuir y por consiguiente los índices de eficiencia terminal mejoraran considerablemente.

1.2.1 Generales

Obtener conocimiento útil y novedoso acerca de las encuestas socioeconómicas y cuestionarios de evaluación realizados por los alumnos de la generación 2010-1 de la Facultad de Ingeniería de la Universidad Nacional Autónoma de México para con ello realizar toma de decisiones más inteligentes y que realmente mejoren el programa de Tutoría “Nueva era”, y con ello la situación académica y humana de los alumnos.

1.2.2 Específicos

Los objetivos expuestos a continuación han de llevarse a cabo con base en la minería de datos que se realizará en el presente trabajo:

- Demostrar la utilidad del programa de tutoría como herramienta en la mejoría del desempeño académico de los alumnos.
- Propuestas para incrementar la asistencia a la tutoría.
- Propuestas para mejorar la tutoría enfocado al tutor.
- Propuestas para ayudar al tutor con su labor.

1.3 Alcances y limitaciones

La minería de datos tiene un alcance y aplicaciones muy amplias. Puede ser utilizada en cualquier situación donde se requiere encontrar conocimiento en vastas cantidades de datos, puede ser utilizada para la explotación de datos, comportamientos predictivos, ventas cruzadas, formaciones taxonómicas y pronósticos; para el presente trabajo será utilizada para los dos primeros rubros.

Así mismo el trabajo realizado en la presente investigación no está exento de ciertas limitaciones. Estas limitaciones se derivan en algunos casos de la información con que se cuenta en la base de datos (la cual es la integración de diversas bases), ya que aunque la información es extensa y buena, una minería de datos optima requiere que el almacén de datos tenga una gran cantidad de datos (entre más datos mejor minería), lo ideal para la minería sería contar con información de más de una generación de alumnos, de esa manera los patrones que se podrían obtener serían mayores y más precisos que aquellos que se obtendrán analizando una sola generación. Cabe mencionar llegados a este punto que la decisión de sólo tomar como base de información una sola generación (2010-1) está basada en el hecho de que con el paso de los años las encuestas realizadas acerca de la tutoría así como la encuesta socioeconómica han ido evolucionando, esto es que conforme a los años se han ido integrando nuevas secciones a las encuestas, siendo estas más significativas en la encuesta de tutoría y la cual es el pilar de este trabajo.

Una limitante más a mencionar es la propia naturaleza del sistema de gestión de base de datos, debido a que la minería de datos requiere que la información a analizar cuente con ciertas características las consultas a realizar podrían volverse considerablemente grandes y en algunos casos confusas.

Aunado a lo anterior existe una limitante todavía más considerable para el presente trabajo, y es el hecho de que en estos momentos no se cuenta con un sistema que permita identificar qué porcentaje de alumnos le dan continuidad a la tutoría, lo cual no nos permite conocer de manera exacta cual es el impacto de la tutoría en los semestres subsecuentes al primero.

Aun con todo lo antes mencionado es importante aclarar que para el presente trabajo se han considerado modos de resolución a cada una de las limitantes, es decir que se considera que aunque el almacén no es lo suficiente grande como se desea, los resultados que se obtendrán permitirán una toma de decisiones satisfactoria, de igual forma se utilizara un manejador de bases de datos que facilite la minería de datos.

Capítulo 2. Marco teórico del problema

2.1 Conceptos principales

Sistema de gestión de base de datos

Un sistema gestor de bases de datos (SGBD) consiste en una colección de datos interrelacionados y una colección de programas para acceder a esos datos. [8]

El objetivo principal de un SGBD es proporcionar un entorno que sea tanto conveniente como eficiente para las personas que lo usan para la recuperación y almacenamiento de la información. Una de las principales razones de usar SGBDs es tener un control centralizado tanto de los datos como de los programas que acceden a esos datos. [8]

Bases de datos relacionales

Es una base de datos que cumple con el modelo relacional, el cual es el modelo más utilizado en la actualidad para implementar bases de datos ya planificadas. Permiten establecer interconexiones (relaciones) entre los datos (que están guardados en tablas), y a través de dichas conexiones relacionar los datos de ambas tablas, de ahí proviene su nombre: "Modelo Relacional".

Aunque las bases de datos relacionales son la fuente de datos para la mayoría de aplicaciones de minería de datos, las técnicas de minería no son capaces de trabajar con toda la base de datos ó almacén de datos, sino que sólo son capaces de tratar con una sola tabla ó vista minable. [9]

Modelo entidad - relación

El modelo de datos entidad-relación (E-R) está basado en una percepción del mundo real que consta de una colección de objetos básicos, llamados entidades, y de relaciones entre estos objetos. Una entidad es una «cosa» u «objeto» en el mundo real que es distinguible de otros objetos. [8]

Sus principales características son:

Entidades: Cualquier información referente a un objeto que queremos almacenar

Atributos: Características del objeto que se quiere almacenar

Relaciones: Como interactúan unos con otros

Cardinalidad: Relación entre entidades

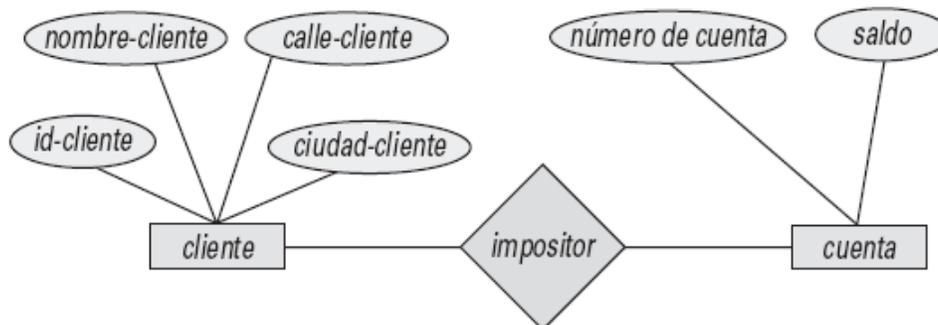


Figura 2.1.a Ejemplo del Modelo E-R [8]

Base de datos no volátil

La información no se modifica ni se elimina una vez almacenado un dato, éste se convierte en información de sólo lectura, y se mantiene para futuras consultas. [10]

Base de datos intratable

Una base de datos intratable es aquella en la que las fases de análisis, diseño y desarrollo de la misma no han sido realizadas por un profesional de sistemas que conozca los métodos y procedimientos de cada una de dichas fases, haciendo que el tratamiento y consulta de los datos sea intratable. [10]

Almacenes de datos

Conjunto de datos históricos, internos o externos y descriptivos de un área de estudio, que están:

- integrados. [9]
- organizados. [9]
- variante en el tiempo. [9]

Para aplicar herramientas para:

- resumir [9]
- describir. [9]
- analizar los datos. [9]

Con el fin de ayudar en la toma de decisiones estratégicas.

Las ventajas fundamentales de un almacén de datos son:

- Su diseño específico. [9]
- Su separación de la base de datos transaccional. [9]

Diccionario de datos

- Descripción externa conceptual e interna de la base de datos. [11]
- Descripción de entidades, atributos y entidades externas. [11]
- Sinónimos de los datos contenidos en la base de datos. [11]
- Códigos de autorización y perfiles de los usuarios, así como los privilegios sobre la base de datos. [11]

Lenguaje de definición de datos

Un lenguaje de definición de datos (DDL) es un lenguaje proporcionado por el sistema de gestión de base de datos que permite a los usuarios de la misma llevar a cabo las tareas de definición de las estructuras que almacenarán los datos así como de los procedimientos o funciones que permitan consultarlos. [8]

Lenguaje de manipulación de datos

Un lenguaje de manipulación de datos (DML) es un lenguaje que permite a los usuarios acceder o manipular los datos organizados mediante el modelo de datos apropiado. [8]

Hay dos tipos básicamente:

- DMLs procedimentales. Requieren que el usuario especifique qué datos se necesitan y cómo obtener esos datos. [8]
- DMLs declarativos (también conocidos como DMLs no procedimentales). Requieren que el usuario especifique qué datos se necesitan sin especificar cómo obtener esos datos. [8]

Son DMLs: Select, Insert, Delete y Update [8]

Lenguaje de control de datos

Un Lenguaje de Control de Datos (DCL) es un lenguaje proporcionado por el Sistema de Gestión de Base de Datos que incluye una serie de comandos SQL que permiten al administrador controlar los privilegios en la Base de Datos. [11]

Lenguaje de control de transacciones

El Lenguaje de Control de Transacciones (Transaction Control Language – *TCL*) se utiliza para administrar los procesos transaccionales en una base de datos en relación a los requerimientos de atomicidad, consistencia, aislamiento y durabilidad, de igual forma permite regresar a su estado original cualquier transacción si esta no se lleva a cabo. [11]

Minería de datos

La minería de datos es el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos. Es decir, la tarea fundamental de la minería de datos es encontrar modelos inteligibles a partir de los datos. Para que este proceso sea efectivo debería ser automático o semi-automático y el uso de los patrones descubiertos debería ayudar a tomar decisiones más seguras que reporten, por tanto, algún beneficio a la organización.

De esta manera, dos son los retos de la minería de datos: por un lado, trabajar con grandes volúmenes de datos, procedentes mayoritariamente de sistemas de información, con los problemas que ello conlleva (ruido, datos ausentes, intratabilidad, volatilidad de los datos) y por el otro usar técnicas adecuadas para analizar los mismos y extraer conocimiento novedoso y útil. En muchos casos la utilidad del conocimiento minado está íntimamente relacionada con la comprensibilidad del modelo inferido.

De una manera simplista pero ambiciosa, se puede decir que el objetivo de la minería de datos es convertir datos en conocimiento. [9]

Modelo de minería de datos

Patrones y tendencias que existen en los datos. Los modelos de minería de datos se pueden aplicar a situaciones empresariales como las siguientes:

- Predecir ventas. [12]
- Dirigir correo a clientes específicos. [12]
- Determinar los productos que se pueden vender juntos. [12]
- Buscar secuencias en el orden en que los clientes agregan productos a una cesta de compra. [12]

Sistema de minería de datos

Es una tecnología de soporte para el usuario final cuyo objetivo es extraer conocimiento útil y utilizable a partir de la información contenida en las bases de datos de las empresas. [9]

Vista SQL

Relaciones que no forman parte del modelo lógico pero se hacen visibles a los usuarios como relaciones virtuales. Se puede trabajar con gran número de vistas sobre cualquier conjunto dado de relaciones reales. [8]

Vista Minable

Es la combinación en una sola tabla de la información de varias tablas que requiramos para cada tarea concreta de minería de datos. [9]

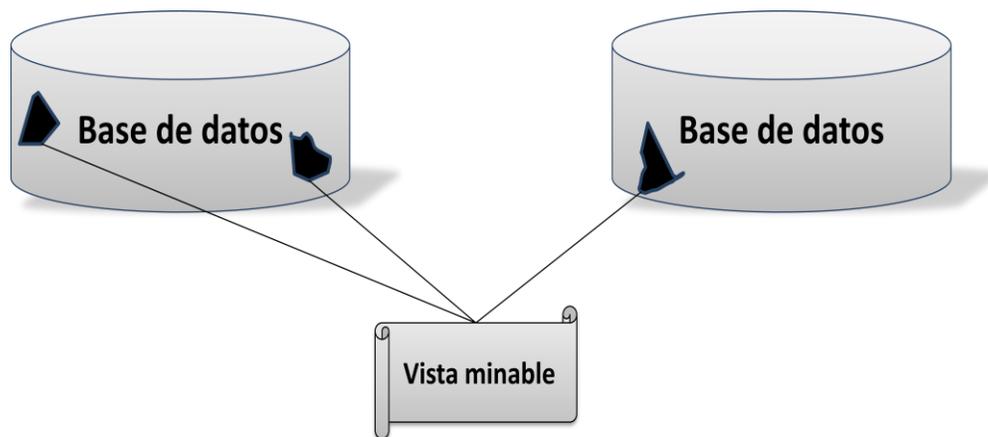


FIGURA 2.1.b Vista Minable

Herramientas de minería de datos

Las herramientas de minería de datos sirven para predecir tendencias y comportamientos. De esta manera permiten a las organizaciones tomar decisiones proactivas para adaptarse rápidamente a los cambios del mercado obteniendo así ventajas competitivas.

Las herramientas de minería de datos pueden responder a preguntas de negocios que tradicionalmente consumen demasiado tiempo para poder ser resueltas por consultas en un sistema tradicional de soporte operacional. La potencialidad de estas herramientas reside en la capacidad de explorar las bases de datos en busca de patrones ocultos, encontrando información predecible que para un experto sería casi imposible debido al gran volumen de información.

Algunas herramientas de minería de datos son:

Matlab: Es un poderoso entorno de cálculo técnico integrado que combina el cálculo numérico, gráficos avanzados y visualización, y un lenguaje de programación de alto nivel. [13]

Weka: Es una extensa colección de algoritmos de máquinas de conocimiento desarrollados por la Universidad de Waikato (Nueva Zelanda) implementados en Java; útiles para ser aplicados sobre datos mediante las interfaces que ofrece. Weka ofrece también las herramientas necesarias para realizar transformaciones sobre los datos, tareas de clasificación, regresión, agrupamiento, asociación y visualización. [14]

Aprendizaje automático

Área de inteligencia artificial que se ocupa de desarrollar algoritmos (programas) capaces de aprender, y constituye, junto con la estadística, el corazón del análisis inteligente de los datos.

Los principios seguidos en el aprendizaje automático y en la minería de datos son los mismos: la máquina aprende un modelo a partir de ejemplos y lo usa para resolver el problema. [9]

Sistemas de aprendizaje automático

Tienen la intención de construir un modelo a partir de datos de entrada y cambiar su comportamiento de tal manera que son capaces de clasificar nuevos datos y desarrollarse mejor que en antiguas situaciones. [15]

Técnicas de minería de datos

Son el resultado de un largo proceso de investigación y desarrollo de productos orientados al almacenamiento, extracción y análisis de datos. [9]

Tipos de Datos

La minería de datos puede aplicarse a cualquier tipo de información, siendo las técnicas de minería diferentes para cada una de ellas. [9]

Datos Estructurados	Datos no Estructurados
Bases de datos relacionales Bases de datos espaciales Bases de datos temporales Bases de datos textuales	Web Multimedia Otros tipos de repositorios de documentos

FIGURA 2.1.c Tipos de datos

Tipos de modelos producidos por la minería de datos

Modelos Predictivos:

Son aquellos que pretenden estimar valores futuros o desconocidos de variables de interés, que denominamos variables objetivo o dependientes (target), usando otras variables o campos de la base de datos, a las que nos referiremos como variables independientes o predictivas. [9]

Modelos Descriptivos:

Son aquellos que pretenden identificar patrones que explican o resumen los datos, además sirven para explorar las propiedades de los datos. [9]

Los sistemas para la toma de decisión

Son las diversas herramientas y sistemas que asisten a los directivos en la resolución de problemas y en la toma de decisiones. Su objetivo es proporcionar la información necesaria para realizar decisiones efectivas en el ámbito empresarial o en tareas de diagnóstico (por ejemplo en medicina). [9]

Visualización de datos

Son técnicas de visualización que permiten al usuario:

- Descubrir.
- Intuir.
- Entender patrones.

Que serían más difíciles de “ver” a partir de descripciones matemáticas o textuales de los resultados. [9]

Fases de Descubrimiento de conocimiento en bases de datos (*Knowledge Discovery from Databases - KDD*)

Se define el KDD como “el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos”. En esta definición se resumen cuáles deben ser las propiedades deseables del conocimiento extraído:

- **Válido:** hace referencia a que los patrones deben seguir siendo válidos para datos nuevos (con un cierto grado de certidumbre), y no sólo para aquellos que han sido usados en su obtención. [9]
- **Novedoso:** que aporte algo desconocido tanto para el sistema y preferiblemente para el usuario. [9]
- **Potencialmente útil:** la información debe conducir a acciones que reporten algún tipo de beneficio para el usuario. [9]
- **Comprensible:** la extracción de patrones no comprensibles dificulta o imposibilita su interpretación, revisión, validación y uso en la toma de decisiones. De hecho, una información incomprensible no proporciona conocimiento (al menos desde el punto de vista de su utilidad). [9]

Como se deduce de la anterior definición, el KDD es un proceso complejo que incluye no sólo la obtención de los modelos o patrones (el objetivo de la minería de datos), sino también la evaluación y posible interpretación de los mismos. [9]

A continuación se muestra un diagrama, el cual muestra las diferentes fases del KDD

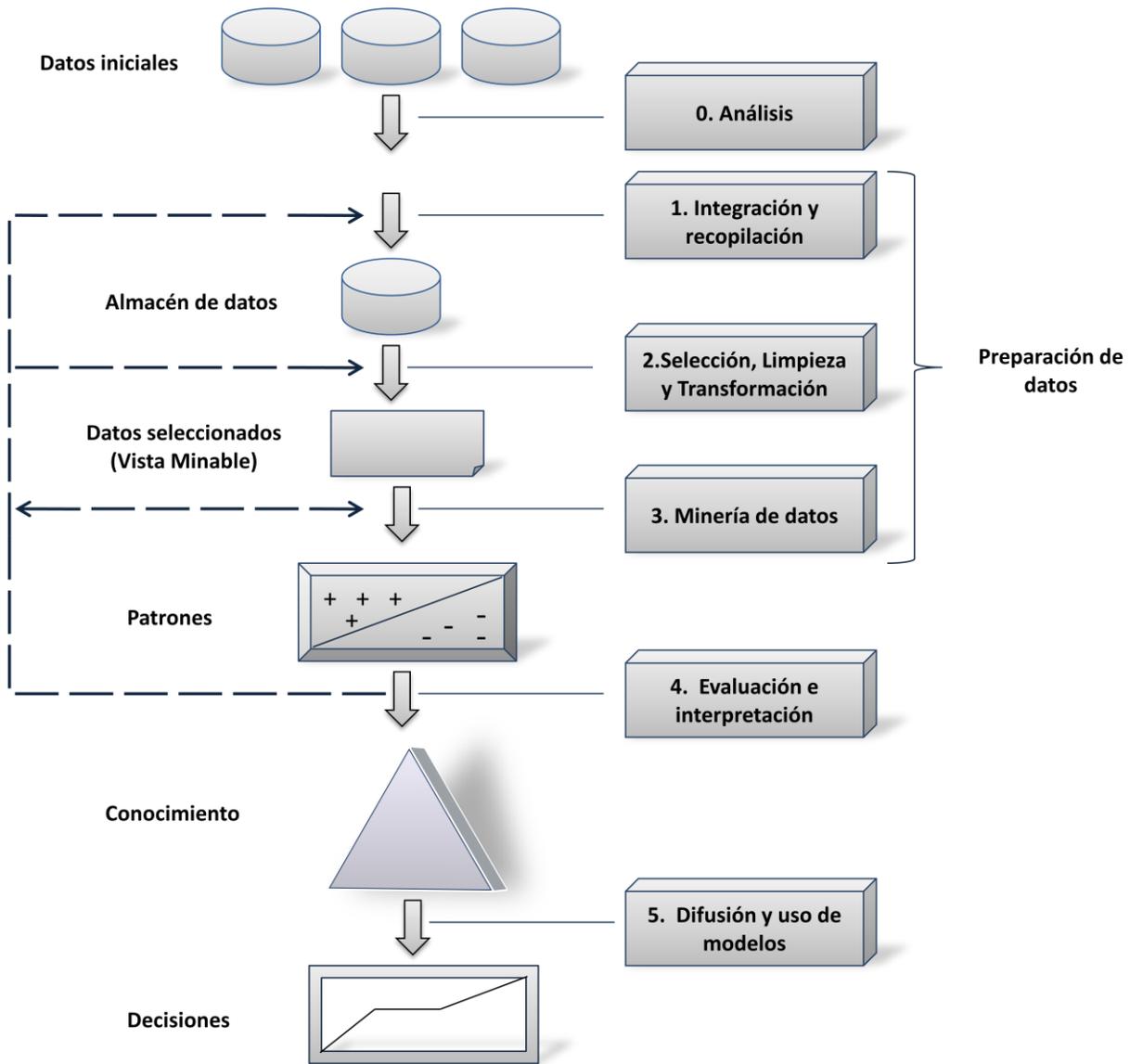


FIGURA 2.1.d KDD [9]

Así, los sistemas de KDD permiten la selección, limpieza, transformación y proyección de los datos; analizar los datos para extraer patrones y modelos adecuados; evaluar e interpretar los patrones para convertirlos en conocimiento; consolidar el conocimiento resolviendo posibles conflictos con conocimiento previamente extraído; y hacer el conocimiento disponible para su uso. [9]

Este proceso propone no sólo la definición de los problemas, sino también la obtención de los modelos, la evaluación y posible interpretación de los mismos. Sin embargo todo lo anterior no se muestra explícitamente, de manera que a continuación se muestra un esquema que se seguirá para cada uno de los análisis principales, cabe mencionar que dicho esquema parte después del segundo bloque y de igual forma que el primero no volverán a ser requeridos, debido a las limitantes mismas de este trabajo :

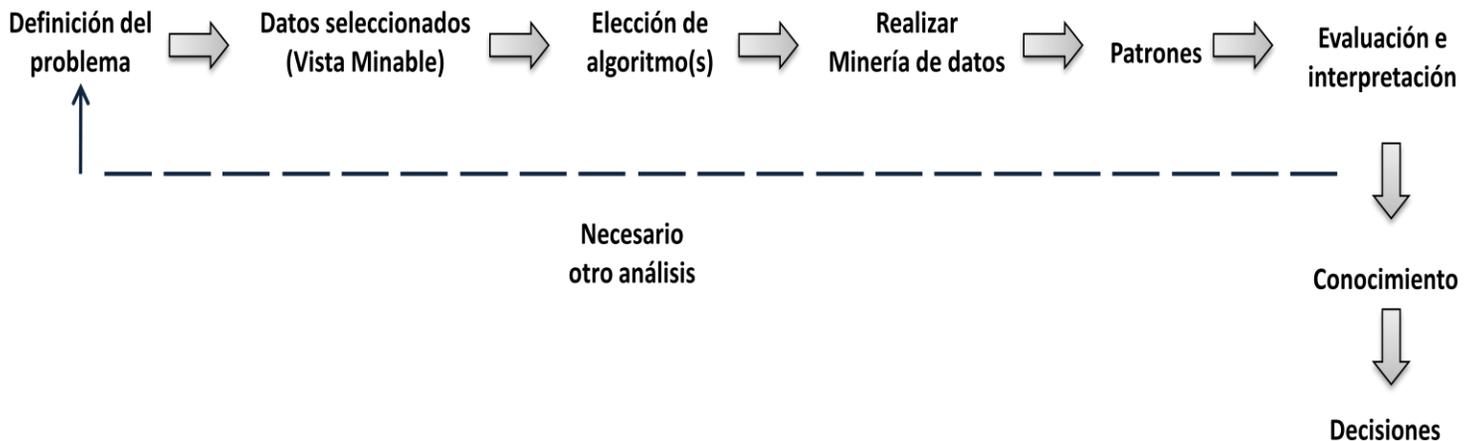


FIGURA 2.1.e Esquema para análisis

Definición del problema: Enunciado que describe la situación a ser tratada.

Datos seleccionados (Vista minable): Descripción (así como código) de la consulta que se utilizará para obtener los datos necesarios para la resolución del problema.

Elección del algoritmo (s): Exposición de razones para utilizar el o los algoritmos que brindaran los patrones a analizar posteriormente.

Minería de datos: Proceso de minería que consiste en la creación de la vista de origen de datos, selección de algoritmo (el cual se eligió en el punto anterior), características del mismo y descripción de los datos a procesar.

Patrones: Graficas, reglas y/o tablas con los resultados de la minería de datos.

Evaluación e interpretación: Define la precisión del análisis realizado utilizando los datos destinados a las pruebas, así mismo define si el análisis es suficiente para obtener el conocimiento necesario para resolver el problema, en caso de no existir patrones significativos es forzoso realizar otro análisis, cambiando los datos de entrada y/o salida o bien cambiando el algoritmo empleado.

Conocimiento: Descripción de la información obtenida durante todo el análisis. En este punto ya no es necesario ningún nuevo análisis pero tampoco se toman decisiones.

Decisiones: Finalización del análisis, en este punto se toman las decisiones basadas en el conocimiento adquirido en el apartado anterior, y las cuales buscan dar una solución al problema inicialmente planteado.

El esquema anterior puede ser asemejado al proceso de 5 pasos [17] para la toma de decisiones y resolución de problemas donde:

- 1^{er} Paso. Definir el problema = Definición del problema y datos seleccionados (Vista Minable)
- 2^o Paso. Buscar alternativas = Elección de algoritmos y realizar minería de datos
- 3^{er} Paso. Valorar las consecuencias de cada alternativa = patrones, evaluación e interpretación
- 4^o Paso. Elegir la mejor alternativa = Conocimiento
- 5^o Paso. Aplicar alternativa escogida = Decisiones

Media

Denominada más frecuentemente como “promedio”. Es la suma de los valores obtenidos entre la cuenta de observaciones. [16]

Mediana

Es el valor medio de un conjunto de números. La mitad de los números tiene un mayor valor que la mediana, y la otra mitad tiene n valor inferior a la mediana. [16]

Desviación estándar

Medida de la dispersión de los datos. De manera específica, las medidas de desviación estándar miden la manera en que se extienden los valores de datos a partir de la media. [16]

Rango

La diferencia entre el valor más pequeño y es más grande de un valor de datos. [16]

Distribución

La serie de valores observada para un atributo determinado en sus datos. La distribución suele mostrarse de manera útil utilizando histogramas o gráficas circulares. [16]

Datos continuos

Datos cuyo espacio muestral especificado como dominio no es numerable. Por ejemplo, todos envejecemos en fracciones de segundo. [16]

Datos discretos

Datos cuyo espacio muestral especificado como dominio es numerable. Algunos ejemplos son país de residencia o estado civil. [16]

Algoritmo

Método para resolver un problema, debe presentarse como una secuencia ordenada de instrucciones que siempre se ejecutan en un tiempo finito y con una cantidad de esfuerzo también finito. En un algoritmo siempre debe haber un punto de inicio y un punto de terminación, estos deben ser únicos y deben ser fácilmente identificables. [9]

Todo algoritmo debe cumplir las siguientes características:

- A. Debe ser Preciso; esto es, debe especificar sin ambigüedad el orden en que se deben ejecutar las instrucciones.
- B. Debe estar Definido; esto es, cada vez que se ejecute bajo las mismas condiciones, la secuencia de ejecución deberá ser la misma proporcionándonos el mismo resultado.
- C. Debe ser Finito; esto es, siempre que sea adecuado se realizarán un número finito de instrucciones, en un tiempo finito y requiriendo una cantidad finita de esfuerzo.

Proceso incremental

Proceso por el cual un proyecto es realizado por partes y que al final terminará siendo la solución completa requerida para un problema. [9]

Procesos Iterativos

Proceso en el que un sistema mejora su funcionalidad durante cada iteración. [9]

2.2 Definición de tareas de minería

Tareas y métodos

Una tarea de minería de datos es un (tipo de) problema de minería de datos. Por ejemplo, “clasificar las piezas del proveedor Minatronix en óptimas, defectuosas reparables y defectuosas irreparables” es una tarea de clasificación, que podría resolverse mediante árboles de decisión o redes neuronales, entre otros métodos. Es muy importante distinguir el problema de los métodos para solucionarlo. Una tarea puede tener muchos métodos diferentes para resolverla y el mismo método puede resolver muchas tareas. [9]

Tareas predictivas

Se trata de problemas en los que hay que predecir uno o más valores para uno o más ejemplos. [9]

Tareas descriptivas

Los ejemplos se presentan como un conjunto sin etiquetar ni ordenar de ninguna manera, el objetivo, es describir los existentes. [9]

Se pueden considerar tareas descriptivas:

- A) Las tablas de frecuencias.
- B) El análisis de componentes principales) .
- C) Agrupamiento (clustering).

Su objetivo es obtener grupos o conjuntos entre los elementos, de tal manera que los elementos asignados al mismo grupo sean similares.



FIGURA 2.2.a Tareas de minería

Tareas de clasificación

La clasificación (predictiva) es quizá la tarea más utilizada.

- Cada instancia (o registro de la base de datos) pertenece a una clase, la cual se indica mediante el valor de un atributo que llamamos la clase de la instancia.
- El resto de los atributos de la instancia (los relevantes a la clase) se utilizan para predecir la clase.
- El objetivo es predecir la clase de nuevas instancias de las que se desconoce la clase.

Ejemplo: Un oftalmólogo desea clasificar nuevos pacientes, para decidir si es conveniente operarlos o no en función de una base de datos de sus antiguos pacientes. [9]

Tareas de regresión

La regresión (predictiva), la principal diferencia respecto a la clasificación es que el valor a predecir es numérico.

- Su objetivo es minimizar el error entre el valor predicho y el valor real.

Ejemplo: Un empresario quiere conocer cuál es el costo de un nuevo contrato basándose en los datos correspondientes a contratos anteriores. [9]

Tareas de agrupamiento (segmentación)

El agrupamiento (clustering) es descriptiva, consiste en obtener grupos “naturales” a partir de los datos. Se forman grupos tales que los objetos de un mismo grupo son muy similares entre sí y, al mismo tiempo, son muy diferentes a los objetos de otro grupo. [9]

Ejemplo, una librería identifica grupos de clientes en base a sus preferencias, para que recomiende otros libros comprados por clientes de su mismo grupo. [9]

Tareas de asociación

Las reglas de asociación (descriptiva) son similares a las de las correlaciones, tienen como objetivo identificar relaciones no explícitas entre atributos categóricos.

Estas reglas pueden ser de muchas formas, aunque la formulación más común es del estilo

“Si el atributo X toma el valor d entonces el atributo Y toma el valor b”.

Las reglas de asociación no implican una relación causa-efecto, es decir, puede no existir una causa para que los datos estén asociados.

Por ejemplo en el análisis de la cesta de la compra, para identificar productos que son frecuentemente comprados juntos, información que puede usarse:

- Para ajustar los inventarios.
- Para la organización física del almacén.

- En campañas publicitarias.

Ejemplo: En una tienda de electrodomésticos, se analizan ventas y se descubre que el 30% de los clientes que compraron un televisor hace 6 meses compraron un DVD en los siguientes 2 meses. [9]

Tareas de correlación

Las correlaciones (descriptiva) se usan para examinar el grado de similitud de los valores de dos variables numéricas, para medir la correlación lineal es con el coeficiente de correlación r , -1 y 1 . [9]

- Si es 0 no hay correlación.
- Cuando r es positivo, las variables tienen un comportamiento similar (ambas crecen o decrecen al mismo tiempo).
- Cuando r es negativo si una variable crece la otra decrece.

Ejemplo: Un inspector de incendios obtiene correlaciones negativas entre el empleo de aisladores y la frecuencia de incendios. [9]

Algoritmos (técnicas) de minería de datos

El algoritmo de minería de datos es el mecanismo que crea modelos de minería de datos. Para crear un modelo, un algoritmo analiza primero un conjunto de datos, buscando patrones y tendencias específicos. Después, el algoritmo utiliza los resultados de este análisis para definir los parámetros del modelo de minería de datos. [19]

El modelo de minería de datos que crea un algoritmo puede tomar diversas formas, incluyendo:

- Un conjunto de reglas que describen cómo se agrupan los productos en una transacción.
- Un árbol de decisión que predice si un cliente determinado comprará un producto.
- Un modelo matemático que predice las ventas.
- Un conjunto de clústeres que describe cómo se relacionan los escenarios de un conjunto de datos.

La elección del algoritmo apropiado para una tarea específica puede ser un trabajo difícil. Aunque se pueden utilizar diferentes algoritmos para realizar una misma tarea, cada uno de ellos genera un resultado diferente, y algunos pueden generar más de un tipo de resultado.

Tampoco es necesario que los algoritmos sean usados de modo independiente: en una solución de minería de datos se pueden utilizar algunos algoritmos para examinar los datos y, después, usar otros para predecir un resultado específico basándose en esos datos. Por ejemplo, se puede utilizar un algoritmo de clústeres, que reconoce patrones, para dividir los datos en grupos que sean más o menos homogéneos, y luego usar los resultados para crear un mejor modelo de árbol de decisión.

La siguiente figura muestra algunas técnicas de minería de datos así como las tareas en las que pueden ser utilizadas para obtener los modelos predictivos y descriptivos.

NOMBRE DE TÉCNICA	TAREAS PREDICTIVAS		TAREAS DESCRIPTIVAS		
	Clasificación	Regresión	Agrupamiento	Reglas de Asociación	Correlaciones
Redes Neuronales	X	X	X		
Puntuación de grado de interés	X				
Bayesiano con prioridad K2	X				
Regresión lineal y logarítmica		X			X
Kmeans			X		
EM			X		
Apriori			X	X	
Naive Bayes	X				

FIGURA 2.2.b Técnicas y tareas de minería [9]

2.3 Hipótesis

Como principales hipótesis para el presente trabajo de investigación podemos encontrar que:

- La tutoría alcanza a un 70% de la población de alumnos de licenciatura en ingeniería.
- Aquellos alumnos que no obtuvieron una buena calificación en su examen diagnóstico son aquellos que asisten con mayor regularidad a la tutoría.
- La tutoría influye de manera considerable en el promedio del alumno.
- La distancia hogar-universidad es uno de los principales factores en la inasistencia del alumno a la tutoría, entre mas distancia mayor inasistencia.
- La asistencia del alumno a la tutoría va ligada directamente a la disponibilidad del alumno, esto es aquellos que trabajan tienen una asistencia menor que aquellos que no lo hacen.
- La clase social de un alumno no influye en la asistencia del alumno en la tutoría ni su promedio del semestre.
- El desempeño de un alumno va de la mano con la efectividad que tenga un tutor en las sesiones de tutoría.
- Los hábitos de estudio de un alumno son parte fundamental en un buen desempeño académico.

Capítulo 3. Fases de descubrimiento en bases de datos

3.1 Sistema de gestión de bases de datos

La elección del sistema de gestión de bases de datos no es una tarea fácil, si bien todos tienen un mismo objetivo que es definir, construir y manipular una base de datos, es cierto también que algunos poseen características específicas que dependiendo del propósito para que se quieran usar son mejores que otros, por ejemplo algunos sistemas cuentan con una forma más sencilla de generar consultas complejas que otros, en algunos los procedimientos son más avanzados y por mencionar otra característica algunos sistemas de gestión no permiten múltiples desencadenadores, cualquiera de las características antes mencionadas u otras más pueden ser suficiente justificación para elegir un sistema de gestión de bases de datos de otro.

Para el presente trabajo se ha elegido utilizar como principal sistema de gestión de bases de datos a SQL Server 2008, las principales razones son expuestas a continuación:

Microsoft SQL Server es un sistema para la gestión de bases de datos producido por Microsoft basado en el modelo relacional. Sus lenguajes para consultas son T-SQL y ANSI SQL. Microsoft SQL Server constituye la alternativa de Microsoft a otros potentes sistemas gestores de bases de datos como son Oracle, PostgreSQL o MySQL. [20]



FIGURA 3.1.a *Microsoft SQLServer*

Ventajas: [20]

- Soporte de transacciones.
- Escalabilidad, estabilidad y seguridad.
- Soporta procedimientos almacenados.
- Incluye también un potente entorno gráfico de administración, que permite el uso de comandos DDL y DML gráficamente.
- Permite trabajar en modo cliente-servidor, donde la información y datos se alojan en el servidor y los terminales o clientes de la red sólo acceden a la información.
- Además permite administrar información de otros servidores de datos.

Desventajas [20]

- MSSQL no maneja compresión de datos (en *SQL Server* 2005 y 2000, solamente la versión 2008 *Enterprise Edition* incluye esta característica), por lo que ocupa mucho espacio en disco.
- MSSQL está atado a la plataforma del sistema operativo sobre la cual se instala.

3.2 Integración, almacenamiento y limpieza (Etapa 1 y 2 del KDD)

Como se mencionó previamente un almacén de datos es un repositorio de información coleccionada desde varias fuentes, almacenada bajo un esquema unificado que normalmente reside en un único emplazamiento¹ y que nos permiten aplicar herramientas que permiten resumir, describir y analizar datos con el fin de ayudar en la toma de decisiones estratégicas.

En el presente trabajo y para buscar la consecución de los objetivos planteados en el primer capítulo del mismo se recolectaron cinco distintas fuentes de información, buscando con ellas y mediante su integración, almacenamiento y limpieza el obtener un almacén de datos que satisfaga la necesidad de contener en una sola base toda la información requerida para comenzar la minería.

Las bases de información recolectadas son:

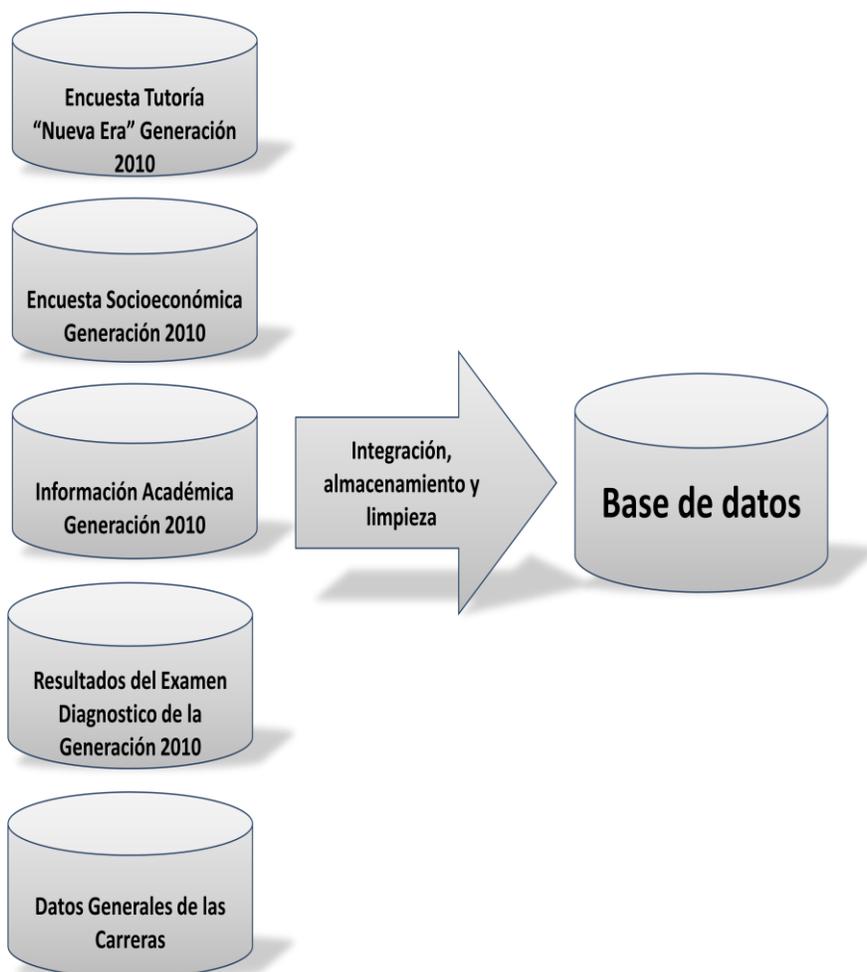


FIGURA 3.2.a Bases de información recolectadas para la minería

¹ A cada uno de los ordenadores que integran un sistema de Bases de Datos distribuido se le conoce como nodo o emplazamiento del sistema y pueden ser administrados de forma diferente.

Los datos de cada una de ellas se muestran en la siguiente tabla:

Base de información	Fuente	Formato	Información contenida
Encuesta tutoría "Nueva Era" Generación 2010	Archivo proporcionado por la Facultad de Ingeniería de la UNAM	Excel	Resultados de la encuesta respecto a la tutoría "Nueva Era" de la generación 2010
Encuesta Socioeconómica Generación 2010	Archivo proporcionado por la Facultad de Ingeniería de la UNAM	Excel	Resultados de la encuesta socioeconómica de la generación 2010
Información Académica Generación 2010	Archivo proporcionado por la Facultad de Ingeniería de la UNAM	Excel	Calificaciones de las materias inscritas de los alumnos de la generación 2010
Datos Generales de las Carreras	Página web de la Facultad de Ingeniería de la UNAM	Contenido web	Información general de cada una de las carreras como nombre, claves, siglas, créditos totales, etc.
Resultados del Examen Diagnostico de la Generación 2010	Archivo proporcionado por la Facultad de Ingeniería de la UNAM	Excel	Resultados del examen diagnostico que realizo la Facultad de Ingeniería a los alumnos de la generación 2010 previo al inicio del semestre

FIGURA 3.2.b Detalle de las bases de información recolectadas para la minería

Hasta este punto se ha hablado acerca de las fuentes de información recolectadas y del sistema de gestión de bases de datos a utilizar, así como las características de los mismos, pero no puede hacerse a un lado uno de los puntos más importantes para que la minería de datos funcione y es la limpieza y transformación de datos.

Partiendo del hecho de que el éxito de un proceso de minería de datos depende, de tener todos los datos necesarios (una buena recopilación) y también una buena limpieza e integración, se puede afirmar que este punto es uno de los pilares más importantes de la minería de dato y es que una mala calidad de la información, trae como consecuencia que los patrones descubiertos contrasten con la realidad, y la realización del proceso sea inútil.

El primer problema al realizar una integración de distintas fuentes de datos son los identificadores únicos. Este problema se conoce como esclarecimiento de la identidad. Para el presente trabajo dicho problema fue resuelto con el número de cuenta del alumno, de esta manera podemos obtener que todas las base de información recolectadas comparten dicho dato, el cual es único e irrepetible.



FIGURA 3.2.c Muestras de las bases de información

El siguiente punto a considerar es que al integrar dos fuentes o más suele suceder que aparezcan datos faltantes o diferentes, por ejemplo para esta investigación tanto la encuesta socioeconómica como la encuesta de tutores contaban con el campo de sexo, al momento de realizar la integración se presentaron casos (pocos) en los que una respuesta para esta opción difirió en ambas encuestas para el mismo identificador del alumno, es decir, mientras en una encuesta el sexo era femenino en otra encuesta aparecía como masculino, la solución que se considero en este trabajo es dejar dicho dato al azar.

En ocasiones para la realización de la minería de datos es recomendable utilizar valores numéricos que valores nominales, por ejemplo en lugar de utilizar un SI o un NO para una respuesta en alguna pregunta de la encuesta sería recomendable utilizar un 1 para SI y un 2 para un NO, es por lo anterior que en la integración de datos se podría incluir una transformación de tal forma que se manejen solo valores numéricos, sin embargo para el presente trabajo y por la descripción del almacén de datos, la elección tomada fue manejar valores nominales, y será hasta la creación de la vista minable cuando dichos valores serán transformados en numéricos. En realidad muchas transformaciones en los datos así como la obtención de otros datos derivados están consideradas para ser realizados en la creación de la vista minable y no antes.

Finalmente y tras realizar todo lo antes mencionado es posible obtener un almacén de datos completo y útil, con información coherente y que permita obtener buenos resultados.

3.3 Descripción del almacén de datos

3.3.1 Modelo Entidad/Relación

A continuación se muestra el modelo entidad/relación de la base de datos relacional diseñada:

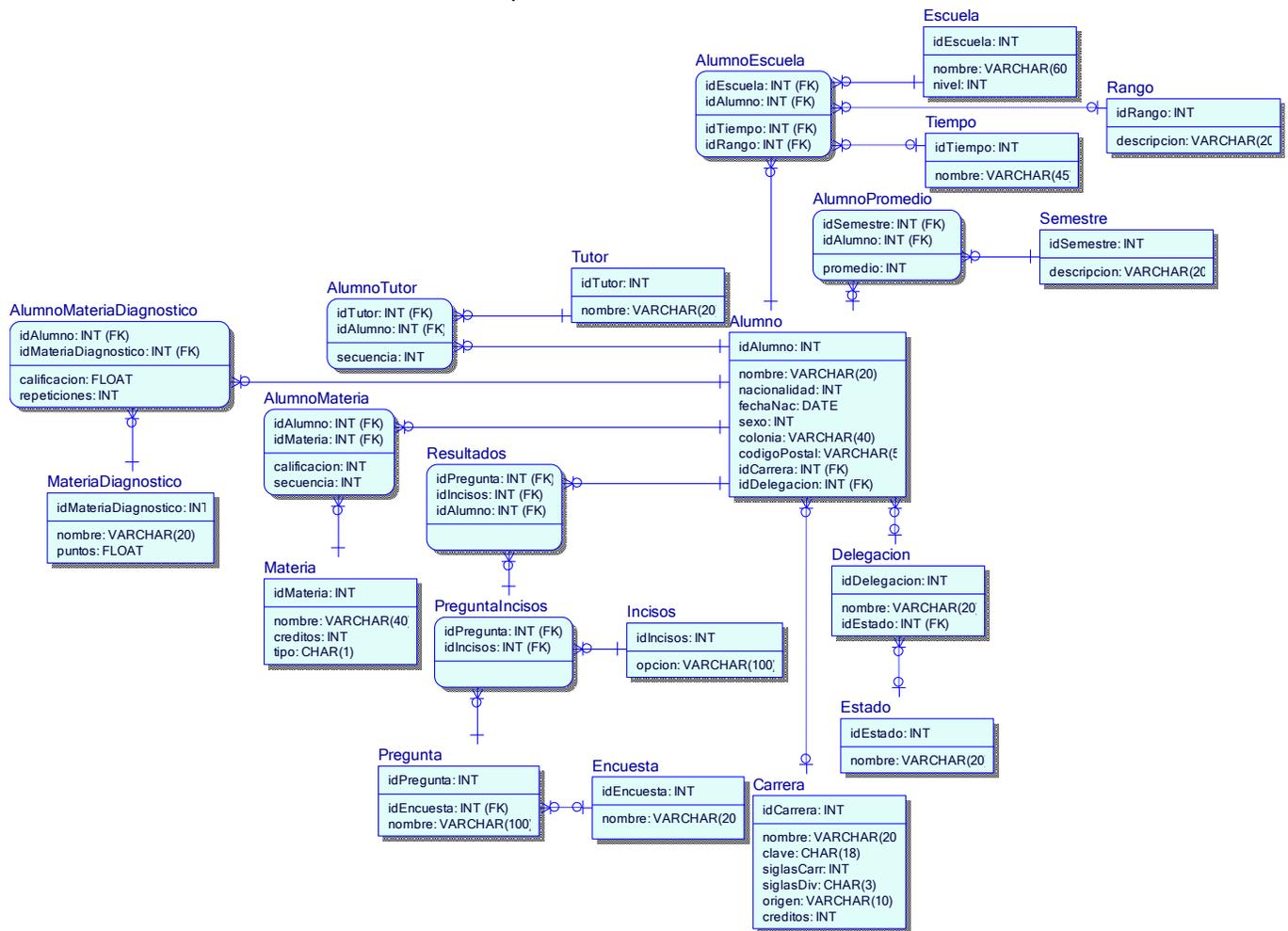


FIGURA 3.3.1.a Modelo Entidad / Relación

3.3.2 Diccionario de datos corporativo

Tabla	Característica
alumno	Tabla que almacena los datos de los alumnos
alumnoescuela	Tabla que almacena alumnos y escuelas a las que asistieron
alumnomateria	Tabla que asocia alumnos y materias
alumnomateriadiagnostico	Tabla que asocia alumnos y materias de diagnostico
alumnopromedio	Tabla que asocia alumnos y semestres (con promedio)
alumnotutor	Tabla que asocia alumnos y tutores
carrera	Catalogo de carreras
delegacion	Catalogo de delegaciones
encuesta	Catalogo de encuestas [1] = En cuesta de tutores [2] = Encuesta Socioeconómica
escuela	Tabla que almacena los datos de las escuelas
estado	Catalogo de estados
incisos	Tabla que almacena los incisos a ser usados en las preguntas
materia	Catalogo de materias
materiadiagnostico	Catalogo de materias de diagnostico
pregunta	Tabla que almacena las preguntas usadas de todas la encuestas
preguntaincisos	Tabla que asocia las preguntas de las encuestas y sus incisos
rango	Tabla que almacena los rango de las calificaciones
resultados	Tabla con las respuestas de los alumnos
semestre	Catalogo de semestres
tiempo	Catalogo con rangos de tiempos en que el alumno curso cada nivel escolar.
tutor	Tabla que almacena los datos de los tutores

Tabla "alumno"

Campo	Tipo de Dato	PK	FK	No nulo	Por defecto	Descripción
idAlumno	int	✓		✓		Identificador del alumno
idDelegacion	int		✓	✓		Identificador de la delegación
nombre	varchar(20)			✓		Nombre del alumno
nacionalidad	int			✓		[1] = Nacionalidad mexicana [2] = Nacionalidad extranjera
fechaNac	date			✓		En formato YYYY-MM-DD
sexo	int			✓		[1] = Sexo masculino [2] = Sexo femenino
colonia	varchar(40)					Colonia
codigoPostal	varchar(5)					CP del alumno
idCarrera	int		✓	✓		Identificador de su carrera

Definición:

```
CREATE TABLE alumno (
  idAlumno      int IDENTITY(1,1) NOT NULL,-- COMMENT 'Identificador del alumno',
  idDelegacion  int NOT NULL,-- COMMENT 'Identificador de la delegación ',
  nombre        varchar(20) NOT NULL,-- COMMENT 'Nombre del alumno',
  nacionalidad  int NOT NULL,-- COMMENT ' [1] = Nacionalidad mexicana [2] = Nacionalidad
extranjera',
  fechaNac      date NOT NULL,-- COMMENT 'En formato YYYY-MM-DD',
  sexo          int NOT NULL,-- COMMENT '[1] = Sexo masculino [2] = Sexo femenino',
  colonia        varchar(40),-- COMMENT 'Colonia',
  codigoPostal  varchar(5),-- COMMENT 'CP del alumno',
  idCarrera     int NOT NULL,-- COMMENT 'Identificador de su carrera',
  /* Keys */
  PRIMARY KEY (idAlumno),
  /* Foreign keys */
  CONSTRAINT alumno_ibfk_1
    FOREIGN KEY (idDelegacion)
    REFERENCES delegacion(idDelegacion)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION,
  CONSTRAINT foreign_key01
    FOREIGN KEY (idCarrera)
    REFERENCES carrera(idCarrera)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION
);

CREATE INDEX Alumno_FKIndex1
  ON alumno
  (idDelegacion);
CREATE INDEX Alumno_FKIndex2
  ON alumno
  (idCarrera);
```

Tabla "alumnoescuela"

Campo	Tipo de Dato	PK	FK	No nulo	Por defecto	Descripción
idEscuela	int	✓	✓	✓		Id de la escuela
idAlumno	int	✓	✓	✓		Id del alumno
idTiempo	int		✓	✓		Id del tiempo cursado
idRango	int		✓	✓		Id del rango de cal.

Definición:

```
CREATE TABLE alumnoescuela (
  idEscuela  int NOT NULL ,--COMMENT 'Id de la escuela',
  idAlumno   int NOT NULL ,--COMMENT 'Id del alumno',
  idTiempo   int NOT NULL ,--COMMENT 'Id del tiempo cursado',
  idRango    int NOT NULL ,--COMMENT 'Id del rango de cal.',
  /* Keys */
  PRIMARY KEY (idEscuela, idAlumno),
  /* Foreign keys */
```

```

CONSTRAINT FK_alumnoescuela_1
  FOREIGN KEY (idAlumno)
  REFERENCES alumno(idAlumno)
  ON DELETE NO ACTION
  ON UPDATE NO ACTION,
CONSTRAINT FK_alumnoescuela_2
  FOREIGN KEY (idTiempo)
  REFERENCES tiempo(idTiempo)
  ON DELETE NO ACTION
  ON UPDATE NO ACTION,
CONSTRAINT FK_alumnoescuela_3
  FOREIGN KEY (idEscuela)
  REFERENCES escuela(idEscuela)
  ON DELETE NO ACTION
  ON UPDATE NO ACTION,
CONSTRAINT FK_alumnoescuela_4
  FOREIGN KEY (idRango)
  REFERENCES rango(idRango)
  ON DELETE NO ACTION
  ON UPDATE NO ACTION
);

CREATE INDEX AlumnoEscuela_FKIndex3
  ON alumnoescuela
  (idRango);

CREATE INDEX AlumnoEscuela_FKIndex4
  ON alumnoescuela
  (idTiempo);

CREATE INDEX Alumno_has_Escuela_FKIndex1
  ON alumnoescuela
  (idAlumno);

CREATE INDEX Alumno_has_Escuela_FKIndex2
  ON alumnoescuela
  (idEscuela);

```

Tabla "alumnomateria"

Campo	Tipo de Dato	PK	FK	No nulo	Por defecto	Descripción
idAlumno	int	✓	✓	✓		Identificador del alumno
idMateria	int	✓	✓	✓		Identificador de la materia
calificacion	int					[0] = NP
secuencia	int			✓	1	VeZ que el alumno la cursa

Definición:

```

CREATE TABLE alumnomateria (
  idAlumno      int NOT NULL ,--COMMENT 'Identificador del alumno',
  idMateria     int NOT NULL ,--COMMENT 'Identificador de la materia',
  calificacion  int ,--COMMENT '[0] = NP',
  secuencia     int NOT NULL DEFAULT '1' ,--COMMENT 'VeZ que el alumno la cursa',
  /* Keys */

```

```

PRIMARY KEY (idAlumno, idMateria),
/* Foreign keys */
CONSTRAINT alumnomateria_ibfk_1
FOREIGN KEY (idAlumno)
REFERENCES alumno(idAlumno)
ON DELETE NO ACTION
ON UPDATE NO ACTION,
CONSTRAINT alumnomateria_ibfk_2
FOREIGN KEY (idMateria)
REFERENCES materia(idMateria)
ON DELETE NO ACTION
ON UPDATE NO ACTION
);

CREATE INDEX Alumno_has_Materia_FKIndex1
ON alumnomateria
(idAlumno);

CREATE INDEX Alumno_has_Materia_FKIndex2
ON alumnomateria
(idMateria);

```

Tabla "alumnomateriadiagnostico"

Campo	Tipo de Dato	PK	FK	No nulo	Por defecto	Descripción
idAlumno	int	✓	✓	✓		Identificador del alumno
idMateriaDiagnostico	int	✓	✓	✓		Identificador de la materia
calificacion	float					Calificación del alumno
repeticiones	int					Repeticiones que realizo

Definición:

```

CREATE TABLE alumnomateriadiagnostico (
  idAlumno int NOT NULL ,--COMMENT 'Identificador del alumno',
  idMateriaDiagnostico int NOT NULL ,--COMMENT 'Identificador de la materia',
  calificacion float ,--COMMENT 'Calificación del alumno',
  repeticiones int ,--COMMENT 'Repeticiones que realizo',
/* Keys */
PRIMARY KEY (idAlumno, idMateriaDiagnostico),
/* Foreign keys */
CONSTRAINT alumnomateriadiagnostico_ibfk_1
FOREIGN KEY (idAlumno)
REFERENCES alumno(idAlumno)
ON DELETE NO ACTION
ON UPDATE NO ACTION,
CONSTRAINT alumnomateriadiagnostico_ibfk_2
FOREIGN KEY (idMateriaDiagnostico)
REFERENCES materiadiagnostico(idMateriaDiagnostico)
ON DELETE NO ACTION
ON UPDATE NO ACTION
);

CREATE INDEX Alumno_has_MateriaDiagnostico_FKIndex1
ON alumnomateriadiagnostico

```

```
(idAlumno);

CREATE INDEX Alumno_has_MateriaDiagnostico_FKIndex2
ON alumnomateriadiagnostico
(idMateriaDiagnostico);
```

Tabla "alumnopromedio"

Campo	Tipo de Dato	PK	FK	No nulo	Por defecto	Descripción
idSemestre	int	✓	✓	✓		Identificador del semestre
idAlumno	int	✓	✓	✓		Identificador del alumno
promedio	int					Promedio del alumno en el semestre

Definición:

```
CREATE TABLE alumnopromedio (
  idSemestre int NOT NULL,--COMMENT 'Identificador del semestre',
  idAlumno int NOT NULL ,--COMMENT 'Identificador del alumno\r\n',
  promedio int ,--COMMENT 'Promedio del alumno en el semestre',
  /* Keys */
  PRIMARY KEY (idSemestre, idAlumno),
  /* Foreign keys */
  CONSTRAINT alumnopromedio_ibfk_1
  FOREIGN KEY (idAlumno)
  REFERENCES alumno(idAlumno)
  ON DELETE NO ACTION
  ON UPDATE NO ACTION,
  CONSTRAINT alumnopromedio_ibfk_2
  FOREIGN KEY (idSemestre)
  REFERENCES semestre(idSemestre)
  ON DELETE NO ACTION
  ON UPDATE NO ACTION
);
```

```
CREATE INDEX Semestre_has_Alumno_FKIndex1
ON alumnopromedio
(idSemestre);
```

```
CREATE INDEX Semestre_has_Alumno_FKIndex2
ON alumnopromedio
(idAlumno);
```

Tabla "alumnotutor"

Campo	Tipo de Dato	PK	FK	No nulo	Por defecto	Descripción
idAlumno	int	✓	✓	✓		Identificador del alumno
idTutor	int	✓	✓	✓		Identificador del tutor
secuencia	in				1	[1] = Primer tutor [2] = Segundo tutor [3] =

Definición:

```
CREATE TABLE alumnotutor (
  idAlumno  int  NOT NULL ,--COMMENT 'Identificador del alumno',
  idTutor   int  NOT NULL ,--COMMENT 'Identificador del tutor\r\n',
  secuencia int  DEFAULT '1' ,--COMMENT '[1] = Primer tutor [2] = Segundo tutor [3] = .....',
  /* Keys */
  PRIMARY KEY (idAlumno, idTutor),
  /* Foreign keys */
  CONSTRAINT alumnotutor_ibfk_1
    FOREIGN KEY (idTutor)
    REFERENCES tutor(idTutor)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION,
  CONSTRAINT alumnotutor_ibfk_2
    FOREIGN KEY (idAlumno)
    REFERENCES alumno(idAlumno)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION
);

CREATE INDEX Tutor_has_Alumno_FKIndex1
  ON alumnotutor
  (idTutor);

CREATE INDEX Tutor_has_Alumno_FKIndex2
  ON alumnotutor
  (idAlumno);
```

Tabla "carrera"

Campo	Tipo de Dato	PK	FK	No nulo	Por defecto	Descripción
idCarrera	int	✓		✓		Identificador de la carrera
nombre	varchar(20)			✓		Nombre de la carrera
clave	int			✓		Clave de la carrera
siglasCarr	char(3)			✓		Siglas de la carrera
siglasDiv	varchar(10)			✓		Siglas de su división
origen	varchar(10)			✓		Origen
creditos	int					Créditos que tiene la carrera

Definición:

```
CREATE TABLE carrera (
  idCarrera  int  IDENTITY(1,1) NOT NULL,-- COMMENT 'Identificador de la carrera\r\n',
  nombre     varchar(20) NOT NULL,-- COMMENT 'Nombre de la carrera\r\n',
  clave      int  NOT NULL,-- COMMENT 'Clave de la carrera',
  siglasCarr char(3) NOT NULL, -- COMMENT 'Siglas de la carrera',
  siglasDiv  varchar(10) NOT NULL,-- COMMENT 'Siglas de su division',
  origen     varchar(10) NOT NULL,-- COMMENT 'Origen',
```

```

    creditos    int -- COMMENT 'Creditos que tiene la carrera',
    /* Keys */
    PRIMARY KEY (idCarrera)
)

```

Tabla "delegacion"

Campo	Tipo de Dato	PK	FK	No nulo	Por defecto	Descripción
idDelegacion	int	✓		✓		Identificador de la delegación
idEstado	int		✓	✓		Identificador del estado
nombre	varchar(20)			✓		Nombre de la delegación

Definición:

```

CREATE TABLE delegacion (
    idDelegacion int IDENTITY(1,1) NOT NULL,-- COMMENT 'Identificador de la delegación\r\n',
    idEstado    int NOT NULL,-- COMMENT 'Identificador del estado',
    nombre      varchar(20) NOT NULL,-- COMMENT 'Nombre de la delegación',
    /* Keys */
    PRIMARY KEY (idDelegacion),
    /* Foreign keys */
    CONSTRAINT delegacion_ibfk_1
        FOREIGN KEY (idEstado)
        REFERENCES estado(idEstado)
        ON DELETE NO ACTION
        ON UPDATE NO ACTION
);

CREATE INDEX Delegacion_FKIndex1
ON delegacion
(idEstado);

```

Tabla "encuesta"

Campo	Tipo Dato	de	PK	FK	Not Null	Default	Descripción
idEncuesta	int		✓		✓		Identificador de la encuesta
nombre	varchar(20)				✓		Nombre de la encuesta

Definición:

```

CREATE TABLE encuesta (
    idEncuesta int IDENTITY(1,1) NOT NULL,-- COMMENT 'Identificador de la encuesta',
    nombre     varchar(20) NOT NULL,-- COMMENT 'Nombre de la encuesta',
    /* Keys */
    PRIMARY KEY (idEncuesta)
)

```

Tabla "escuela"

Campo	Tipo de Dato	PK	FK	No nulo	Por defecto	Descripción
idEscuela	int(✓		✓		Identificador de la escuela
nombre	varchar(60)			✓		Nombre de la escuela
nivel	int			✓		[1] = Primaria [2] = Secundaria [3] = Bachillerato

Definición:

```
CREATE TABLE escuela (
  idEscuela int IDENTITY(1,1) NOT NULL,-- COMMENT 'Identificador de la escuela',
  nombre varchar(60) NOT NULL,--COMMENT 'Nombre de la escuela',
  nivel int NOT NULL,-- COMMENT '[1] = Primaria [2] = Secundaria [3] = Bachillerato',
  /* Keys */
  PRIMARY KEY (idEscuela)
);
```

Tabla "estado"

Campo	Tipo de Dato	PK	FK	No nulo	Por defecto	Descripción
idEstado	int	✓		✓		Identificador del Estado
nombre	varchar(20)			✓		Nombre del estado

Definición:

```
CREATE TABLE estado (
  idEstado int IDENTITY(1,1) NOT NULL,-- COMMENT 'Identificador del Estado\r\n',
  nombre varchar(20) NOT NULL,-- COMMENT 'Nombre del estado',
  /* Keys */
  PRIMARY KEY (idEstado)
);
```

Tabla "incisos"

Campo	Tipo de Dato	PK	FK	No nulo	Por defecto	Descripción
idIncisos	int	✓		✓		Identificador del inciso
opcion	varchar(100)			✓		Texto del inciso

Definición:

```
CREATE TABLE incisos (
  idIncisos int IDENTITY(1,1) NOT NULL,-- COMMENT 'Identificador del inciso',
  opcion varchar(100) NOT NULL,-- COMMENT 'Texto del inciso',
  /* Keys */
  PRIMARY KEY (idIncisos)
);
```

Tabla "materia"

Campo	Tipo de Dato	PK	FK	No nulo	Por defecto	Descripción
idMateria	int	✓		✓		Identificador de la materia
nombre	varchar(40)					Nombre de la materia
creditos	int					Créditos de la materia
tipo	char					Tipo de materia

Definición:

```
CREATE TABLE materia (
  idMateria int IDENTITY(1,1) NOT NULL, -- COMMENT 'Identificador de la materia'
  nombre varchar(40) , -- COMMENT 'Nombre de la materia'
  creditos int, -- COMMENT 'Creditos de la materia'
  tipo char , -- COMMENT 'Tipo de materia'
  /* Keys */
  PRIMARY KEY (idMateria)
);
```

Tabla "materiadiagnostico"

Campo	Tipo de Dato	PK	FK	No nulo	Por defecto	Descripción
idMateriaDiagnostico	int	✓		✓		Identificador de la materia de diagnostico
nombre	varchar(20)			✓		Nombre de la materia de diagnostico
puntos	float			✓		Puntos de la materia

Definición:

```
CREATE TABLE materiadiagnostico (
  idMateriaDiagnostico int IDENTITY(1,1) NOT NULL,-- COMMENT 'Identificador de la materia de diagnostico',
  nombre varchar(20) NOT NULL,-- COMMENT 'Nombre de la materia de diagnostico',
  puntos float NOT NULL,-- COMMENT 'Puntos de la materia',
  /* Keys */
  PRIMARY KEY (idMateriaDiagnostico)
);
```

Tabla "pregunta"

Campo	Tipo de Dato	PK	FK	No nulo	Por defecto	Descripción
idPregunta	int	✓		✓		Identificador de la pregunta
idEncuesta	int		✓	✓		Identificador de la encuesta a la que pertenece
nombre	varchar(140)			✓		Texto de la pregunta

Definición:

```
CREATE TABLE pregunta (
  idPregunta int IDENTITY(1,1) NOT NULL,-- COMMENT 'Identificador de la pregunta',
  idEncuesta int NOT NULL,-- COMMENT 'Identificador de la encuesta a la que pertenece',
  nombre varchar(140) NOT NULL,-- COMMENT 'Texto de la pregunta',
  /* Keys */
  PRIMARY KEY (idPregunta),
  /* Foreign keys */
  CONSTRAINT pregunta_ibfk_1
  FOREIGN KEY (idEncuesta)
  REFERENCES encuesta(idEncuesta)
  ON DELETE NO ACTION
  ON UPDATE NO ACTION
);

CREATE INDEX Pregunta_FKIndex1
ON pregunta
(idEncuesta);
```

Tabla "preguntaincisos"

Campo	Tipo de Dato	PK	FK	No nulo	Por defecto	Descripción
idPregunta	int	✓	✓	✓		Identificador de la pregunta
idIncisos	int	✓	✓	✓		Identificador del inciso que tendrá la pregunta

Definición:

```
CREATE TABLE preguntaincisos (
  idPregunta int IDENTITY(1,1) NOT NULL,-- COMMENT 'Identificador de la pregunta',
  idIncisos int NOT NULL,-- COMMENT 'Identificador del inciso que tendra la pregunta',
  /* Keys */
  PRIMARY KEY (idPregunta, idIncisos),
  /* Foreign keys */
  CONSTRAINT preguntaincisos_ibfk_1
  FOREIGN KEY (idPregunta)
  REFERENCES pregunta(idPregunta)
  ON DELETE NO ACTION
  ON UPDATE NO ACTION,
  CONSTRAINT preguntaincisos_ibfk_2
  FOREIGN KEY (idIncisos)
```

```
REFERENCES incisos(idIncisos)
ON DELETE NO ACTION
ON UPDATE NO ACTION);
CREATE INDEX Pregunta_has_Incisos_FKIndex1
ON preguntaincisos
(idPregunta);

CREATE INDEX Pregunta_has_Incisos_FKIndex2
ON preguntaincisos
(idIncisos);
```

Tabla "rango"

Campo	Tipo de Dato	PK	FK	No nulo	Por defecto	Descripción
idRango	int	✓		✓		Identificador del rango
descripcion	varchar(20)			✓		Rango de promedio de calificación con que salió el alumno de esa escuela

Definición:

```
CREATE TABLE rango (
  idRango int IDENTITY(1,1) NOT NULL,-- COMMENT 'Identificador del rango',
  descripcion varchar(20) NOT NULL,-- COMMENT 'Rango de promedio de calificacion con que
salió el alumno de esa escuela',
/* Keys */
PRIMARY KEY (idRango));
```

Tabla "resultados"

Campo	Tipo de Dato	PK	FK	No nulo	Por defecto	Descripción
idAlumno	int	✓	✓	✓		Identificador del alumno
idIncisos	int	✓	✓	✓		Identificador del inciso
idPregunta	int	✓	✓	✓		Identificador de la pregunta

Definición:

```
CREATE TABLE resultados (
  idAlumno int NOT NULL,--COMMENT 'Identificador del alumno',
  idIncisos int NOT NULL,--COMMENT 'Identificador del inciso',
  idPregunta int NOT NULL,--COMMENT 'Identificador de la pregunta',
/* Keys */
PRIMARY KEY (idAlumno, idIncisos, idPregunta),
/* Foreign keys */
CONSTRAINT resultados_ibfk_1
FOREIGN KEY (idPregunta, idIncisos)
REFERENCES preguntaincisos(idPregunta, idIncisos)
ON DELETE NO ACTION
```

```

ON UPDATE NO ACTION,
CONSTRAINT resultados_ibfk_2
FOREIGN KEY (idAlumno)
REFERENCES alumno(idAlumno)
ON DELETE NO ACTION
ON UPDATE NO ACTION);

CREATE INDEX PreguntaIncisos_has_Alumno_FKIndex1
ON resultados
(idPregunta, idIncisos);

CREATE INDEX PreguntaIncisos_has_Alumno_FKIndex2
ON resultados
(idAlumno);

```

Tabla "semestre"

Campo	Tipo de Dato	PK	FK	No nulo	Por defecto	Descripción
idSemestre	int	✓		✓		Identificador del semestre
descripcion	varchar(20)					Semestre en cuestión

Definition:

```

CREATE TABLE semestre (
  idSemestre int IDENTITY(1,1) NOT NULL,-- COMMENT 'Identificador del semestre\r\n',
  descripcion varchar(20),-- COMMENT 'Semestre en cuestión',
  /* Keys */
  PRIMARY KEY (idSemestre)
);

```

Tabla "tiempo"

Campo	Tipo de Dato	PK	FK	No nulo	Por defecto	Descripción
idTiempo	int	✓		✓		Identificador de la tabla
nombre	varchar(45)					Años cursados

Definición:

```

CREATE TABLE tiempo (
  idTiempo int IDENTITY(1,1) NOT NULL,-- COMMENT 'Identificador de la tabla\r\n',
  nombre varchar(45),-- COMMENT 'Años cursados',
  /* Keys */
  PRIMARY KEY (idTiempo)
);

```

Tabla "tutor"

Campo	Tipo de Dato	PK	FK	No nulo	Por defecto	Descripción
idTutor	int	✓		✓		Identificador del tutor
Nombre	varchar(20)					Nombre del tutor

Definición:

```
CREATE TABLE tutor (
  idTutor int IDENTITY(1,1) NOT NULL, --COMMENT 'Identificador del tutor',
  Nombre varchar(20), -- COMMENT 'Nombre del tutor',
  /* Keys */
  PRIMARY KEY (idTutor)
);
```

3.4 Herramientas y técnicas de análisis de datos y evaluación de las mismas



Microsoft SQL Server proporciona varios algoritmos que se pueden usar en las soluciones de minería de datos. Estos algoritmos son un subconjunto de todos los algoritmos que pueden utilizarse en la minería de datos. También permite utilizar algoritmos de minería de datos desarrollados por terceros que cumplan la especificación OLE DB para minería de datos. [21]

El siguiente diagrama describe las relaciones entre cada paso del proceso y las tecnologías de Microsoft SQL Server que se pueden utilizar para completar cada paso.

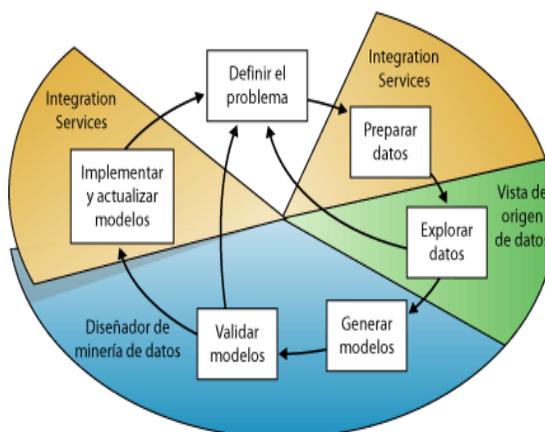


FIGURA 3.4.a Diagrama de procesos SQL Server

Aunque el proceso que se ilustra en el diagrama es circular, esto no significa que cada paso conduzca directamente al siguiente. La creación de un modelo de minería de datos es un proceso dinámico e iterativo. Una vez que ha explorado los datos, puede que descubra que resultan insuficientes para crear los modelos de minería

de datos adecuados y que, por tanto, debe buscar más datos. Pueden ser generados varios modelos y descubrir que no responden al problema planteado cuando fue definido y que, por tanto, se debe volver a definir el problema. Es posible que en ocasiones se tenga la necesidad de actualizar los modelos una vez implementados debido a que haya más datos disponibles.

SQL Server ofrece así un entorno integrado para crear y trabajar con modelos de minería de datos denominado *Business Intelligence Development Studio*. El entorno incluye algoritmos y herramientas de minería de datos que facilitan la generación de una solución completa para diversos proyectos. [21]

Ventajas:

- Permite a los usuarios más avanzados optimizar los modelos al cambiar valores de parámetros de algoritmos.
- Incluye un juego completo de herramientas de visualización.
- Permite graficar la relativa exactitud de todos sus modelos para columnas de predicción.

La siguiente tabla nos proporciona sugerencias sobre qué algoritmos usar en tareas específicas dentro de SQL Server 2008:

Tarea	Algoritmos de Microsoft que se pueden usar
Predecir un atributo discreto. Por ejemplo, predecir si el destinatario de una campaña de correo directo adquirirá un producto.	Algoritmo de árboles de decisión de Microsoft Algoritmo Naive Bayes de Microsoft Algoritmo de clústeres de Microsoft Algoritmo de red neuronal de Microsoft
Predecir un atributo continuo. Por ejemplo, prever las ventas del año próximo.	Algoritmo de árboles de decisión de Microsoft Algoritmo de serie temporal de Microsoft
Predecir una secuencia. Por ejemplo, realizar un análisis del flujo de clics en el sitio Web de una empresa.	Algoritmo de clústeres de secuencia de Microsoft
Buscar grupos de elementos comunes en transacciones. Por ejemplo, utilizar el análisis de la cesta de compra para sugerir a un cliente la compra de productos adicionales.	Algoritmo de asociación de Microsoft Algoritmo de árboles de decisión de Microsoft
Buscar grupos de elementos similares. Por ejemplo, segmentar datos demográficos en grupos a fin de comprender mejor las relaciones entre los atributos.	Algoritmo de clústeres de Microsoft Algoritmo de clústeres de secuencia de Microsoft

FIGURA 3.4.b Tabla de algoritmos [21]

Microsoft SQL Server Analysis Services

Microsoft SQL Server Analysis Services contiene las características y herramientas necesarias para crear complejas soluciones de minería de datos:

- Un conjunto de algoritmos de minería de datos estándar del sector.
- El diseñador de minería de datos sirve para crear, administrar y examinar modelos de minería de datos para, a continuación, crear predicciones a partir de dichos modelos.
- El lenguaje DMX (Extensiones de minería de datos), que sirve para administrar modelos de minería de datos y crear complejas consultas predictivas.

Se pueden usar varias de estas características y herramientas a la vez para detectar las tendencias y los patrones

existentes en los datos; después, se pueden usar esas tendencias y los patrones para tomar decisiones informadas sobre los problemas que se tengan que resolver. [22]

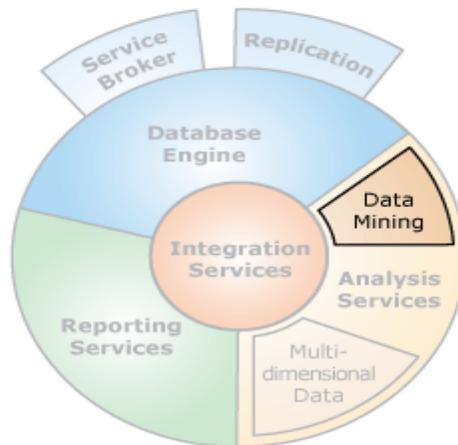


FIGURA 3.4.c *Integration Services* [22]

La imagen anterior muestra los distintos componentes que conforman a *Integration Services*, uno de estos componentes es el *Analysis Services*, el cual contiene las herramientas de minería de datos antes mencionadas.

Business Intelligence Development Studio

SQL Server Business Intelligence Development Studio es un entorno integrado para desarrollar construcciones de inteligencia empresarial, como orígenes de datos, cubos, informes y paquetes de *Integration Services*. *Business Intelligence Development Studio* incluye plantillas de proyecto que proporcionan un contexto para desarrollar construcciones específicas. También es posible desarrollar proyectos que formen parte de una solución independiente de un servidor concreto. [23]

En *Business Intelligence Development Studio*, el asistente para minería de datos facilita la creación de estructuras y de modelos de minería de datos basados en orígenes de datos OLAP y relacionales. Uno puede utilizar el asistente para definir estructuras y modelos que utilicen técnicas de minería de datos específicas para analizar datos. Así mismo es posible utilizar el diseñador de minería de datos para perfeccionar la definición de modelos de minería de datos y explorar y trabajar con los resultados del modelo. [23]

SQL Server Management Studio

SQL Server Management Studio proporciona herramientas que pueden ser utilizadas para administrar y explorar los modelos de minería de datos una vez creados. [24]

Transformaciones y tareas de minería de datos en Integration Services

SQL Server Integration Services proporciona herramientas que pueden ser utilizadas para automatizar tareas comunes de minería de datos, como procesar un modelo de minería de datos y crear consultas de predicción. Por ejemplo, si se dispone de un modelo de minería de datos generado a partir de un conjunto de datos de posibles clientes, se puede crear un paquete de *Integration Services* que actualice automáticamente el modelo cada vez que el conjunto de datos se actualice con nuevos clientes. A continuación se podría utilizar el paquete para crear una predicción, separando los clientes potenciales en dos tablas. Una tabla contendría los clientes probables y la otra los clientes que posiblemente no adquirirán ningún producto. [24]

Capítulo 4. Análisis de datos y presentación de resultados

4.1 Métodos de minería de datos

Como se mencionó anteriormente un método de minería de datos es el mecanismo que crea modelos de minería de datos. En este apartado y previo a crear los modelos de datos es importante hacer mención de los métodos y algoritmos que habrán de usarse posteriormente.

4.1.1 Árboles de decisión (Clasificación y Regresión)

El método de árboles de decisión utiliza algoritmos de clasificación y regresión proporcionados por MicrosoftSQL ServerAnalysis Services para el modelado de predicción de atributos discretos y continuos. [25]

Para los atributos discretos, el método hace predicciones basándose en las relaciones entre las columnas de entrada de un conjunto de datos. Utiliza los valores, conocidos como estados, de estas columnas para predecir los estados de una columna que se designa como elemento de predicción. Específicamente, identifica las columnas de entrada que se correlacionan con la columna de predicción. Por ejemplo, en un escenario para predecir qué clientes van a adquirir probablemente una bicicleta, si nueve de diez clientes jóvenes compran una bicicleta, pero sólo lo hacen dos de diez clientes de edad mayor, el algoritmo infiere que la edad es un buen elemento de predicción en la compra de bicicletas. El árbol de decisión realiza predicciones basándose en la tendencia hacia un resultado concreto. [25]

Para los atributos continuos, se usa el algoritmo de regresión lineal para determinar dónde se divide un árbol de decisión. [25]

Si se define más de una columna como elemento de predicción, o si los datos de entrada contienen una tabla anidada que se haya establecido como elemento de predicción, se genera un árbol de decisión independiente para cada columna de predicción. [25]

El método de árboles de decisión de Microsoft genera un modelo de minería de datos mediante la creación de una serie de divisiones en el árbol. Estas divisiones se representan como *nodos*. Agrega un nodo al modelo cada vez que una columna de entrada tiene una correlación significativa con la columna de predicción. La forma en que se determina una división varía en función de si predice una columna continua o una columna discreta. El método de árboles de decisión de Microsoft utiliza la selección de características para guiar la selección de los atributos más útiles. [25]

Predecir columnas discretas

La forma en que el método de árboles de decisión genera un árbol para una columna de predicción discreta puede mostrarse mediante un histograma. El siguiente diagrama muestra un histograma que traza una columna de predicción, Bike Buyers, con una columna de entrada, Age. El histograma muestra que la edad de una persona ayuda a distinguir si esa persona comprará una bicicleta. [25]

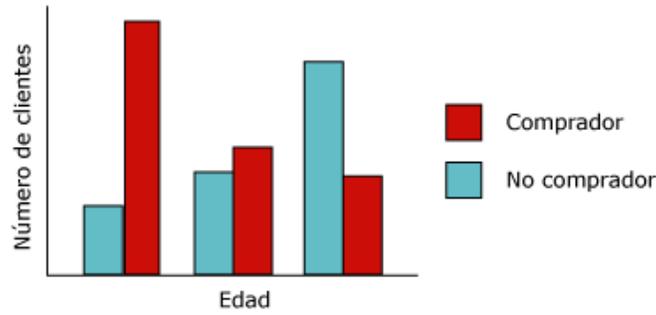


Figura 4.1.1.a Histograma

La correlación que aparece en el diagrama hará que el método de árboles de decisión cree un nuevo nodo en el modelo.

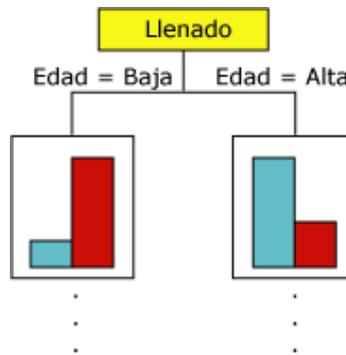


Figura 4.1.1.b Árbol basado en histograma

A medida en que se agregan nuevos nodos a un modelo, se forma una estructura en árbol. El nodo superior del árbol describe el desglose de la columna de predicción para la población global de clientes. A medida que el modelo crece, se consideran todas las columnas. [25]

Predecir columnas continuas

Cuando el método de árboles de decisión genera un árbol basándose en una columna de predicción continua, cada nodo contiene una fórmula de regresión. Se produce una división en un punto de no linealidad de la fórmula de regresión. Por ejemplo, considérese el siguiente diagrama.

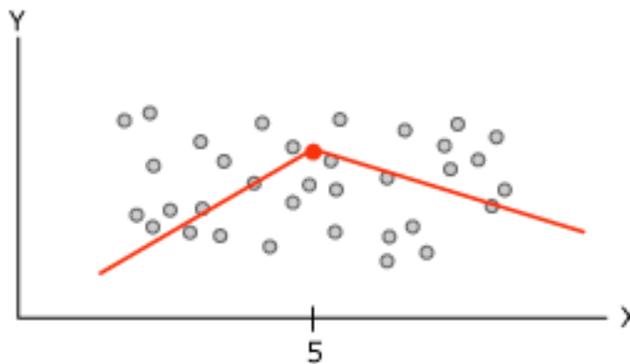


Figura 4.1.1.c Diagrama de datos

El diagrama contiene los datos que pueden modelarse utilizando una sola línea o dos líneas conectadas. Sin embargo, una sola línea realizará un pobre trabajo en la representación de los datos. En su lugar, si se usan dos líneas, el modelo hará un mejor trabajo en la aproximación a los datos. El punto donde las dos líneas se unen es el punto de no linealidad y donde se dividiría un nodo de un modelo de árbol de decisión. Por ejemplo, el nodo que corresponde al punto de no linealidad del gráfico anterior podría representarse mediante el siguiente diagrama. Las dos ecuaciones representan las ecuaciones de regresión de las dos líneas. [25]

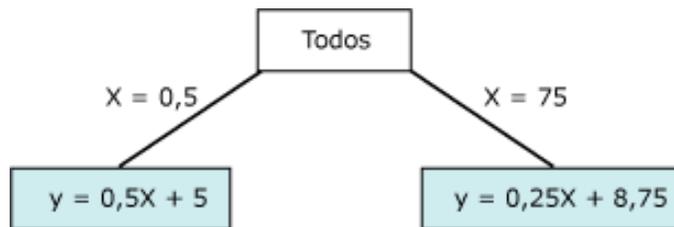


Figura 4.1.1.d Árbol basado en diagrama

Selección de características

Hay muchas maneras de implementar la selección de características, dependiendo del tipo de datos con los que se esté trabajando. SQL Server Analysis Services proporciona varias técnicas conocidas y consolidadas para puntuar los atributos. Aquella que se aplique en los conjuntos de datos depende de los tipos de datos, así como del uso de las columnas.

Puntuación de grado de interés

Una característica es interesante si ofrece información útil.

La medida de la cualidad del grado de interés que se utiliza en SQL Server Analysis Services está basada en la entropía, lo que significa que los atributos con distribuciones aleatorias tienen una entropía más alta y obtienen menos información; por tanto, esos atributos son menos interesantes. La entropía para cualquier atributo se compara con la entropía de todos los demás atributos de la manera siguiente:

$$\text{Interestingness}(\text{Atributo}) = - (m - \text{Entropy}(\text{Atributo})) * (m - \text{Entropy}(\text{Atributo}))$$

La entropía central (m) es la entropía de todo el conjunto de características. Al restar la entropía del atributo de destino de la entropía central, se puede evaluar cuánta información proporciona el atributo.

Esta puntuación se utiliza de forma predeterminada cada vez que la columna contiene datos numéricos continuos no binarios. [26]

Bayesiano con prioridad K2

Analysis Services proporciona dos puntuaciones de selección de características basadas en las redes bayesianas. Una red bayesiana es un gráfico *dirigido* o acíclico de estados y de transiciones entre ellos; esto significa que algunos estados siempre son anteriores al estado actual y otros son posteriores, y que el gráfico no se repite ni realiza bucles. Por definición, las redes bayesianas permiten el uso del conocimiento previo. Sin embargo, la pregunta sobre qué estados anteriores se deben utilizar para calcular las probabilidades de los estados posteriores es importante para la precisión, el rendimiento y el diseño del algoritmo.

Esta técnica de puntuación está disponible para los atributos discretos y discretizados. [26]

4.1.2 Clústeres (Agrupamiento)

Es un método que utiliza técnicas iterativas para agrupar los escenarios de un conjunto de datos dentro de clústeres que contienen características similares. Estas agrupaciones son útiles para la exploración de datos, la identificación de anomalías en los datos y la creación de predicciones. [27]

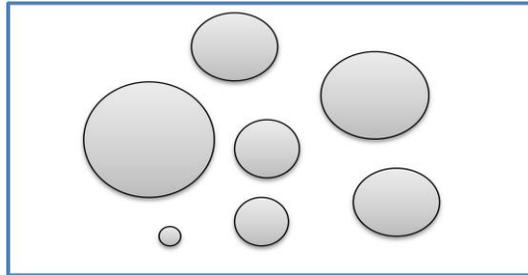


Figura 4.1.2.a Grupos

Los modelos de clústeres identifican las relaciones en un conjunto de datos que no se podrían derivar lógicamente a través de la observación casual. Primero se identifican relaciones de un conjunto de datos y genera una serie de clústeres basándose en ellas. Una forma de verlo es mediante un gráfico de dispersión, el cual representa todos los escenarios del conjunto de datos; cada escenario es un punto del gráfico. Los clústeres agrupan los puntos del gráfico e ilustran las relaciones que identifica el algoritmo. [27]

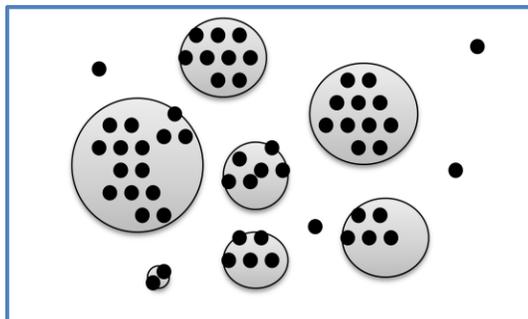


Figura 4.1.2.b Clústeres

Después de definir los clústeres, el algoritmo calcula el grado de perfección con que los clústeres representan las agrupaciones de puntos y, a continuación, intenta volver a definir las agrupaciones para crear clústeres que representen mejor los datos. El algoritmo establece una iteración en este proceso hasta que ya no es posible mejorar los resultados mediante la redefinición de los clústeres. [27]

El método de clústeres de Microsoft proporciona dos algoritmos para crear clústeres y asignar puntos de datos a dichos clústeres. El primero, el algoritmo K-means, que es un algoritmo de agrupación en clústeres duro. Esto significa que un punto de datos puede pertenecer a un solo clúster, y que únicamente se calcula una probabilidad de pertenencia de cada punto de datos de ese clúster. El segundo, el algoritmo de Expectation Maximization (EM), que es un algoritmo de agrupación en clústeres blando. Esto significa que un punto de datos siempre pertenece a varios clústeres, y que se calcula una probabilidad para cada combinación de punto de datos y clúster. [28]

Agrupación en clústeres EM

En el algoritmo de agrupación en clústeres EM, el algoritmo refina de forma iterativa un modelo de clústeres inicial para ajustar los datos y determina la probabilidad de que un punto de datos exista en un clúster. El algoritmo finaliza el proceso cuando el modelo probabilístico ajusta los datos. La función usada para determinar el ajuste es el logaritmo de la probabilidad de los datos dado el modelo. [28]

Si durante el proceso se generan clústeres vacíos, o si la pertenencia de uno o varios de los clústeres cae por debajo del umbral especificado, los clústeres con poblaciones bajas se reinician en los nuevos puntos y vuelve a ejecutarse el algoritmo EM. [28]

Los resultados del proceso son probabilísticos. Esto significa que cada punto de datos pertenece a todos los clústeres, pero cada asignación de un punto de datos a un clúster tiene una probabilidad diferente. Dado que el método permite que los clústeres se superpongan, la suma de los elementos de todos los clústeres puede superar la totalidad de los elementos existentes en el conjunto de entrenamiento. En los resultados del modelo de minería de datos, las puntuaciones que indican soporte se ajustan para tener en cuenta este hecho. [28]

El algoritmo EM es el algoritmo predeterminado usado en los modelos de agrupación en clústeres de Microsoft. Este algoritmo se usa como algoritmo predeterminado porque proporciona numerosas ventajas comparado con la agrupación en clústeres K-means:

- Requiere examinar la base de datos como máximo una vez.
- Funciona incluso si la cantidad de memoria (RAM) es limitada.
- Tiene la capacidad de usar un cursor de solo avance.
- Sus resultados superan los obtenidos por los métodos de muestreo.

La implementación de Microsoft proporciona dos opciones: EM escalable y no escalable. De forma predeterminada, en EM escalable, los primeros 50.000 registros se usan para inicializar el examen inicial. Si esta operación se realiza correctamente, el modelo solo usa estos datos. Si el modelo no se puede ajustar con 50.000 registros, se leen otros 50.000. En EM no escalable, se lee el conjunto de datos completo independientemente de su tamaño. Este método puede crear clústeres más precisos, pero los requisitos de memoria pueden ser significativos. Dado que EM escalable funciona en un búfer local, recorrer los datos en iteración es mucho más rápido, y el algoritmo hace un mejor uso de la caché de memoria de la CPU que EM no escalable. Es más, EM escalable es tres veces más rápido que EM no escalable, incluso si todos los datos caben en la memoria principal. En la mayoría de casos, la mejora en el rendimiento no significa una reducción de la calidad del modelo completo. [28]

Agrupación en clústeres K-means

La agrupación en clústeres K-means es un algoritmo muy conocido para asignar la pertenencia al clúster que consiste en minimizar las diferencias entre los elementos de un clúster al tiempo que se maximiza la distancia entre los clústeres. El término "mediana" hace referencia al *centroide* del clúster, que es un punto de datos que se elige arbitrariamente y que se refina de forma iterativa hasta que representa la verdadera media de todos los puntos de datos del clúster. La "K" hace referencia a un número arbitrario de puntos que se usan para inicializar el proceso de agrupación en clústeres. El algoritmo K-means calcula las distancias euclidianas cuadradas entre los registros de datos de un clúster y el vector que representa la media de clústeres, y converge en un conjunto final

de K clústeres cuando la suma alcanza su valor mínimo. [28]

El algoritmo K-means asigna cada punto de datos a un solo clúster y no permite la incertidumbre en la pertenencia. En un clúster, la pertenencia se expresa como una distancia desde el centroide. [28]

El algoritmo mediana-K proporciona dos métodos para realizar un muestreo en el conjunto de datos: mediana-K no escalable, que carga el conjunto de datos completo y realiza una pasada de agrupación en clústeres, y mediana-K escalable, donde el algoritmo usa los primeros 50.000 casos y lee más casos únicamente si necesita más datos para lograr un buen ajuste del modelo a los datos. [28]

4.1.3 Reglas de asociación y dependencia (Agrupamiento, Reglas de asociación)

Los modelos de asociación se generan basándose en conjuntos de datos que contienen identificadores para escenarios individuales y para los elementos que contienen los escenarios. Un grupo de elementos de un escenario se denomina un *conjunto de elementos*. Un modelo de asociación se compone de una serie de conjuntos de elementos y de las reglas que describen cómo estos elementos se agrupan dentro de los escenarios. [29]

El algoritmo Apriori de asociación puede encontrar potencialmente muchas reglas dentro de un conjunto de datos. El algoritmo usa dos parámetros, compatibilidad y probabilidad, para describir los conjuntos de elementos y las reglas que genera. [29]

El algoritmo de asociación recorre un conjunto de datos para hallar elementos que aparezcan juntos en un escenario. A continuación, agrupa en conjuntos de elementos todos los elementos asociados que aparecen. Por ejemplo, un conjunto de elementos podría ser "Mountain 200=Existing, Sport 100=Existing" y podría tener una compatibilidad de 710. A continuación, el algoritmo genera las reglas a partir de los conjuntos de elementos. Estas reglas se usan para predecir la presencia de un elemento en la base de datos, basándose en la presencia de otros elementos específicos que el algoritmo ha identificado como importantes. Por ejemplo, una regla puede ser "if Touring 1000=existing and Road bottle cage=existing, then Water bottle=existing", y puede tener una probabilidad de 0.812. En este ejemplo, el algoritmo identifica que la presencia en la cesta del neumático Touring 1000 y del soporte de la botella de agua predice que probablemente la cesta de compra incluirá también una botella de agua. [29]

El algoritmo Apriori no analiza patrones, sino que genera y después cuenta conjuntos de elementos candidatos. Un elemento puede representar un evento, un producto o el valor de un atributo, dependiendo del tipo de datos que se analice. [30]

En el tipo más común de modelo de asociación, las variables booleanas, que representan un valor Sí/No o Falta/Existe, se asignan a cada atributo, como un nombre de producto o evento. Un análisis de cesta de la compra es un ejemplo de un modelo de reglas de asociación que utiliza variables booleanas para representar la presencia o ausencia de determinados productos en la cesta de la compra de un cliente. [30]

Para cada conjunto de elementos, el algoritmo crea puntuaciones que representan el soporte y la confianza. Estas puntuaciones se pueden usar para clasificar y derivar reglas interesantes de los conjuntos de elementos. [30]

Los modelos de asociación se pueden crear también para atributos numéricos. Si los atributos son continuos, los números se pueden discretizar o agruparse en depósitos. A continuación, los valores discretizados se pueden tratar como booleanos o como pares atributo-valor. [30]

4.1.4 Naive Bayes (Clasificación)

El método Naive Bayes es un método de clasificación para el modelado de predicción. Este calcula la probabilidad condicional entre columnas de entrada y de predicción y supone que las columnas son independientes. Esta suposición de independencia implica, de manera un tanto ingenua en ocasiones, que este método no tiene en cuenta las dependencias que puedan existir. [31]

Desde el punto de vista computacional, su algoritmo es menos complejo que otros algoritmos y, por tanto, resulta útil para generar rápidamente modelos de minería de datos para descubrir relaciones entre columnas de entrada y columnas de predicción. Puede ser utilizado para realizar exploraciones iniciales de datos y, más adelante, aplicar los resultados para crear modelos de minería de datos adicionales con otros algoritmos más complejos y precisos desde el punto de vista computacional. [31]

Su algoritmo Naive Bayes calcula la probabilidad de cada estado de cada columna de entrada, dado cada posible estado de la columna de predicción. [31]

Atributos	Estados	Población... Tamaño: 18484	0 Tamaño: 9352	1 Tamaño: 9132	ausente Tamaño: 0
Age	<ul style="list-style-type: none"> ● 38 - 43 ● 29 - 34 ● 43 - 48 ● Other 				
Commute Distance	<ul style="list-style-type: none"> ● 0-1 Miles ● 2-5 Miles ● 1-2 Miles ● Other 				
Education	<ul style="list-style-type: none"> ● Bachelors ● Partial College ● High School ● Other 				
Marital Status	<ul style="list-style-type: none"> ● M ● S ● Missing 				
Number Cars Owned	<ul style="list-style-type: none"> ● 2 ● 1 ● 0 ● Other 				
Number Children At Home	<ul style="list-style-type: none"> ● 0 ● 1 ● 2 ● Other 				
Occupation	<ul style="list-style-type: none"> ● Professional ● Skilled Manual ● Management 				

Figura 4.1.4.a Resultados del modelo Naive Bayes

En la figura se muestran los resultados después de procesar Naive Bayes, la primer columna muestra el atributo, en la segunda se muestran los estados que tuvo dicho atributo, la tercera muestra el tamaño de la población analizada y las siguientes columnas muestra la probabilidad de que el atributo tome dicho valor con el atributo dado. [30]

4.1.5 Redes Neuronales (Clasificación, Regresión, Agrupamiento)

El método de red neuronal Microsoft crea modelos de minería de datos de clasificación y regresión mediante la generación de una red de neuronas de tipo perceptrón multicapa. Su algoritmo calcula las probabilidades para cada posible estado del atributo de entrada cuando se da cada estado del atributo de predicción. Posteriormente, puede utilizar estas probabilidades para predecir un resultado del atributo predicho basado en los atributos de entrada. [32]

El método de red neuronal es útil para analizar datos de entrada complejos, como los datos de un proceso comercial o de producción, o problemas empresariales para los que hay una cantidad importante de datos de entrenamiento disponibles pero en los que no es fácil derivar reglas mediante otros algoritmos.

Los casos sugeridos para utilizar la red neuronal son:

- Análisis de comercialización y promoción, como medir el éxito de una promoción por correo directo o una campaña publicitaria en la radio. [32]
- Predecir los movimientos de las acciones, la fluctuación de la moneda u otra información financiera con gran número de cambios a partir de los datos históricos. [32]
- Analizar los procesos industriales y de producción. [32]
- Cualquier modelo de predicción que analice relaciones complejas entre muchas entradas y relativamente pocas salidas. [32]

El método de red neuronal utiliza una red de tipo perceptrón multicapa, que también se denomina red de tipo regla delta de propagación hacia atrás, compuesta por tres capas de neuronas o perceptrones. Estas capas son una capa de entrada, una capa oculta opcional y una capa de salida. En una red de tipo perceptrón multicapa, cada neurona recibe una o más entradas y genera una o más salidas idénticas. Cada salida es una función no lineal simple de la suma de las entradas a la neurona. Las entradas sólo pasan de los nodos de la capa de entrada a los nodos de la capa oculta y, finalmente, a la capa de salida; no existe ninguna conexión entre neuronas de la misma capa. (Las entradas pasan de los nodos de la capa de entrada a los nodos de la capa de salida si no se ha incluido ninguna capa oculta.) [33]

Los modelos de minería de datos construidos con la red neuronal pueden contener varias redes, en función del número de columnas que se utilizan para la entrada y la predicción, o sólo para la predicción. El número de redes que contiene un único modelo de minería de datos depende del número de estados que contienen las columnas de entrada y las columnas de predicción que utiliza el modelo. [33]

Existen tres tipos de neuronas en una red neuronal creada con el algoritmo de red neuronal:

- Neuronas de entrada

Las neuronas de entrada proporcionan valores de atributo de entrada para el modelo de minería de datos. En el caso de los atributos de entrada discretos, las neuronas de entrada suelen representar un único estado del atributo de entrada, incluidos los valores que faltan. Por ejemplo, un atributo de entrada binario produce un nodo de entrada que describe un estado ausente o existente, que indica si existe un valor para ese atributo. Una columna booleana que se utiliza como atributo de entrada genera tres neuronas de entrada: una neurona para un valor verdadero, una neurona para un valor falso y una neurona para un estado ausente o existente. Un atributo de entrada discreto que tiene más de dos estados genera una neurona de entrada por cada estado y una neurona de entrada para un estado ausente o existente. Un atributo de entrada continuo genera dos neuronas de entrada: una neurona para un estado ausente o existente y una neurona para el valor del propio atributo continuo. Las neuronas de entrada proporcionan entradas para una o más neuronas ocultas. [33]

- Neuronas ocultas

Las neuronas ocultas reciben entradas de las neuronas de entrada y proporcionan salidas a las neuronas de salida. [33]

- Neuronas de salida

Las neuronas de salida representan valores de atributo de predicción para el modelo de minería de datos. En el caso de los atributos de entrada discretos, una neurona de salida suele representar un único estado de predicción para un atributo de predicción, incluidos los valores que faltan. Por ejemplo, un atributo de predicción binario produce un nodo de salida que describe un estado ausente o existente, que indica si existe un valor para ese atributo. Una columna booleana que se utiliza como atributo de predicción genera tres neuronas de salida: una neurona para un valor verdadero, una neurona para un valor falso y una neurona para un estado ausente o existente. Un atributo de predicción discreto que tiene más de dos estados genera una neurona de salida por cada estado y una neurona de salida para un estado ausente o existente. Las columnas de predicción continuas generan dos neuronas de salida: una neurona para un estado ausente o existente y una neurona para el valor de la propia columna continua. [33]

Una neurona recibe varias entradas: en las neuronas de entrada, una neurona recibe entradas de los datos originales; en las neuronas ocultas y de salida, una neurona recibe entradas de la salida de otras neuronas de la red neuronal. Las entradas establecen relaciones entre neuronas; estas relaciones sirven como ruta de análisis para un conjunto específico de escenarios. [33]

Cada entrada tiene un valor asignado denominado *peso*, que describe la relevancia o importancia de una determinada entrada en la neurona oculta o en la neurona de salida. Cuanto mayor es el peso asignado a una entrada, más relevante o importante es el valor de esa entrada para la neurona que la recibe cuando el algoritmo determina si dicha entrada clasifica correctamente un escenario específico. Los pesos también pueden ser negativos, lo cual implica que la entrada puede desactivar, en lugar de activar, una neurona específica. El valor de la entrada se multiplica por el peso de realce de la entrada de una neurona específica. (En el caso de pesos negativos, el valor de la entrada se multiplica por el valor de no realce del peso.) [33]

Por consiguiente, cada neurona tiene una función no lineal sencilla asignada denominada función de activación, que describe la relevancia o importancia de una neurona determinada para la capa de una red neuronal. [33]

Redes neuronales de entrenamiento

Existen varios pasos implicados en el entrenamiento de un modelo de minería de datos que utiliza el algoritmo de red neuronal. Estos pasos están muy influenciados por los valores que se especifican en los parámetros disponibles para el algoritmo. [33]

En primer lugar, el algoritmo evalúa y extrae los datos de entrenamiento del origen de datos. Un porcentaje de los datos de entrenamiento, denominado *datos de exclusión*, se reserva para medir la precisión de la estructura del modelo resultante. Durante el proceso de entrenamiento, el modelo se evalúa frente a los datos de exclusión después de cada iteración en los datos de entrenamiento. Cuando la precisión del modelo deja de aumentar, el proceso de entrenamiento se detiene. [33]

A continuación, el algoritmo determina el número y la complejidad de las redes que admite el modelo de minería de datos. Si el modelo contiene uno o más atributos que sólo se utilizan para la predicción, el algoritmo crea una

única red que representa todos estos atributos. Si el modelo contiene uno o más atributos que se utilizan para la entrada y la predicción, el proveedor de algoritmos construye una red para cada uno de estos atributos. [33]

En el caso de los atributos de entrada y de predicción que tienen valores discretos, cada neurona de entrada o de salida representa respectivamente un único estado. En el caso de los atributos de entrada y de predicción que tienen atributos continuos, cada neurona de entrada o de salida representa respectivamente el intervalo y la distribución de valores del atributo. [33]

A continuación, el algoritmo determina el número inicial de neuronas que se crearán para la capa oculta.

El proveedor de algoritmos evalúa iterativamente el peso de todas las entradas de la red simultáneamente, tomando el conjunto de datos de entrenamiento reservado anteriormente y comparando el valor real conocido de cada escenario de los datos de exclusión con la predicción de la red, en un proceso conocido como *aprendizaje por lotes*. Una vez que el algoritmo ha evaluado el conjunto completo de los datos de entrenamiento, revisa el valor predicho y real de cada neurona. El algoritmo calcula el grado de error, si lo hay, y ajusta los pesos asociados con las entradas de esa neurona, trabajando hacia atrás desde las neuronas de salida a las de entrada en un proceso conocido como *propagación hacia atrás*. A continuación, el algoritmo repite el proceso en todo el conjunto de datos de entrenamiento. Dado que el algoritmo puede admitir múltiples pesos y neuronas de salida, el algoritmo de gradiente conjugado se utiliza para guiar el proceso de entrenamiento en la asignación y evaluación de los pesos de las entradas. [33]

4.1.6 Algoritmo de regresión lineal

El algoritmo Regresión lineal es una variación del algoritmo de árboles de decisión, donde el mínimo de nodos S se establece para ser mayor o igual que el número total de escenarios del conjunto de datos que el algoritmo utiliza para realizar el entrenamiento del modelo de minería de datos. Con el parámetro así establecido, el algoritmo no crea nunca una división y, por tanto, lleva a cabo una regresión lineal. [34]

Se puede utilizar la regresión lineal para determinar una relación entre dos columnas continuas. La relación toma la forma de una ecuación para la línea que mejor represente una serie de datos. Por ejemplo, la línea del siguiente diagrama muestra la mejor representación lineal de los datos. [34]

La ecuación que representa la línea en el diagrama toma la forma general de $y = ax \pm b$ y es conocida como la ecuación de regresión. La variable Y representa la variable de salida, X representa la variable de entrada y a y b son coeficientes ajustables. Cada punto de datos del diagrama tiene un error asociado con su distancia con respecto a la línea de regresión. Los coeficientes a y b de la ecuación de regresión ajustan el ángulo y la ubicación de la línea de regresión. Puede obtener la ecuación de regresión ajustando a y b hasta que la suma de los errores asociados con los puntos alcance su cifra más baja. [34]

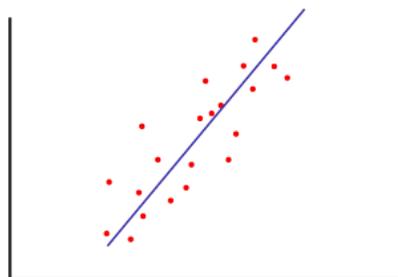


Figura 4.1.6.a Regresión Lineal

4.2 Primer análisis: Los tutorados

La presente sección está enfocada a realizar el análisis predictivo/descriptivo para la toma de decisiones dentro del programa tutoría “Nueva Era”, como se ha venido mencionando la minería de datos es un proceso iterativo, haciendo que la realización de un modelo “proponga” la realización de otro, y así sucesivamente hasta lograr obtener el conocimiento necesario para tomar la decisión más adecuada con base en todos ellos.

Este análisis tiene como objetivo el establecer la influencia que tiene la tutoría sobre el desempeño académico de los alumnos, de igual forma se pretende definir las características de los alumnos que asisten a la tutoría y en que difieren de los que no lo hacen, finalmente se busca encontrar si dichas características en común son compartidas para el desempeño de los alumnos durante el semestre, lo cual validaría o no el primer análisis. Todo lo anterior permitirá tomar una decisión adecuada para mejorar la asistencia de los alumnos, asegurando en gran medida sus beneficios.

Los análisis que han de llevarse a cabo para la toma de decisión son:

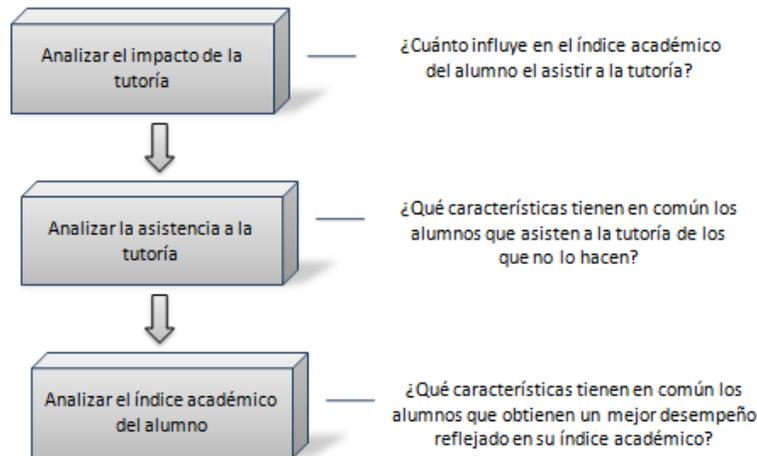


Figura 4.2.a Puntos a analizar para los tutorados

4.2.1 Analizar el impacto de la tutoría

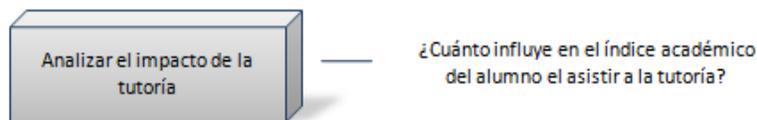


Figura 4.2.1.a El impacto

4.2.1.1 Definición del problema

Uno de los principales retos de un programa de tutoría es lograr que una proporción elevada de alumnos alcance niveles de desempeño académico favorables, esto reflejado directamente en el promedio del alumno, considerando además el número de materias aprobadas. Dicho lo anterior el siguiente análisis estará enfocado en encontrar si existe una relación directa entre asistir a las tutorías y el desempeño escolar del alumno, en otras palabras el objetivo de este análisis será:

Evaluar la importancia que tiene para los alumnos el asistir a la tutoría reflejado en su índice académico.

4.2.1.2 Vista minable

```

CREATE VIEW [TablaMinableRes] AS
-- Creando las tablas temporales para ser usadas mas tarde
WITH
    mc AS (SELECT idAlumno, count(*) as asignaturasCursadas FROM alumnomateria a
           group by idAlumno),
    mr AS (SELECT idAlumno, count(*) as asignaturasAprobadas FROM alumnomateria a
           where calificacion>=6 group by idAlumno)
-- Seleccionar todos los datos de la subconsulta y procesar el indice academico
select *,
CASE WHEN indiceAcademico >=9 THEN '9-10'
      WHEN indiceAcademico>=8 AND indiceAcademico<9 THEN '8-8.99'
      WHEN indiceAcademico>=7 AND indiceAcademico<8 THEN '7-7.99'
      WHEN indiceAcademico>=6 AND indiceAcademico<7 THEN '6-6.99'
      ELSE '0-5.99'
END as resultados
from (
-- Realizar la subconsulta
select idAlumno,
promedioPrimerSemestre,
asignaturasAprobadas,
asignaturasCursadas,
cast(asignaturasAprobadas as float) / cast(asignaturasCursadas as float) as avanceAcademico,
round(promedioPrimerSemestre * (cast(asignaturasAprobadas as float) / cast(asignaturasCursadas
as float)),2) as indiceAcademico,
CASE [¿A cuántas sesiones de tutoría grupal asististe?] WHEN '1' THEN 'SI' WHEN '2' THEN 'SI'
WHEN '3' THEN 'SI' WHEN '4' THEN 'SI' WHEN '5' THEN 'SI' WHEN '6' THEN 'SI' WHEN '7' THEN 'SI'
WHEN '8' THEN 'SI' WHEN 'Mas de 8' THEN 'SI' ELSE 'NO' END as AsistioTutGrupal,
CASE [¿A cuántas sesiones de tutoría individual asististe?] WHEN '1' THEN 'SI' WHEN '2' THEN
'SI' WHEN '3' THEN 'SI' WHEN '4' THEN 'SI' WHEN '5' THEN 'SI' WHEN '6' THEN 'SI' WHEN '7' THEN
'SI' WHEN '8' THEN 'SI' WHEN 'Mas de 8' THEN 'SI' ELSE 'NO' END as AsistioTutIndividual
from (
-- Esta nueva subconsulta es la que realiza el pivote
SELECT
r.idAlumno as idAlumno,
p.nombre as nombre,
i.opcion as opcion,
d.nombre as delegacion,
ap.promedio as promedioPrimerSemestre,
mc.asignaturasCursadas AS asignaturasCursadas,
CASE WHEN mr.asignaturasAprobadas IS NULL THEN '0' Else mr.asignaturasAprobadas END As
asignaturasAprobadas,
2009 - YEAR(fechaNac) as edad,
idCarrera as carrera
FROM resultados r
inner join pregunta p on (r.idPregunta=p.idPregunta)
inner join incisos i on (r.idIncisos=i.idIncisos)
inner join alumno a on (r.idAlumno=a.idAlumno)
inner join delegacion d on (d.idDelegacion=a.idDelegacion)
inner join alumnopromedio ap on (ap.idAlumno=a.idAlumno)
inner join mc on (mc.idAlumno=a.idAlumno)
left join mr on mr.idAlumno=mc.idAlumno
where ap.idSemestre = 1
) piv
PIVOT(
    Max(piv.opcion)
    FOR piv.nombre IN
    (

```

```

-- Filas que seran pivoteadas (convertidas en columnas)
[¿A cuántas sesiones de tutoría grupal asististe?],
[¿A cuántas sesiones de tutoría individual asististe?]
)
) as consulta_pivote) as tbl
GO

```

Nótese que se está haciendo uso de una de las ventajas que ofrece *SQL Server* en comparación con otros manejadores de bases de datos que es permitir consultas de referencia cruzada, esto mediante la instrucción *PIVOT*, la cual simplifica mucho la manera de realizar la creación de la vista minable. [35]

Nota: La descripción de los datos obtenidos en la vista minable se realizara posteriormente en el subtema Realizar Minería de Datos

4.2.1.3 Elección del algoritmo

Microsoft SQL Server proporciona varios algoritmos que se pueden usar en las soluciones de minería de datos, y los cuales se encuentran implementados por la herramienta de *Analysis Services* del mismo distribuidor. Para este análisis se utilizará el algoritmo de red neuronal, con el cual mediante las entradas de asistencia se buscará obtener las salidas “favorecidas”, lo cual permitirá saber la influencia que tiene la asistencia a las tutorías sobre el promedio del alumno.

4.2.1.4 Minería de datos

Una vez que se tiene la vista creada es necesario crear un origen de datos en *Business Intelligence Development*.

Un *origen de datos* es una conexión de datos en donde se almacena, se administra en el proyecto y se implementa en la base de datos de *Microsoft SQL Server Analysis Services*. El origen de datos contiene los nombres del servidor y la base de datos donde residen los datos de origen, además de otras propiedades de conexión necesarias. [36]

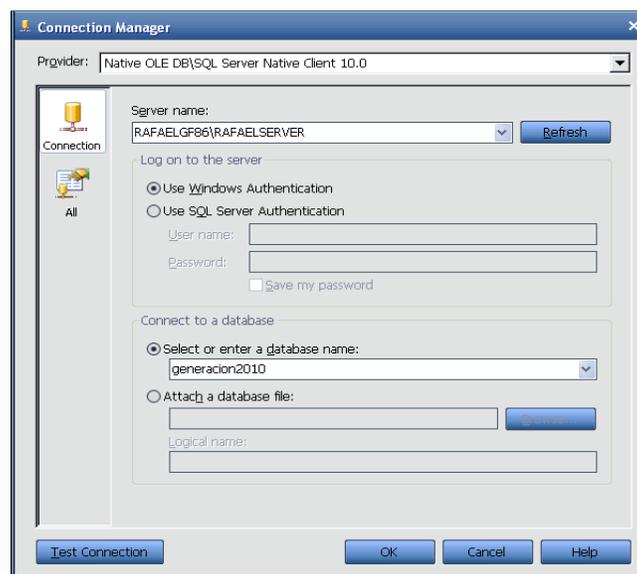


Figura 4.2.1.4.a Origen de datos

Nota: El origen de datos solo tiene que ser creado una única vez, solo que se decida usar otra base de datos, es necesario crear uno nuevo (Para este trabajo solo se creara una vez apuntando a la base de 'generacion2010').

De igual manera es necesario crear la vista del origen de datos que se utilizara para crear las estructuras de minería de datos. Una vista del origen de datos se genera en un origen de datos y define un subconjunto de los datos que rellenan un almacén de datos. De esta forma se puede modificar la estructura de los datos para que sean más significativos en el contexto del proyecto.

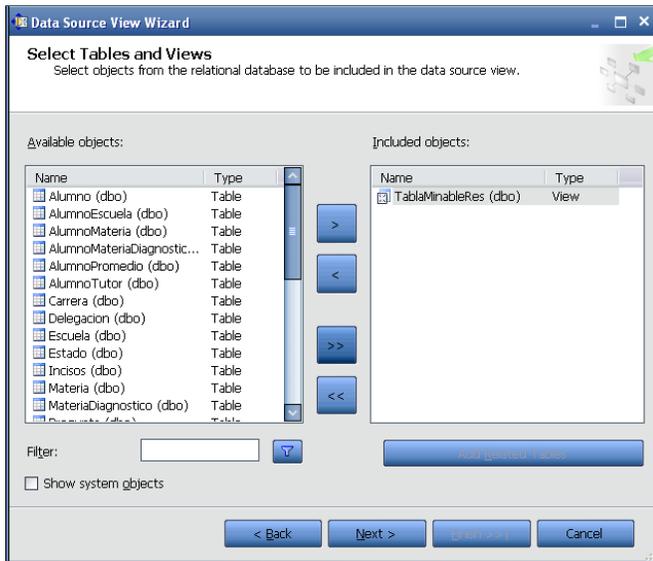


Figura 4.2.1.4.b Vista de origen de datos [1]

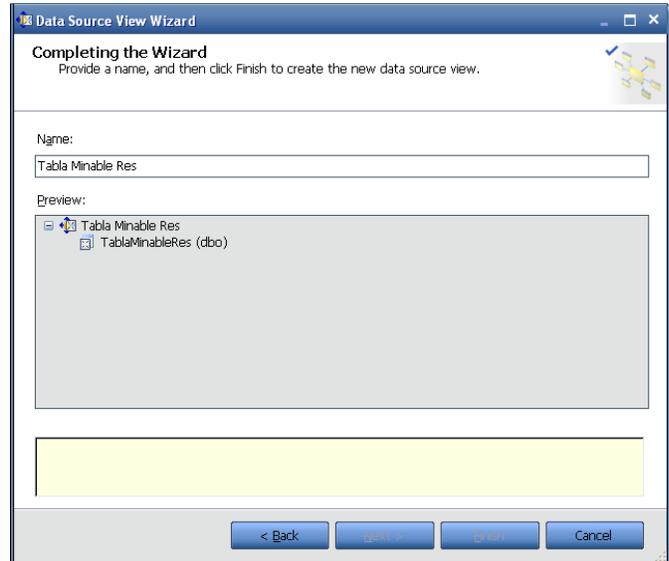


Figura 4.2.1.4.c Vista de origen de datos [2]

En ocasiones y previo a crear y procesar algún modelo es recomendable hacer un análisis de los datos existentes en la base, lo cual nos permitirá interpretar de mejor manera los resultados a obtener. *Microsoft SQL Server* permite ver de manera gráfica los datos de interés utilizando la vista del origen de datos, en este caso los datos a predecir son los resultados de los alumnos (su índice académico).

La siguiente gráfica muestra los resultados de los alumnos, los cuales han sido clasificados para facilitar su estudio.

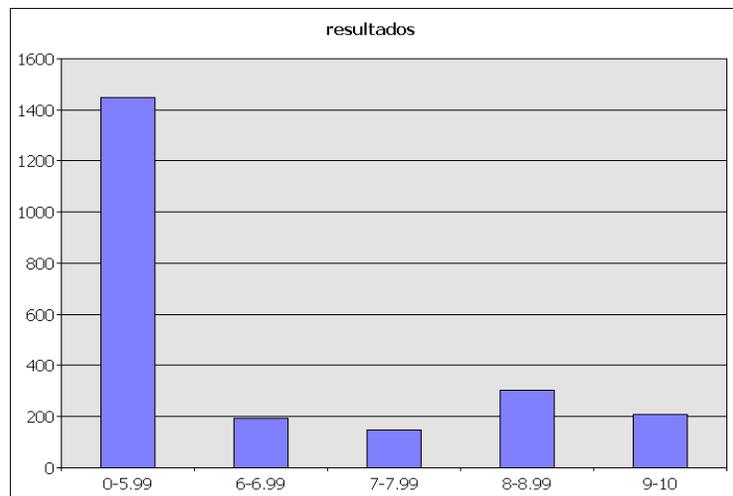


Figura 4.2.1.4.d Gráfica de resultados [1]

Como se puede observar en la gráfica la mayoría de los alumnos no han tenido un buen desempeño académico, esto es mas de 1400 de 2302 alumnos no consiguieron los resultados óptimos.

El modelo a utilizar se describe a continuación:

Características principales del modelo:

Estructura a la que pertenece: Tabla Minable Res

Modelo: Tabla Minable Res – Neuronal Network

Objetivo del modelo: Encontrar la influencia que tiene la tutoría sobre el promedio del alumno.

Atributo a predecir: resultados

Structure	Tabla Minable Res - Neuronal Network
Asistio Tut Grupal	Microsoft_Neural_Network
Asistio Tut Individual	Input
Id Alumno	Input
Promedio Primer Semestre	Key
Resultados	PredictOnly
	PredictOnly

Descripción de los datos a procesar

Indica si el alumno asistió al menos a una tutoría grupal. [Si] [No]
 Indica si el alumno asistió al menos a una tutoría individual. [Si] [No]
 Id del alumno
 Promedio del alumno durante el primer semestre
 Clasificación del resultado del alumno (índice académico¹)

Una vez hecho lo anterior es posible procesar el modelo con la estructura creada:

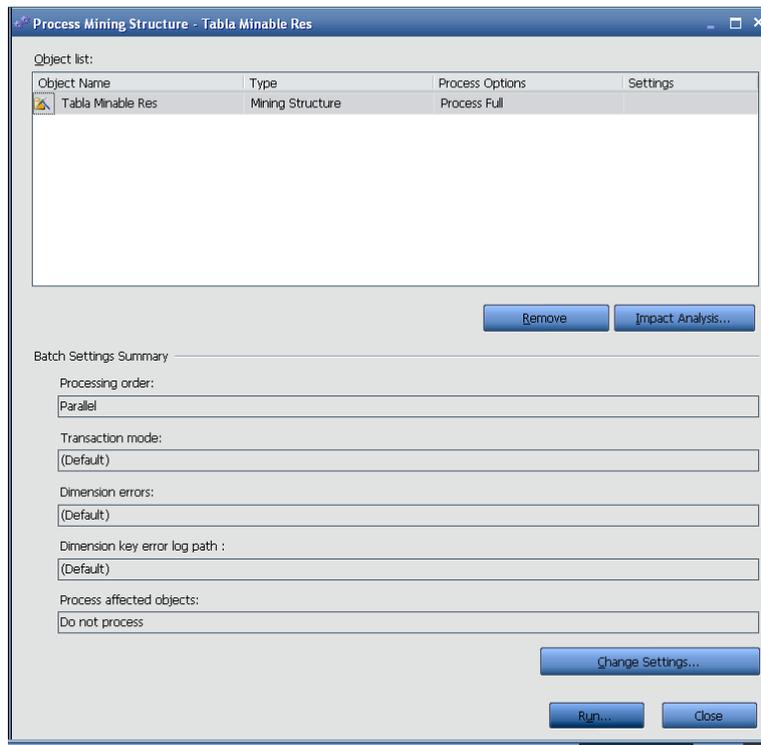


Figura 4.2.1.4.e Procesamiento del modelo

¹ El índice académico es la referencia en la cual es basado el número de inscripción, su fórmula es:

$$IA = Promedio\ del\ alumno * \frac{Asignaturas\ aprobadas}{Asignaturas\ cursadas}$$

Fórmula proporcionada por el Prof. José Enrique Larios Canale (Codirector de Tesis)

4.2.1.5 Patrones

Ya procesado el modelo, el visualizador del mismo presenta los resultados, dependiendo el algoritmo utilizado las opciones pueden variar.

En el caso de las redes neuronales la salida obtenida es la siguiente:

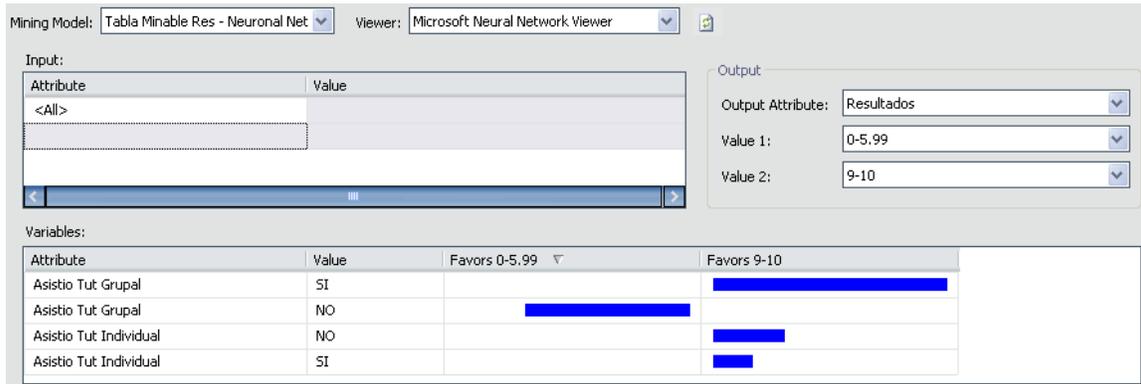


Figura 4.2.1.5.a Tendencia para el atributo 'Resultados'

En donde para un atributo de salida se pueden obtener la tendencia de los resultados para los atributos que fueron colocados como entrada en el proceso de minería. Así es como puede observarse que aquellos que asistieron a la tutoría grupal tuvieron una mayor tendencia a obtener un índice académico mucho más alto que aquellos que no asistieron. También es posible observar que la asistencia individual no influye en los resultados de los alumnos.

De igual manera es posible analizar otros valores como el promedio:

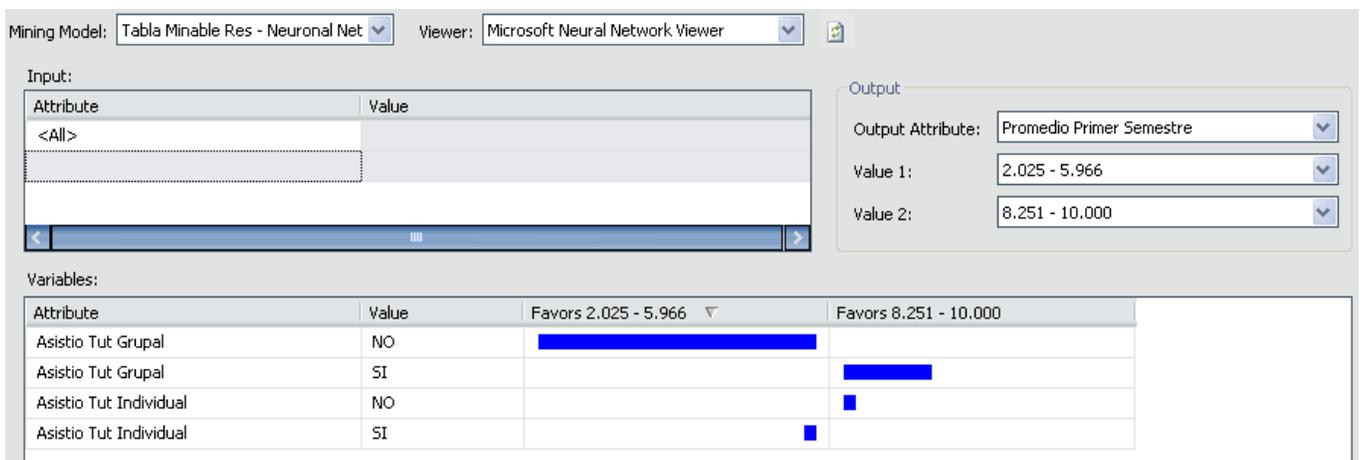


Figura 4.2.1.5.b Tendencia para el atributo 'Promedio Primer Semestre'

Donde se mantiene la misma tendencia “Aquellos que asistieron a la tutoría grupal obtuvieron un mayor promedio”.

4.2.1.6 Evaluación e interpretación del modelo

Para evaluar el modelo se utilizará un gráfico de elevación. En un gráfico de elevación se puede calcular la precisión de las predicciones de los modelos y comparar los resultados entre ellos. Normalmente, la exactitud de la predicción de un modelo de minería de datos se cuantifica mediante la elevación o la exactitud de la clasificación. El gráfico de elevación traza un modelo de suposición aleatorio así como un modelo ideal. [37]

Es posible crear dos tipos de gráficos de elevación:

- Un gráfico en el que se especifique un valor de destino para la columna de predicción: Este gráfico muestra en modo en que el modelo se comporta específicamente para un estado del atributo de predicción.
- Un gráfico en el que no se especifique el valor: Este gráfico muestra el modo en que el modelo se comporta para todos los estados del atributo de predicción.

Para este caso solo importa saber que tan preciso es el gráfico para medir el índice académico sin importarnos el rango.



Figura 4.2.1.6.a Gráfico de elevación

Explicación del gráfico de elevación

El eje X del gráfico representa el porcentaje del conjunto de datos de prueba que se utiliza para comparar las predicciones. El eje Y representa el porcentaje de predicciones correctas. Por consiguiente, la línea ideal es la línea diagonal, que muestra que en el 50 por ciento de los datos, el modelo predice correctamente el 50 por ciento de los casos, el máximo que se puede esperar.

De igual forma se puede observar la precisión de un modelo mediante una matriz de clasificación, la cual muestra el número de datos predichos correctamente.

Counts for Tabla Minable Res - Neuronal Network on Resultados:				
Predicted	0-5.99 (Actual)	6-6.99 (Actual)	7-7.99 (Actual)	8-8.99 (Actual)
0-5.99	242	37	24	50

Figura 4.2.1.6.b Matriz de clasificación

En la imagen anterior se puede observar que 242 casos fueron predichos correctamente y 111 fueron predichos de forma incorrecta.

Debido a la precisión de los datos (buena), no es necesario hacer otro análisis respecto al impacto de la tutoría, lo cual permite iniciar con el siguiente “subanálisis”: la asistencia.

4.2.2 La asistencia a la tutoría

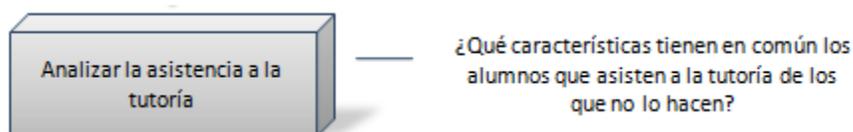


Figura 4.2.2.a La asistencia

4.2.2.1 Definición del problema

Dentro de cualquier programa de tutoría, un factor importante para su funcionamiento es sin duda alguna la confianza que logre establecer el tutor con el alumno en las sesiones. Dentro del programa Tutoría “Nueva Era” se trabajan 8 sesiones con temáticas específicas y sesiones adicionales que pueden ser trabajadas en forma individual o en pequeños grupos dependiendo de las necesidades y requerimientos de los estudiantes y el tutor. Sin embargo y como es de esperarse, existe una cantidad de alumnos que nunca asisten a las sesiones ya sean de tipo individual o grupal, así como en ocasiones no asisten a ninguna de las dos.

Con base en la información recopilada y como se mostro en el capítulo anterior, los datos obtenidos referentes a los alumnos de la generación 2010 permiten conocer muchas de sus características fuera y dentro de la facultad de Ingeniería de la UNAM, de igual manera se conocen sus respuestas para cada una de las preguntas de la encuesta de tutores, entre las que se incluyen la cantidad de veces que asistieron a las tutorías grupales e individuales.

Es por todo lo anterior que este análisis puede definirse como:

Clasificar a los alumnos que asisten a la tutoría, esto es que características tienen en común y en que difieren de aquellos que no lo hacen.

4.2.2.2 Vista Mizable

```

create [TablaMizable] as
-- Creando las tablas temporales para ser usadas mas tarde
WITH
amd AS (select * from (SELECT idAlumno,ROUND(avg(calificacion),2) as t1 FROM
alumnomateriadiagnostico
group by idAlumno) as t1),
ae AS (select * from (select i.idAlumno,e.nombre as escuela,r.descripcion as
PromedioBachillerato from alumno i
inner join alumnosescuela ae on (i.idAlumno=ae.idAlumno)
inner join rango r on (ae.idRango=r.idRango)
inner join escuela e on (ae.idEscuela=e.idEscuela)
where nivel=3) as t2)
-- Seleccionar elementos
select idAlumno,delegacion,promedioED,escuela,
promedioBachillerato,edad,carrera,
CASE [¿A cuántas sesiones de tutoría grupal asististe?] WHEN '1' THEN 'SI' WHEN '2' THEN 'SI'
WHEN '3' THEN 'SI' WHEN '4' THEN 'SI' WHEN '5' THEN 'SI' WHEN '6' THEN 'SI' WHEN '7' THEN 'SI'
WHEN '8' THEN 'SI' WHEN 'Mas de 8' THEN 'SI' ELSE 'NO' END as AsistioTutGrupal,
CASE [¿A cuántas sesiones de tutoría individual asististe?] WHEN '1' THEN 'SI' WHEN '2' THEN
'SI' WHEN '3' THEN 'SI' WHEN '4' THEN 'SI' WHEN '5' THEN 'SI' WHEN '6' THEN 'SI' WHEN '7' THEN
'SI' WHEN '8' THEN 'SI' WHEN 'Mas de 8' THEN 'SI' ELSE 'NO' END as AsistioTutIndividual,
[¿Cuál es la razón principal por la que elegiste la carrera de ingeniería?] as razonPricipal,
CASE [¿Cuál es la razon principal por la que elegiste la unam para estudiar ingeniería?] WHEN
' Por que aquí han estudiado tus familiares' THEN '1' ELSE '2' END as familiaresUNAM,
[Al terminar la carrera desearías:] as alTerminarCarrera,
CASE [Disfrutas en tu vivienda de computadora] WHEN 'SI' THEN '1' WHEN 'NO' THEN '2' END as
Computadora,
CASE [Disfrutas en tu vivienda de internet] WHEN 'SI' THEN '1' WHEN 'NO' THEN '2' END as
Internet,
CASE [Disfrutas en tu vivienda de automovil propio] WHEN 'SI' THEN '1' WHEN 'NO' THEN '2' END
as AutPropio,
CASE [Disfrutas en tu vivienda de automóvil familiar] WHEN 'SI' THEN '1' WHEN 'NO' THEN '2'
END as AutFamiliar,
CASE [¿Trabajas?] WHEN 'Sí, permanentemente' THEN '1' WHEN 'Sí, eventualmente' THEN '1' WHEN
'No' THEN '2' END as Trabaja,
CASE [¿Cuántas horas trabajas en promedio, a la semana?] WHEN 'Menos de 6 horas' THEN '1-5'
WHEN 'Entre 6 y 12 horas' THEN '6-12' WHEN 'Entre 13 y 20 horas' THEN '13-20' WHEN 'Entre 21 y
25 horas' THEN '21-25' WHEN 'Entre 26 y 30 horas' THEN '26-30' WHEN 'Entre 31 y 35 horas' THEN
'31-35' WHEN 'Entre 36 y 40 horas' THEN '36-40' WHEN 'No aplica' THEN '0' END as HrsTrabajo,
[Respecto a tus padres :] as Padres,
[Señala cuál es el nivel máximo de estudios de tu padre:] as NivelPadre,
[Señala cuál es el nivel máximo de estudios de tu madre:] as NivelMadre,
[Tu orientación vocacional está:] as OrientacionVocacional,
[Considerando tu vida escolar, ¿cómo te calificarías como estudiante?] as ComoEstudiante,
CASE [En promedio, el tiempo al día, que usas computadora es:] WHEN 'Máximo una hora' THEN
'0-60' WHEN 'Entre una y dos horas' THEN '61-120' WHEN 'Entre dos y tres horas' THEN '121-180'
WHEN 'Entre dos y tres horas' THEN '121-180' WHEN 'Más de tres horas' THEN '180->' WHEN 'No
utilizas la computadora' THEN '0' END as UsoComputadora,
CASE [En promedio, el tiempo al día, que navegas en internet es:] WHEN 'Máximo media hora'
THEN '0-30' WHEN 'Entre media hora y una hora' THEN '31-60' WHEN 'Entre una y dos horas' THEN
'61-120' WHEN 'Entre dos y tres horas' THEN '121-180' WHEN 'Más de tres horas' THEN '180->'
WHEN 'No navegas en Internet' THEN '0' END as UsoIntenet,
[Señala cuál es la principal ocupación de tu padre:] as OcPadre,
[Señala cuál es la principal ocupación de tu madre:] as OcMadre,
[¿Principalmente, de quién dependes económicamente?] as DepEconomica,
[¿Cuántas personas contribuyen al sostenimiento de tu hogar?] as PersonasContribuyen,
CASE [¿A cuánto asciende el ingreso mensual total de tu hogar?] WHEN 'Menos de $3000' THEN
'$0-$3,000' WHEN 'Entre $3001 y $5000' THEN '$3001-$5000' WHEN 'Entre $5001 y $7000' THEN
'$5001-$7000' WHEN 'Entre $7001 y $9000' THEN '$7001-$9000' WHEN 'Entre $9001 y $11000' THEN
'$9001-$11000' WHEN 'Más de $11000' THEN '$11000-'END as IngMensual,

```

```

[¿Cómo te transportarás a la universidad?] as Transporte,
[¿Cuánto tiempo emplearas diariamente en transporte para ir y venir a la universidad?] as
TiempoTransporte,
CASE [¿A cuánto asciende el ingreso mensual total de tu hogar?] WHEN 'Menos de $3000' THEN 'E'
WHEN 'Entre $3001 y $5000' THEN 'D' WHEN 'Entre $5001 y $7000' THEN 'D' WHEN 'Entre $7001 y
$9000' THEN 'D+' WHEN 'Entre $9001 y $11000' THEN 'D+' WHEN 'Más de $11000' THEN 'C'END as
claseSocioeconomica
from (
-- Esta nueva subconsulta es la que realiza el pivote
SELECT
r.idAlumno as idAlumno,
p.nombre as nombre,
i.opcion as opcion,
d.nombre as delegacion,
ap.promedio as promedioPrimerSemestre,
amd.tl as promedioED,
ae.escuela as escuela,
ae.PromedioBachillerato as promedioBachillerato,
2009 - YEAR(fechaNac) as edad,
idCarrera as carrera
FROM resultados r
inner join pregunta p on (r.idPregunta=p.idPregunta)
inner join incisos i on (r.idIncisos=i.idIncisos)
inner join alumno a on (r.idAlumno=a.idAlumno)
inner join delegacion d on (d.idDelegacion=a.idDelegacion)
inner join alumнопromedio ap on (ap.idAlumno=a.idAlumno)
inner join alumнопromedio ap2 on (ap2.idAlumno=a.idAlumno)
inner join amd on (amd.idAlumno=a.idAlumno)
inner join ae on (ae.idAlumno=a.idAlumno)
where ap.idSemestre = 1
) piv
PIVOT(
Max(piv.opcion)
FOR piv.nombre IN
(
[¿A cuántas sesiones de tutoría grupal asististe?],
[¿A cuántas sesiones de tutoría individual asististe?],
[Disfrutas en tu vivienda de computadora],
[Disfrutas en tu vivienda de internet],
[Disfrutas en tu vivienda de automovil propio],
[Disfrutas en tu vivienda de automóvil familiar],
[¿Trabajas?],
[¿Cuántas horas trabajas en promedio, a la semana?],
[Respecto a tus padres :],
[Señala cuál es el nivel máximo de estudios de tu padre:],
[Señala cuál es el nivel máximo de estudios de tu madre:],
[Tu orientación vocacional está:],
[Considerando tu vida escolar, ¿cómo te calificarías como estudiante?],
[En promedio, el tiempo al día, que usas computadora es:],
[En promedio, el tiempo al día, que navegas en internet es:],
[Señala cuál es la principal ocupación de tu padre:],
[Señala cuál es la principal ocupación de tu madre:],
[¿Principalmente, de quién dependes económicamente?],
[¿Cuántas personas contribuyen al sostenimiento de tu hogar?],
[¿A cuánto asciende el ingreso mensual total de tu hogar?],
[¿Cómo te transportarás a la universidad?],
[¿Cuánto tiempo emplearas diariamente en transporte para ir y venir a la
universidad?],
[¿Cuál es la razón principal por la que elegiste la carrera de ingeniería?],
[¿Cuál es la razón principal por la que elegiste la unam para estudiar ingeniería?],
[Al terminar la carrera desearias:]
)
) as consulta_pivote

```

4.2.2.3 Elección del algoritmo

Debido a la gran cantidad de datos y variedad de los mismos (tipos de datos), en este planteamiento se utilizaron tres tipos diferentes de algoritmos, dos de ellos pertenecientes al agrupamiento y uno a la clasificación: el de clústeres y el algoritmo de árboles de decisión para el primer grupo y el algoritmo de Naive Bayes para el segundo. Después de analizarlos de forma individual (una vez hecha su evaluación), es necesario realizar una comparación de los mismos utilizando gráficos de elevación, el cual permita elegir a uno de ellos como el más adecuado para resolver el problema.

4.2.2.4 Minería de Datos - Clústeres

Como se mencionó anteriormente previo a crear y procesar algún modelo es recomendable hacer un análisis de los datos existentes en la base, lo cual nos permitirá interpretar de mejor manera los resultados a obtener. *Microsoft SQL Server* permite ver de manera grafica los datos de interés utilizando la vista del origen de datos, en este caso como lo que nos interesa saber son las principales características de los alumnos que asistieron, los atributos a graficar serán los de "AsistioTutGrupal" y el de "AsistioTutIndividual" el cual nos indica con 1 si el alumno asistió al menos a una tutoría y con un 2 si no fue así, dichos atributos son de tipo discreto por lo cual el resultado será mostrado con graficas de barras

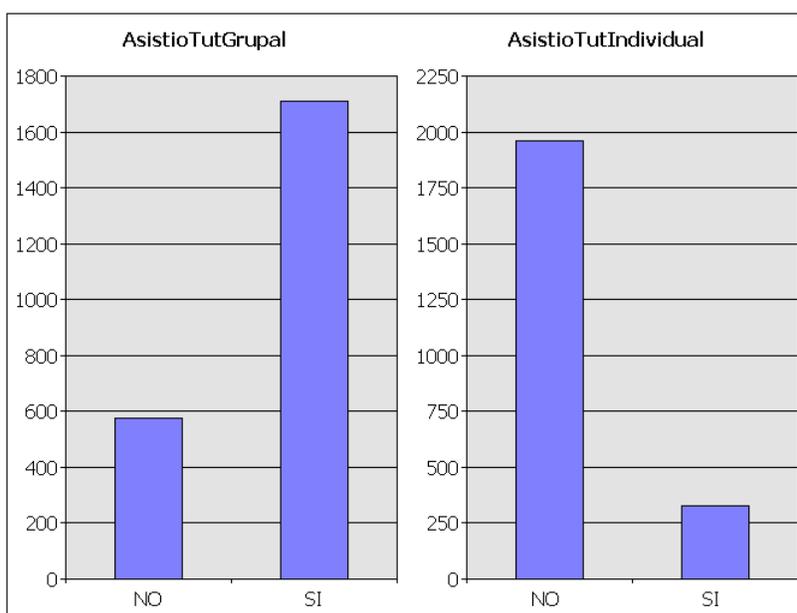


Figura 4.2.2.4.a Gráficas de asistencia

Las gráficas anteriores nos permiten saber que la asistencia a las tutorías grupales fue mucho mayor a la individual, por lo cual los datos obtenidos al procesar el modelo de minería de datos serán más relevantes en la tutoría grupal que en la individual, (como se observó en el primer análisis).

Con los pasos realizados anteriormente, ya puede ser creado el modelo de estructura de datos, en el cual se puede elegir el algoritmo a utilizar y otras características importantes, entre las que se encuentran: el atributo que será considerado como la llave primaria, el o los atributos a predecir, el o los atributos de entrada y cuales serán ignorados.

El modelo a utilizar se describe a continuación:

Características principales del modelo:

Estructura a la que pertenece: Tabla Minable

Modelo: **Tabla Minable – Clústeres**

Objetivo del modelo: Obtener las principales características de los alumnos que asistieron a las tutorías individuales y/o grupales, así como aquellos que no lo hicieron.

Atributos a predecir: Asistió Tut Grupal, Asistió Tut Individual

Structure	Tabla Minable - Clusters	Descripción de los datos a procesar
	Microsoft_Clustering	
Al Terminar Carrera	Input	Indica las aspiraciones del alumno al terminar la carrera
Asistio Tut Grupal	PredictOnly	Indica si el alumno asistió al menos a una tutoría grupal
Asistio Tut Individual	PredictOnly	Indica si el alumno asistió al menos a una tutoría individual
Aut Familiar	Input	Indica si el alumno tiene auto familiar [1] = Si, [2] = No
Aut Propio	Input	Indica si el alumno tiene auto propio [1] = Si, [2] = No
Carrera	Input	Indica el id de la carrera del alumno
Clase Socioeconomica	Input	Indica la clase socioeconómica del alumno
Como Estudiante	Input	Indica cómo se considera el alumno
Computadora	Input	Indica si el alumno tiene computadora [1] = Si, [2] = No
Delegacion	Input	Indica la delegación donde reside actualmente el alumno
Dep Economica	Input	Indica de quien depende económica el alumno
Edad D	Ignore	Indica la edad del alumno (atributo discreto)
Edad C	Input	Indica la edad del alumno (atributo continuo)
Escuela	Input	Indica la escuela a que asistió el alumno en preparatoria
Familiares UNAM	Input	Indica si el alumno tiene familiares que estudiaron en la UNAM [1] = Si, [2] = No
Hrs Trabajo	Input	Indica las horas que trabaja el alumno
Id Alumno	Key	Id del alumno
Ing Mensual	Input	Rango de ingresos mensuales de todos los integrantes en el hogar del alumno
Internet	Input	Indica si el alumno tiene internet en su hogar [1] = Si, [2] = No
Nivel Madre	Input	Indica el nivel de estudios de la madre del alumno
Nivel Padre	Input	Indica el nivel de estudios del padre del alumno
Oc Madre	Input	Indica la ocupación de la madre del alumno
Oc Padre	Input	Indica la ocupación del padre del alumno
Orientacion Vocacional	Input	Indica la orientación vocacional del alumno
Padres	Input	Indica el estado civil de los padres del alumno
Personas Contribuyen	Input	Indica cuantas personas contribuyen económicamente en el hogar del alumno
Promedio Bachillerato	Input	Indica el rango del promedio en bachillerato del alumno
Promedio ED	Input	Indica el promedio del alumno en el examen diagnostico
Razon Prcipal	Input	Indica la razón principal de que el alumno este estudiando su carrera
Tiempo Transporte	Input	Indica el tiempo de transporte desde el hogar del alumno hasta la facultad
Trabaja	Input	Indica si el alumno trabaja o no [1] = Si, [2] = No
Transporte	Input	Indica el tipo de transporte que utiliza el alumno para llegara a la facultad
Uso Computadora	Input	Indica el rango de horas al día que el alumno utiliza la computadora
Uso Internet	Input	Indica el rango de horas al día que el alumno utiliza el internet

Nota: El atributo de la edad fue duplicado utilizando las facilidades de SQL Server para agregar una nueva columna (indicándole el tipo de valor). Lo anterior debido a que la edad como atributo continuo no es permitido para el algoritmo de Bayes (que se utilizara más adelante), por lo cual se ha declarado como parte del modelo y que para esta estructura será simplemente ignorado.

4.2.2.5 Patrones - Clústeres

A continuación se muestran los resultados obtenidos al procesar la estructura antes mencionada y en la cual ya se han identificado los grupos de interés, que son los de mayor y menor asistencia para las tutorías individuales y colectivas, en este caso el clúster de menor asistencia es el mismo para ambas tutorías.

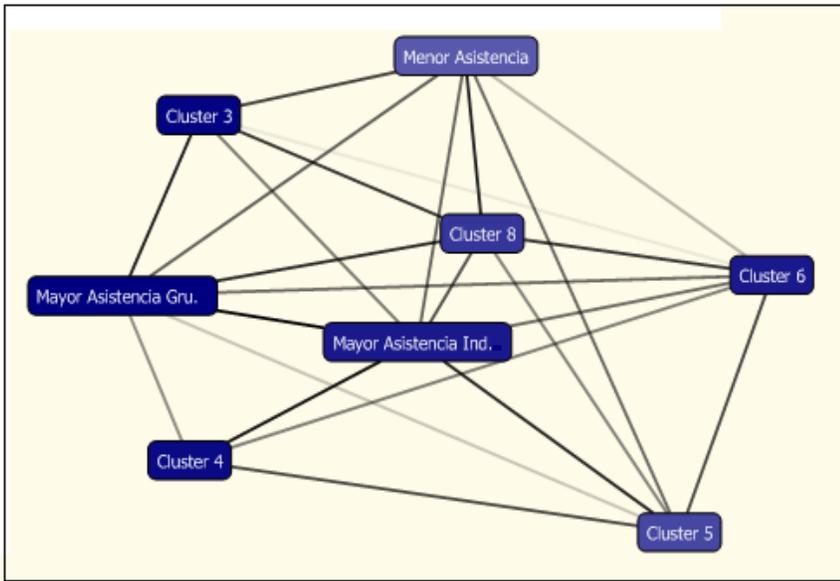


Figura 4.2.2.5.a Diagrama del clúster

La imagen es conocida como diagrama del clúster y en él se muestran todos los clústeres de un modelo de minería de datos. Los clústeres más sombreados son aquellos que tienen más probabilidad de obtener el valor a predecir para un atributo dado (En este caso la asistencia Grupal e Individual).

El sombreado de la línea que conecta un clúster con otro representa la importancia de la similitud de los clústeres. Si el sombreado es claro o inexistente, los clústeres no son muy similares. A medida que la línea se va oscureciendo, va aumentando la similitud de los vínculos. [27]

Al analizar los Clústeres de interés “Mayor Asistencia Grupal” y “Menor Asistencia”, los datos obtenidos fueron para cada uno respectivamente:

Cluster: Mayor Asistencia Gr.		
Characteristics for Mayor Asistencia Grupal		
Variables	Values	Probability
Familiares UNAM	2	
Aut Propio	2	
Transporte	En transporte urbano	
Computadora	1	
Padres	Viven ambos, juntos	
Internet	1	
Hrs Trabajo	0	
Trabaja	2	
Edad C	18 - 19	
Aut Familiar	1	
Personas Contribuyen	Dos	
Orientacion Vocacional	Bien definida y corresponde plenam...	
Escuela	Escuela Nacional Preparatoria de la ...	
Dep Economica	De ambos padres	
Oc Padre	Empleado (a) de base o confianza e...	
Como Estudiante	Bueno (a)	
Al Terminar Carrera	Recibirte y estudiar un posgrado en ...	
Clase Socioeconomica	D	
Oc Madre	Empleado (a) de base o confianza e...	
Como Estudiante	Muy bueno (a)	
Uso Internet	61-120	
Clase Socioeconomica	D+	

Figura 4.2.2.5.b Clúster de mayor asistencia grupal

Cluster: Menor Asistencia		
Characteristics for Menor Asistencia		
Variables	Values	Probability
Computadora	1	
Familiares UNAM	2	
Asisto Tut Individual	NO	
Internet	1	
Aut Familiar	1	
Aut Propio	2	
Transporte	En transporte urbano	
Padres	Viven ambos, juntos	
Trabaja	1	
Como Estudiante	Bueno (a)	
Orientacion Vocacional	Bien definida y corresponde plenam...	
Edad C	21 - 24	
Personas Contribuyen	Dos	
Asisto Tut Grupal	NO	
Oc Madre	Labores del hogar	
Al Terminar Carrera	Recibirte y estudiar un posgrado en ...	
Promedio Bachillerato	Entre 7.1 y 7.5	
Al Terminar Carrera	Recibirte y estudiar un posgrado en ...	
Dep Economica	De tu padre	
Uso Internet	61-120	
Orientacion Vocacional	Bien definida y corresponde parcial...	
Clase Socioeconomica	C	

Figura 4.2.2.5.c Clúster de menor asistencia grupal

Adicionado a lo anterior, es posible realizar una discriminación entre *Clústeres*, lo cual permite comparar los atributos de dos clústeres. El sistema determina las diferencias más importantes entre los clústeres y muestra los estados de atributo asociados con las diferencias por orden de importancia. Una barra a la derecha del atributo muestra el clúster que favorece el estado; el tamaño de la barra muestra la intensidad con la que lo favorece. Al realizar dicha distinción entre el clúster de Mayor Asistencia Grupal y el de Menor Asistencia se observa que:

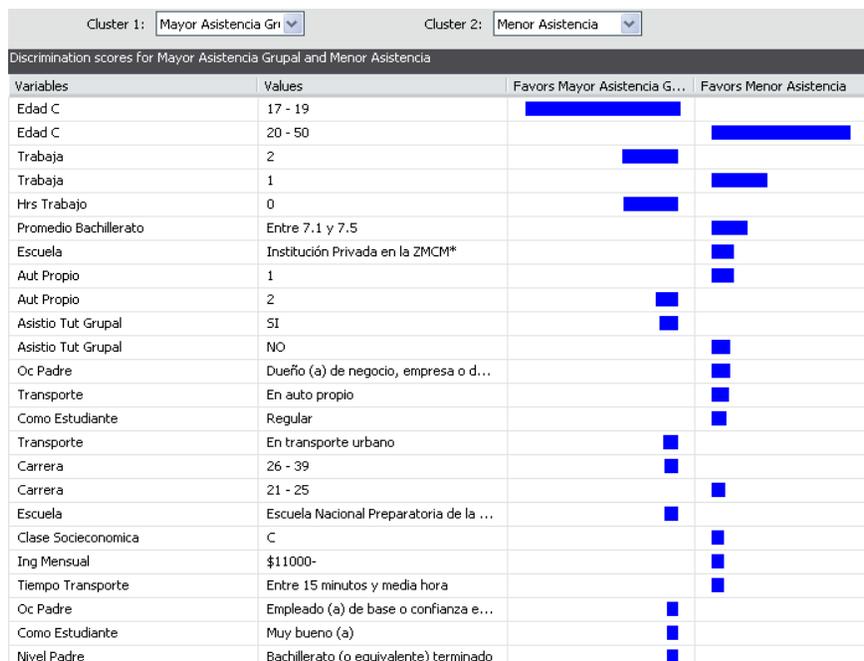


Figura 4.2.2.5.d Discriminación de clústeres de asistencia grupal

A continuación se procede a analizar los *Clústeres* “Mayor Asistencia Individual” y “Menor Asistencia”:

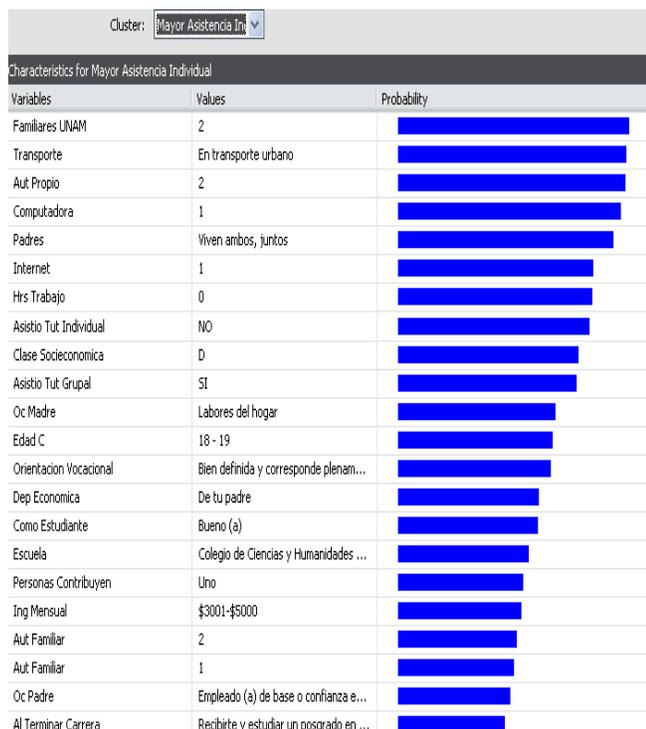


Figura 4.2.2.5.e Clúster de mayor asistencia individual

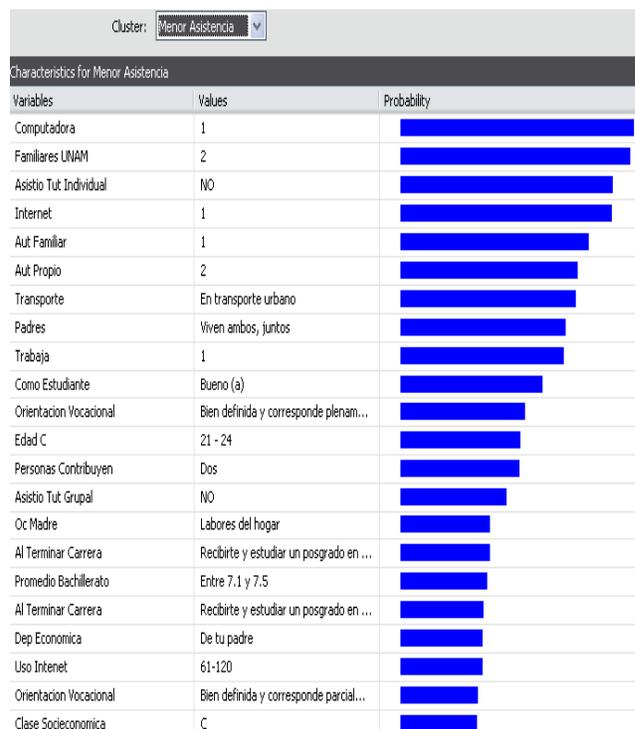


Figura 4.2.2.5.f Clúster de menor asistencia individual

Al realizar la distinción del Clúster para obtener las principales diferencias entre uno y otro podremos obtener que:

Cluster 1: Mayor Asistencia Ind		Cluster 2: Menor Asistencia	
Discrimination scores for Mayor Asistencia Individual and Menor Asistencia			
Variables	Values	Favors Mayor Asistencia I...	Favors Menor Asistencia
Edad C	17 - 19	██████████	
Edad C	20 - 50		██████████
Trabaja	1		██████████
Trabaja	2	██████████	
Ing Mensual	\$11000-		██████████
Clase Socioeconomica	C		██████████
Hrs Trabajo	0	██████████	
Ing Mensual	\$3001-\$5000	██████████	
Clase Socioeconomica	D	██████████	
Escuela	Institución Privada en la ZMCM*		██████████
Promedio Bachillerato	Entre 7.1 y 7.5		██████████
Escuela	Colegio de Ciencias y Humanidades ...	██████████	
Nivel Madre	Licenciatura titulado (a)		██████████
Transporte	En transporte urbano	██████████	
Aut Propio	1		██████████
Aut Propio	2	██████████	
Oc Padre	Dueño (a) de negocio, empresa o d...		██████████
Transporte	En auto propio		██████████
Nivel Padre	Estudios de posgrado		██████████
Aut Familiar	1		██████████
Aut Familiar	2	██████████	
Oc Madre	Labores del hogar	██████████	
Padres	Viven ambos, juntos	██████████	
Promedio Bachillerato	Entre 8.6 y 9.0	██████████	

Figura 4.2.2.5.g Discriminación de clústeres de asistencia individual

Como se pudo observar en las imágenes anteriores, las principales diferencias entre los alumnos que asisten a las **tutorías grupales** y los que no, son el factor edad y el factor trabajo, de forma más clara aquellos alumnos que trabajan y tiene una edad mayor a los 20 años tienen una asistencia menor o casi nula en comparación de aquellos que no trabajan y tienen menos de 20 años.

De igual forma una característica que se presento en el clúster de mayor asistencia fue la escuela, que nos indica en qué tipo de institución estudio el alumno el nivel bachillerato, y cuyo valor para dicho clúster fue la Escuela Nacional Preparatoria, y al ver la discriminación de los *Clústeres*, aquellos que estudiaron en una escuela privada tuvieron una menor asistencia.

La importancia de los datos restantes es mucho menor que los antes descritos y pueden ser producto de los primeros, por ejemplo el atributo que hace referencia a los ingresos mensuales y que sobrepasa los \$11,000 se encuentra en el clúster de menor asistencia, esto puede ser debido a que el alumno trabaja y por lo cual sus ingresos son mayores sin embargo esta misma causa limita el tiempo que un alumno tendría para las tutorías.

Para la **tutoría individual** ocurrió algo muy semejante, los factores trabajo y edad fueron los más importantes para definir la asistencia, sin embargo en este análisis el tercer factor más importante fue el de ingresos mensuales y no el promedio de bachillerato como ocurrió en el caso de las asistencias grupales.

4.2.2.6 Minería de datos – Árboles de Decisión

Las características principales del modelo se describen a continuación.

Características principales del modelo:

Estructura a la que pertenece: Tabla Minable

Modelo: Tabla Minable – Trees

Objetivo del modelo: Obtener las principales características de los alumnos que asistieron a las tutorías individuales y/o grupales, así como aquellos que no lo hicieron.

Atributos a predecir: Asistió Tut Grupal, Asistió Tut Individual

Structure	Tabla Minable - Trees 1
	Microsoft_Decision_Trees
Al Terminar Carrera	Input
Asistio Tut Grupal	PredictOnly
Asistio Tut Individual	PredictOnly
Aut Familiar	Input
Aut Propio	Input
Carrera	Input
Clase Socioeconomica	Input
Como Estudiante	Input
Computadora	Input
Delegacion	Input
Dep Economica	Input
Edad D	Ignore
Edad C	Input
Escuela	Input
Familiares UNAM	Input
Hrs Trabajo	Input
Id Alumno	Key
Ing Mensual	Input
Internet	Input
Nivel Madre	Input
Nivel Padre	Input
Oc Madre	Input
Oc Padre	Input
Orientacion Vocacional	Input
Padres	Input
Personas Contribuyen	Input
Promedio Bachillerato	Input
Promedio ED	Input
Razon Pricpal	Input
Tiempo Transporte	Input
Trabaja	Input
Transporte	Input
Uso Computadora	Input
Uso Intenet	Input

Descripción de los datos a procesar

- Indica las aspiraciones del alumno al terminar la carrera
- Indica si el alumno asistió al menos a una tutoría grupal
- Indica si el alumno asistió al menos a una tutoría individual
- Indica si el alumno tiene auto familiar [1] = Si, [2] = No
- Indica si el alumno tiene auto propio [1] = Si, [2] = No
- Indica el id de la carrera del alumno
- Indica la clase socioeconómica del alumno
- Indica cómo se considera el alumno
- Indica si el alumno tiene computadora [1] = Si, [2] = No
- Indica la delegación donde reside actualmente el alumno
- Indica de quien depende económica el alumno
- Indica la edad del alumno (atributo discreto)
- Indica la edad del alumno (atributo continuo)
- Indica la escuela a que asistió el alumno en preparatoria
- Indica si el alumno tiene familiares que estudiaron en la UNAM [1] = Si, [2] = No
- Indica las horas que trabaja el alumno
- Id del alumno
- Rango de ingresos mensuales de todos los integrantes en el hogar del alumno
- Indica si el alumno tiene internet en su hogar [1] = Si, [2] = No
- Indica el nivel de estudios de la madre del alumno
- Indica el nivel de estudios del padre del alumno
- Indica la ocupación de la madre del alumno
- Indica la ocupación del padre del alumno
- Indica la orientación vocacional del alumno
- Indica el estado civil de los padres del alumno
- Indica cuantas personas contribuyen económicamente en el hogar del alumno
- Indica el rango del promedio en bachillerato del alumno
- Indica el promedio del alumno en el examen diagnostico
- Indica la razón principal de que el alumno este estudiando su carrera
- Indica el tiempo de transporte desde el hogar del alumno hasta la facultad
- Indica si el alumno trabaja o no [1] = Si, [2] = No
- Indica el tipo de transporte que utiliza el alumno para llegara a la facultad
- Indica el rango de horas al día que el alumno utiliza la computadora
- Indica el rango de horas al día que el alumno utiliza el internet

4.2.2.7 Patrones – Árboles de decisión

Al procesar el modelo el árbol obtenido para la asistencia grupal e individual fueron:

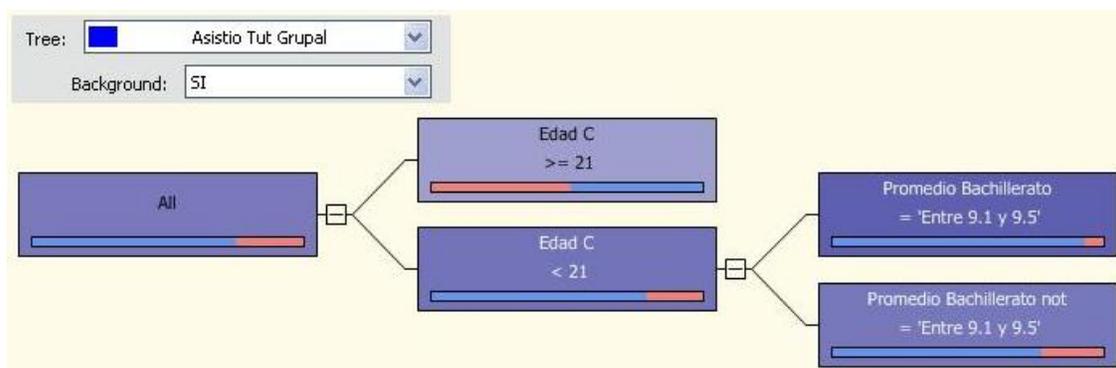


Figura 4.2.2.7.a Árbol de decisión asistencia grupal

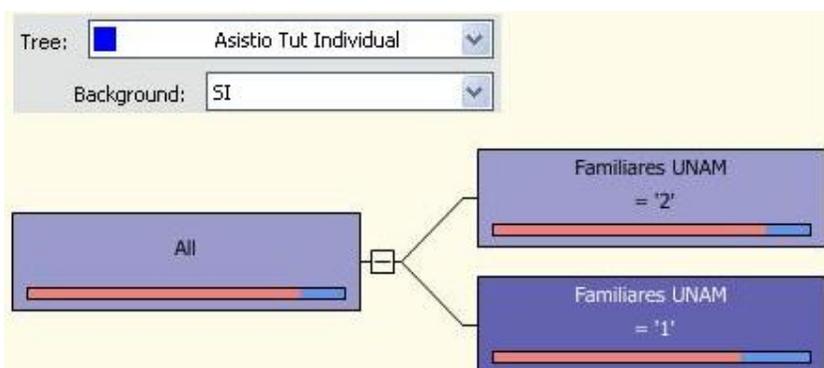


Figura 4.2.2.7.b Árbol de decisión asistencia individual

El color de fondo de cada nodo representa la concentración de casos del estado del atributo concreto que selecciona, y que en este caso fue la opción 'SI'. Además la línea dentro de cada nodo representa el total de alumnos que tiene, mostrando en azul a los que asistieron a la tutoría y el color rosa a aquellos que no lo hicieron.

Considerando el modelo Tabla Minable – *Clústeres* se puede observar que la semejanza principal entre ambos es la edad, y al igual que el primer modelo a menor edad mayor asistencia en la tutoría grupal, sin embargo este modelo no considero importante si el alumno trabaja o no, para este modelo el factor más importante después de la edad fue el promedio que el alumno obtuvo en el nivel bachillerato, donde si el promedio estuvo en un rango elevado mayor fue la asistencia a la tutoría grupal, esto puede deberse al compromiso que adquiere un alumno al haber obtenido buenas notas en sus estudios previos a la carrera y por consiguiente querer mantenerlos en el transcurso de la misma.

Para la asistencia a la tutoría individual el atributo más importante fue si el alumno tenía o no parientes que hubieran estudiado alguna carrera de la UNAM, sin embargo la diferencia entre uno y otro es muy poca debido a que como se analizo antes de procesar los modelos el número de alumnos que asistió a la tutoría individual fue mucho menor que el que asistió a la grupal, lo cual limita el análisis para este tipo de tutoría.

4.2.2.8 Realizar minería de datos – Naive Bayes

El siguiente y último modelo para este planteamiento es el que utiliza el algoritmo de Naive Bayes:

Características principales del modelo:

Estructura a la que pertenece: Tabla Minable

Modelo: Tabla Minable – Bayes

Objetivo del modelo: Obtener las principales características de los alumnos que asistieron a las tutorías individuales y/o grupales, así como aquellos que no lo hicieron.

Atributos a predecir: Asistió Tut Grupal, Asistió Tut Individual

Structure	Tabla Minable - Bayes
	Microsoft_Naive_Bayes
Al Terminar Carrera	Input
Asistio Tut Grupal	PredictOnly
Asistio Tut Individual	PredictOnly
Aut Familiar	Input
Aut Propio	Input
Carrera	Ignore
Clase Socioeconomica	Input
Como Estudiante	Input
Computadora	Input
Delegacion	Input
Dep Economica	Input
Edad D	Input
Edad C	Ignore
Escuela	Input
Familiares UNAM	Input
Hrs Trabajo	Input
Id Alumno	Key
Ing Mensual	Input
Internet	Input
Nivel Madre	Input
Nivel Padre	Input
Oc Madre	Input
Oc Padre	Input
Orientacion Vocacional	Input
Padres	Input
Personas Contribuyen	Input
Promedio Bachillerato	Input
Promedio ED	Ignore
Razon Pricipal	Input
Tiempo Transporte	Input
Trabaja	Input
Transporte	Input
Uso Computadora	Input
Uso Internet	Input

Descripción de los datos a procesar

- Indica las aspiraciones del alumno al terminar la carrera
- Indica si el alumno asistió al menos a una tutoría grupal
- Indica si el alumno asistió al menos a una tutoría individual
- Indica si el alumno tiene auto familiar [1] = Si, [2] = No
- Indica si el alumno tiene auto propio [1] = Si, [2] = No
- Indica el id de la carrera del alumno
- Indica la clase socioeconómica del alumno
- Indica cómo se considera el alumno
- Indica si el alumno tiene computadora [1] = Si, [2] = No
- Indica la delegación donde reside actualmente el alumno
- Indica de quien depende económica el alumno
- Indica la edad del alumno (atributo discreto)
- Indica la edad del alumno (atributo continuo)
- Indica la escuela a que asistió el alumno en preparatoria
- Indica si el alumno tiene familiares que estudiaron en la UNAM [1] = Si, [2] = No
- Indica las horas que trabaja el alumno
- Id del alumno
- Rango de ingresos mensuales de todos los integrantes en el hogar del alumno
- Indica si el alumno tiene internet en su hogar [1] = Si, [2] = No
- Indica el nivel de estudios de la madre del alumno
- Indica el nivel de estudios del padre del alumno
- Indica la ocupación de la madre del alumno
- Indica la ocupación del padre del alumno
- Indica la orientación vocacional del alumno
- Indica el estado civil de los padres del alumno
- Indica cuantas personas contribuyen económicamente en el hogar del alumno
- Indica el rango del promedio en bachillerato del alumno
- Indica el promedio del alumno en el examen diagnostico
- Indica la razón principal de que el alumno este estudiando su carrera
- Indica el tiempo de transporte desde el hogar del alumno hasta la facultad
- Indica si el alumno trabaja o no [1] = Si, [2] = No
- Indica el tipo de transporte que utiliza el alumno para llegara a la facultad
- Indica el rango de horas al día que el alumno utiliza la computadora
- Indica el rango de horas al día que el alumno utiliza el internet

Nota: Este algoritmo no acepta valores continuos, por tal razón se duplico el atributo edad con otro tipo de dato (discreto), lo anterior fue realizado utilizando las facilidades de SQL Server para agregar una nueva columna (indicándole el tipo de valor).

4.2.2.9 Patrones – Naive Bayes

Este algoritmo proporciona una red de dependencia la cual nos indica de cuales atributos depende aquel que se va a predecir:

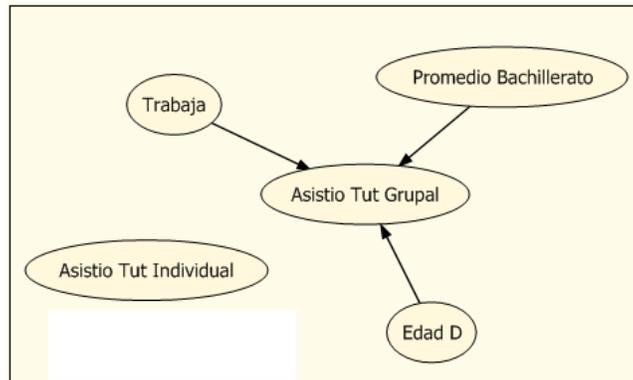


Figura 4.2.2.9.a Red de dependencia

De igual forma podemos analizar el porqué de esas dependencias analizando los perfiles de los atributos obtenidos por el mismo algoritmo.

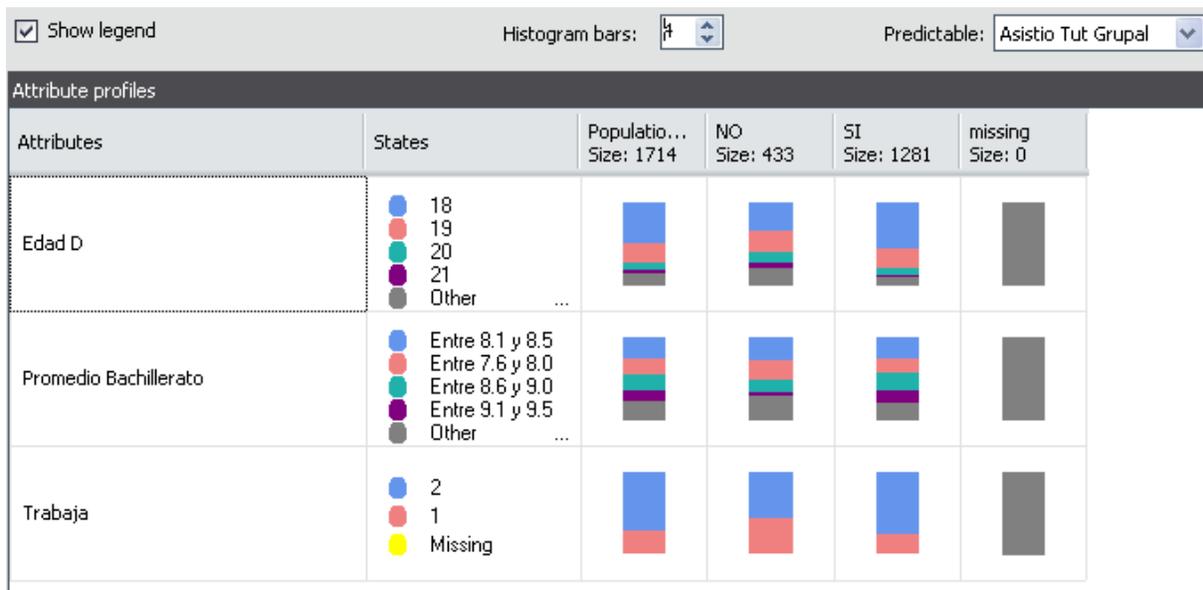


Figura 4.2.2.9.b Perfiles de los atributos

El algoritmo de *Bayes* muestra mediante la red de dependencia que atributos son los más significativos para aquellos que se intentaron predecir, este algoritmo encontró que la asistencia grupal depende de tres atributos solamente: si el alumno trabaja, su edad y su promedio en bachillerato, asemejándose de alguna manera a los anteriores, sin embargo este algoritmo no encontró ninguna dependencia para la asistencia individual y por consiguiente ningún perfil a analizar.

4.2.2.10 Evaluación e interpretación

Una vez analizados los anteriores algoritmos lo que resta es medir la precisión de cada uno de ellos, como los tres modelos pertenecen a una sola estructura es posible graficar la precisión de cada uno de ellos al intentar predecir un valor para el atributo, en este caso y por el análisis anterior el atributo que nos daría los datos más reales sería el de la asistencia a la tutoría grupal, además el grafico de elevación debe tener como valor destino el 'NO' de la asistencia grupal, esto porque para este análisis se desea que el modelo sea lo más preciso posible para detectar quienes no van a asistir y de esta manera enfocarse en ellos.

Al obtener el grafico de elevación obtenemos que:



Figura 4.2.2.10.a Grafico de elevación

Explicación del grafico de elevación

La línea recta diagonal, mostrada en azul, aparece en cada gráfico y representa los resultados de la estimación aleatoria y es la línea base con la que se evalúa la elevación, la línea verde por su parte representa el modelo ideal. [37]

En estos resultados se puede ver que, cuando se mide en el 44% de todos los casos (La línea gris vertical que cruza todas las demás líneas), el modelo que más puede predecir el comportamiento de asistencia es el de Clústeres, ya que logra capturar el 57.64% de la población de destino. En otras palabras si se predijera el comportamiento de los alumnos que no asistirán con solo un 44% de la población de la base de datos, el modelo sería capaz de detectar en un 57.64% de los alumnos potenciales de no asistir.

El valor de probabilidad de predicción representa el umbral necesario para incluir a un estudiante entre los casos "de posible no asistencia". Para cada caso, el modelo calcula la exactitud de cada predicción y almacena ese valor, que puede utilizar para filtrar o elegir estudiantes. Por ejemplo, para identificar los estudiantes que posiblemente no asistan a la tutoría con el modelo de Clústeres, se utilizaría una consulta para recuperar los casos con una probabilidad de predicción de al menos el 23.77%. Es interesante comparar los modelos. El modelo de Clústeres parece capturar más estudiantes potenciales, pero requiere una mayor probabilidad para incluir al alumno como predicción correcta.

El valor de puntuación ayuda a comparar los modelos calculando la efectividad del modelo a través de una población normalizada. Una mayor puntuación es mejor, de modo que en este caso podría decidir el modelo "Tabla Minable- Bayes" fue el que nos dio mejores resultados.

4.2.3 El índice académico

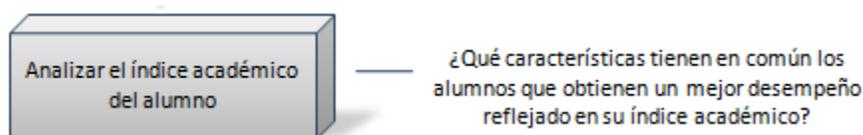


Figura 4.2.3.a Índice académico

4.2.3.1 Definición del problema

El índice académico (como se ha explicado anteriormente) es la referencia que se utiliza para asignar un número de inscripción a un alumno y cuya fórmula es:

$$IA = \text{Promedio del alumno} * \frac{\text{Asignaturas aprobadas}}{\text{Asignaturas cursadas}}$$

Este análisis está enfocado a encontrar las características que tienen en común aquellos alumnos que obtuvieron un índice académico satisfactorio de los que no.

4.2.3.2 Vista minable

```
create view TablaMinableRes2 as
-- Creando las tablas temporales para ser usadas mas tarde
WITH
amd AS (select * from (SELECT idAlumno,ROUND(avg(calificacion),2) as t1
FROM alumnomateriadiagnostico group by idAlumno) as t1),
ae AS (select * from (select i.idAlumno,e.nombre as escuela,r.descripcion as
PromedioBachillerato
from alumno i
inner join alumnoescuela ae on (i.idAlumno=ae.idAlumno)
inner join rango r on (ae.idRango=r.idRango)
inner join escuela e on (ae.idEscuela=e.idEscuela)
where nivel=3) as t2),
mr AS (SELECT idAlumno,count(*) as asignaturasAprobadas FROM alumnomateria a
where calificacion>=6
group by idAlumno),
mc AS (SELECT idAlumno,count(*) as asignaturasCursadas FROM alumnomateria a
group by idAlumno)
-- Seleccionar elementos y generar rango del indice académico
select *,
CASE WHEN indiceAcademico>=9 THEN '9-10'
```

CAPÍTULO 4. ANÁLISIS DE DATOS Y PRESENTACIÓN DE RESULTADOS

```
    WHEN indiceAcademico>=8 AND indiceAcademico<9 THEN '8-8.99'
    WHEN indiceAcademico>=7 AND indiceAcademico<8 THEN '7-7.99'
    WHEN indiceAcademico>=6 AND indiceAcademico<7 THEN '6-6.99'
    ELSE '0-5.99'
END as resultados
from (
-- Subconsulta
select
consulta_pivote.idAlumno,delegacion,promedioPrimerSemestre,promedioSegundoSemestre,promedioED,
escuela,promedioBachillerato,edad,
asignaturasCursadas,
asignaturasAprobadas,
cast(asignaturasAprobadas as float) / cast(asignaturasCursadas as float) as avanceAcademico,
round(promedioPrimerSemestre * (cast(asignaturasAprobadas as float) / cast(asignaturasCursadas
as float)),2) as indiceAcademico,
[¿Cuál es la razón principal por la que elegiste la carrera de ingeniería?] as razonPricipal,
CASE [¿Cuál es la razon principal por la qué elegiste la unam para estudiar ingeniería?] WHEN
' Por que aquí han estudiado tus familiares' THEN '1' ELSE '2' END as familiaresUNAM,
[Al terminar la carrera desearías:] as alTerminarCarrera,
CASE [Disfrutas en tu vivienda de computadora] WHEN 'SI' THEN '1' WHEN 'NO' THEN '2' END as
Computadora,
CASE [Disfrutas en tu vivienda de internet] WHEN 'SI' THEN '1' WHEN 'NO' THEN '2' END as
Internet,
CASE [Disfrutas en tu vivienda de automovil propio] WHEN 'SI' THEN '1' WHEN 'NO' THEN '2' END
as AutPropio,
CASE [Disfrutas en tu vivienda de automóvil familiar] WHEN 'SI' THEN '1' WHEN 'NO' THEN '2'
END as AutFamiliar,
CASE [¿Trabajas?] WHEN 'Sí, permanentemente' THEN '1' WHEN 'Si, eventualmente' THEN '1' WHEN
'No' THEN '2' END as Trabaja,
CASE [¿Cuántas horas trabajas en promedio, a la semana?] WHEN 'Menos de 6 horas' THEN '1-5'
WHEN 'Entre 6 y 12 horas' THEN '6-12' WHEN 'Entre 13 y 20 horas' THEN '13-20' WHEN 'Entre 21 y
25 horas' THEN '21-25' WHEN 'Entre 26 y 30 horas' THEN '26-30' WHEN 'Entre 31 y 35 horas' THEN
'31-35' WHEN 'Entre 36 y 40 horas' THEN '36-40' WHEN 'No aplica' THEN '0' END as HrsTrabajo,
[Respecto a tus padres :] as Padres,
[Señala cuál es el nivel máximo de estudios de tu padre:] as NivelPadre,
[Señala cuál es el nivel máximo de estudios de tu madre:] as NivelMadre,
[Tu orientación vocacional está:] as OrientacionVocacional,
[Considerando tu vida escolar, ¿cómo te calificarías como estudiante?] as ComoEstudiante,
CASE [En promedio, el tiempo al día, que usas computadora es:] WHEN 'Máximo una hora' THEN
'0-60' WHEN 'Entre una y dos horas' THEN '61-120' WHEN 'Entre dos y tres horas' THEN '121-180'
WHEN 'Entre dos y tres horas' THEN '121-180' WHEN 'Más de tres horas' THEN '180->' WHEN 'No
utilizas la computadora' THEN '0' END as UsoComputadora,
CASE [En promedio, el tiempo al día, que navegas en internet es:] WHEN 'Máximo media hora'
THEN '0-30' WHEN 'Entre media hora y una hora' THEN '31-60' WHEN 'Entre una y dos horas' THEN
'61-120' WHEN 'Entre dos y tres horas' THEN '121-180' WHEN 'Más de tres horas' THEN '180->'
WHEN 'No navegas en Internet' THEN '0' END as UsoIntenet,
[Señala cuál es la principal ocupación de tu padre:] as OcPadre,
[Señala cuál es la principal ocupación de tu madre:] as OcMadre,
[¿Principalmente, de quién dependes económicamente?] as DepEconomica,
[¿Cuántas personas contribuyen al sostenimiento de tu hogar?] as PersonasContribuyen,
CASE [¿A cuánto asciende el ingreso mensual total de tu hogar?] WHEN 'Menos de $3000' THEN
'$0-$3,000' WHEN 'Entre $3001 y $5000' THEN '$3001-$5000' WHEN 'Entre $5001 y $7000' THEN
'$5001-$7000' WHEN 'Entre $7001 y $9000' THEN '$7001-$9000' WHEN 'Entre $9001 y $11000' THEN
'$9001-$11000' WHEN 'Más de $11000' THEN '$11000-'END as IngMensual,
[¿Cómo te transportarás a la universidad?] as Transporte,
[¿Cuánto tiempo emplearas diariamente en transporte para ir y venir a la universidad?] as
TiempoTransporte,
CASE [¿A cuánto asciende el ingreso mensual total de tu hogar?] WHEN 'Menos de $3000' THEN 'E'
WHEN 'Entre $3001 y $5000' THEN 'D' WHEN 'Entre $5001 y $7000' THEN 'D' WHEN 'Entre $7001 y
$9000' THEN 'D+' WHEN 'Entre $9001 y $11000' THEN 'D+' WHEN 'Más de $11000' THEN 'C'END as
claseSocioeconomica
from (
-- Esta nueva subconsulta es la que realiza el pivote
SELECT
```

```

r.idAlumno as idAlumno,
p.nombre as nombre,
i.opcion as opcion,
d.nombre as delegacion,
ap.promedio as promedioPrimerSemestre,
ap2.promedio as promedioSegundoSemestre,
amd.tl as promedioED,
ae.escuela as escuela,
ae.PromedioBachillerato as promedioBachillerato,
YEAR(GETDATE()) - YEAR(fechaNac) as edad
FROM resultados r
inner join pregunta p on (r.idPregunta=p.idPregunta)
inner join incisos i on (r.idIncisos=i.idIncisos)
inner join alumno a on (r.idAlumno=a.idAlumno)
inner join delegacion d on (d.idDelegacion=a.idDelegacion)
inner join alumnopromedio ap on (ap.idAlumno=a.idAlumno)
inner join alumnopromedio ap2 on (ap2.idAlumno=a.idAlumno)
inner join amd on (amd.idAlumno=a.idAlumno)
inner join ae on (ae.idAlumno=a.idAlumno)
where ap.idSemestre = 1
) piv
PIVOT(
    Max(piv.opcion)
    FOR piv.nombre IN
    (
        [Disfrutas en tu vivienda de computadora],
        [Disfrutas en tu vivienda de internet],
        [Disfrutas en tu vivienda de automovil propio],
        [Disfrutas en tu vivienda de automóvil familiar],
        [¿Trabajas?],
        [¿Cuántas horas trabajas en promedio, a la semana?],
        [Respecto a tus padres :],
        [Señala cuál es el nivel máximo de estudios de tu padre:],
        [Señala cuál es el nivel máximo de estudios de tu madre:],
        [Tu orientación vocacional está:],
        [Considerando tu vida escolar, ¿cómo te calificarías como estudiante?],
        [En promedio, el tiempo al día, que usas computadora es:],
        [En promedio, el tiempo al día, que navegas en internet es:],
        [Señala cuál es la principal ocupación de tu padre:],
        [Señala cuál es la principal ocupación de tu madre:],
        [¿Principalmente, de quién dependes económicamente?],
        [¿Cuántas personas contribuyen al sostenimiento de tu hogar?],
        [¿A cuánto asciende el ingreso mensual total de tu hogar?],
        [¿Cómo te transportarás a la universidad?],
        [¿Cuánto tiempo emplearás diariamente en transporte para ir y venir a la
universidad?],
        [¿Cuál es la razón principal por la que elegiste la carrera de ingeniería?],
        [¿Cuál es la razon principal por la que elegiste la unam para estudiar ingeniería?],
        [Al terminar la carrera desearías:]
    )
) as consulta_pivote
inner join mc on (mc.idAlumno=consulta_pivote.idAlumno)
left join mr on (mr.idAlumno=mc.idAlumno) as tbl2

```

4.2.3.3 Elección del algoritmo

Debido a la gran cantidad de datos los algoritmos que mejor podrían resolver el problema son: el algoritmo de Naive y el algoritmo de Clústeres, sin embargo el algoritmo elegido será el primero ya que este arroja mejores resultados para el primer análisis, sumado a lo anterior el algoritmo de Naive proporciona una red de dependencia que puede resultar muy útil para encontrar los principales factores que afectan al índice académico.

4.2.3.4 Minería de datos

Características principales del modelo:

Estructura a la que pertenece: Tabla Minable Res2

Modelo: Tabla Minable Res2 – Naive Bayes

Objetivo del modelo: Obtener las principales características de los alumnos que obtuvieron un mejor índice académico, así como las características de los alumnos que no lo obtuvieron.

Atributos a predecir: Resultados

Structure ^A	Tabla Minable Res2 - Naive	Descripción de los datos a procesar
	Microsoft_Naive_Bayes	
Al Terminar Carrera	Input	Indica las aspiraciones del alumno al terminar la carrera
Aut Familiar	Input	Indica si el alumno tiene auto familiar [1] = Si, [2] = No
Aut Propio	Input	Indica si el alumno tiene auto propio [1] = Si, [2] = No
Clase Socioeconomica	Input	Indica la clase socioeconómica del alumno
Como Estudiante	Input	Indica cómo se considera el alumno
Computadora	Input	Indica si el alumno tiene computadora [1] = Si, [2] = No
Delegacion	Input	Indica la delegación donde reside actualmente el alumno
Dep Economica	Input	Indica de quien depende económica el alumno
Edad	Input	Indica la edad del alumno
Escuela	Input	Indica la escuela a que asistió el alumno en preparatoria
Familiares UNAM	Input	Indica si el alumno tiene familiares que estudiaron en la UNAM [1] = Si, [2] = No
Hrs Trabajo	Input	Indica las horas que trabaja el alumno
Id Alumno	Key	Id del alumno
Ing Mensual	Input	Rango de ingresos mensuales de todos los integrantes en el hogar del alumno
Internet	Input	Indica si el alumno tiene internet en su hogar [1] = Si, [2] = No
Nivel Madre	Input	Indica el nivel de estudios de la madre del alumno
Nivel Padre	Input	Indica el nivel de estudios del padre del alumno
Oc Madre	Input	Indica la ocupación de la madre del alumno
Oc Padre	Input	Indica la ocupación del padre del alumno
Orientacion Vocacional	Input	Indica la orientación vocacional del alumno
Padres	Input	Indica el estado civil de los padres del alumno
Personas Contribuyen	Input	Indica cuantas personas contribuyen económicamente en el hogar del alumno
Promedio Bachillerato	Input	Indica el rango del promedio en bachillerato del alumno
Promedio ED	Input	Indica el promedio del alumno en el examen diagnostico
Razon Prcipal	Input	Indica la razón principal de que el alumno este estudiando su carrera
Resultados	PredictOnly	Clasificación del resultado del alumno (Índice académico ²)
Tiempo Transporte	Input	Indica el tiempo de transporte desde el hogar del alumno hasta la facultad
Trabaja	Input	Indica si el alumno trabaja o no [1] = Si, [2] = No
Transporte	Input	Indica el tipo de transporte que utiliza el alumno para llegara a la facultad
Uso Computadora	Input	Indica el rango de horas al día que el alumno utiliza la computadora
Uso Internet	Input	Indica el rango de horas al día que el alumno utiliza el internet

² El índice académico es la referencia en la cual es basado el número de inscripción, su fórmula es:

$$IA = Promedio\ del\ alumno * \frac{Asignaturas\ aprobadas}{Asignaturas\ cursadas}$$

Formula proporcionada por el Prof. José Enrique Larios Canale (Codirector de Tesis)

4.2.3.5 Patrones

Una vez procesado el modelo con el algoritmo seleccionado la regla de dependencia obtenida es:

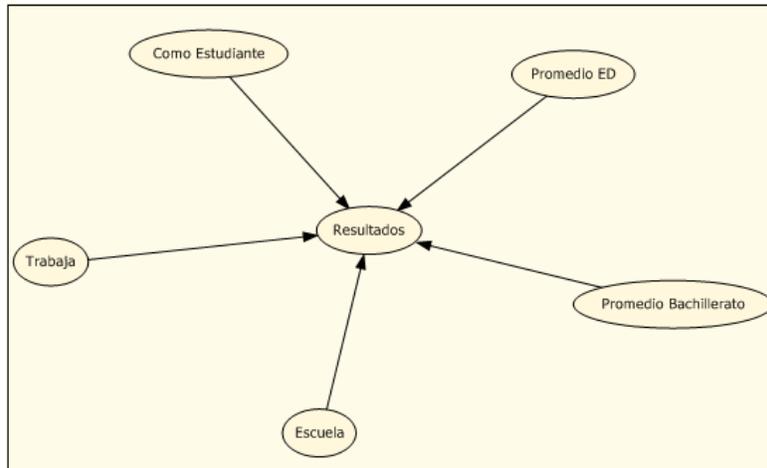


Figura 4.2.3.5.a Regla de dependencia

En ella podemos observar que los factores que más influyen en el índice académico son el promedio en el examen diagnóstico, el promedio en el bachillerato, si el alumno trabaja o no, de que escuela (a nivel bachillerato) viene el alumno y como se considera él como estudiante.

En adición el procesamiento del modelo permite realizar una discriminación entre atributos, donde es posible comparar los extremos del índice académico para de esta manera conocer las principales diferencias entre un óptimo índice académico y uno que muestra un pésimo desempeño académico.

Attribute:	Value 1:	Value 2:	Discrimination scores for 0-5.99 and 9-10	
Resultados	0-5.99	9-10	Favors 0-5.99	Favors 9-10
Promedio ED	>= 6,6537200128			██████████
Promedio Bachillerato	Entre 9.1 y 9.5			██████████
Escuela	Colegio de Ciencias y Humanidades de la ...		██████████	
Promedio Bachillerato	Entre 9.6 y 10.0			██████████
Promedio ED	< 2,9550656096		██████████	
Promedio ED	2,9550656096 - 3,926806116		██████████	
Promedio Bachillerato	Entre 8.1 y 8.5		██████████	
Como Estudiante	Bueno (a)		██████████	
Como Estudiante	Excelente			██████████
Promedio ED	5,0863390328 - 6,6537200128			██████████
Como Estudiante	Muy bueno (a)			██████████
Escuela	Institución Pública fuera de la ZMCM* (Col...			██████████
Promedio Bachillerato	Entre 7.1 y 7.5		██████████	
Como Estudiante	Regular		██████████	
Promedio Bachillerato	Entre 7.6 y 8.0		██████████	
Trabaja	2			██████████
Trabaja	1		██████████	
Escuela	Institución Privada en la ZMCM*			██████████
Escuela	Institución en el extranjero			██████████
Promedio Bachillerato	No lo sabe			██████████
Escuela	Institución Privada fuera de la ZMCM*			██████████
Escuela	Otra			██████████

Figura 4.2.3.5.b Discriminación

4.2.3.6 Evaluación e interpretación

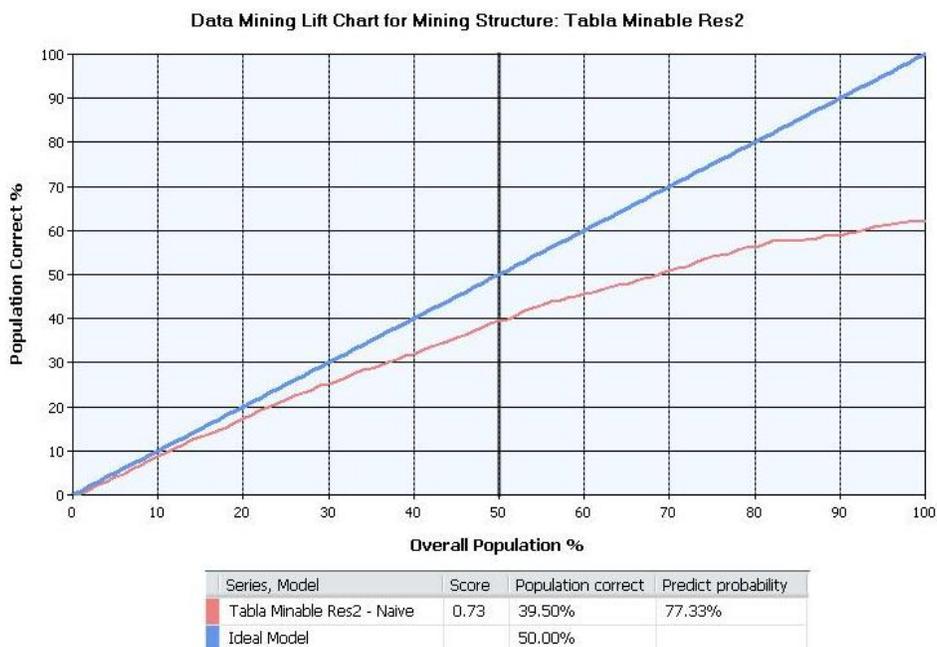


Figura 4.2.3.6.a Grafico de elevación

La gráfica de elevación muestra que el modelo tiene una precisión considerada buena lo que permite saber que el modelo cumple para el objetivo propuesto.

Una vez que se han realizado los tres análisis (impacto, asistencia e índice académico) es posible llegar a una serie de conclusiones:

- En el primer análisis se comprobó que aquellos alumnos que asistieron a la tutoría grupal tuvieron un mejor índice académico que aquellos que no lo hicieron. Sin embargo respecto a la tutoría individual no pudo ser obtenida información relevante, debido a que pocos estudiantes asisten a dicha modalidad de la tutoría.
- En el segundo análisis se llegó a la conclusión de que los alumnos que asisten a la tutoría grupal tienen como características principales: la edad del alumno, si este trabaja al mismo tiempo que realiza sus estudios y finalmente su promedio en nivel bachillerato.
- Por último el tercer análisis arrojó que las características que tienen en común los alumnos que tienen un mejor índice académico de los que no (Sin importar si asistieron a la tutoría o no) fueron: cómo se considera como estudiante, su promedio en examen diagnóstico, si trabaja, su promedio en el bachillerato y la escuela en la que curso dicho nivel.

Lo anterior muestra en primera instancia que la decisión que se ha de tomar tiene que ser sobre el grupo que tenga las características semejantes entre el segundo y tercer análisis que fueron: el trabajo y el promedio en bachillerato. Pero tomar una decisión sobre aquellos alumnos que trabajen puede ser algo delicado, con base en esto es recomendable hacer el análisis siguiente extra, el cual permite demostrar que aquellos alumnos que trabajan y que pese a esto asistieron a la tutoría tuvieron un mejor índice académico.

4.2.4 Los que trabajan

4.2.4.1 Definición del problema

Este análisis está enfocado a apoyar la toma de decisión que será tomada posteriormente, la finalidad del mismo es demostrar que aquellos alumnos que realizan algún trabajo y aún con esta limitante (en cuestión de tiempo) asisten a las tutorías obtienen un mejor desempeño académico reflejado directamente en su índice académico.

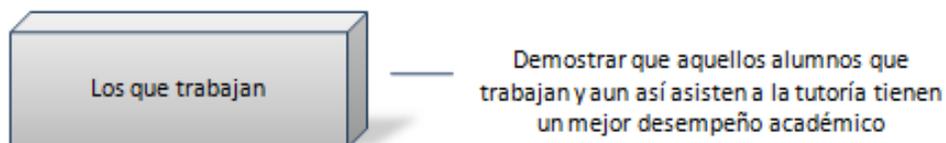


Figura 4.2.4.1.a Los que trabajan

4.2.4.2 Vista minable

La vista minable en cuestión, es una variación de las anteriores, simplemente con menos datos seleccionados.

Nota. También era posible utilizar los filtros que SQL Server proporciona para los queries o colocar los atributos que no se necesitan como ignorados (propiedad previa al proceso). Sin embargo para separar las estructuras y facilitar su comprensión en la plataforma se decidió hacer una nueva vista al igual que una nueva estructura (con su modelo correspondiente).

```

create view TablaMinableResVariacion as
-- Creando las tablas temporales para ser usadas mas tarde
WITH mc AS (SELECT idAlumno, count(*) as asignaturasCursadas FROM alumnomateria a
             group by idAlumno),
     mr AS (SELECT idAlumno, count(*) as asignaturasAprobadas FROM alumnomateria a
             where calificacion>=6
             group by idAlumno),
     ae AS(select * from (select i.idAlumno,e.nombre as escuela,r.descripcion as
                             PromedioBachillerato from alumno i
                             inner join alumnosescuela ae on (i.idAlumno=ae.idAlumno)
                             inner join rango r on (ae.idRango=r.idRango)
                             inner join escuela e on (ae.idEscuela=e.idEscuela)
                             where nivel=3) as t2)
-- Seleccionar elementos y generar rango del indice académico
select *,
CASE WHEN indiceAcademico>=9 THEN '9-10'
      WHEN indiceAcademico>=8 AND indiceAcademico<9 THEN '8-8.99'
      WHEN indiceAcademico>=7 AND indiceAcademico<8 THEN '7-7.99'
      WHEN indiceAcademico>=6 AND indiceAcademico<7 THEN '6-6.99'
      ELSE '0-5.99'
END as resultados
from (
-- Subconsulta
select idAlumno,
promedioPrimerSemestre,
asignaturasAprobadas,
asignaturasCursadas,
cast(asignaturasAprobadas as float) / cast(asignaturasCursadas as float) as avanceAcademico,
round(promedioPrimerSemestre * (cast(asignaturasAprobadas as float) / cast(asignaturasCursadas
as float)),2) as indiceAcademico,

```

```

CASE [¿A cuántas sesiones de tutoría grupal asististe?] WHEN '1' THEN 'SI' WHEN '2' THEN 'SI'
WHEN '3' THEN 'SI' WHEN '4' THEN 'SI' WHEN '5' THEN 'SI' WHEN '6' THEN 'SI' WHEN '7' THEN 'SI'
WHEN '8' THEN 'SI' WHEN 'Mas de 8' THEN 'SI' ELSE 'NO' END as AsistioTutGrupal,
CASE [¿A cuántas sesiones de tutoría individual asististe?] WHEN '1' THEN 'SI' WHEN '2' THEN
'SI' WHEN '3' THEN 'SI' WHEN '4' THEN 'SI' WHEN '5' THEN 'SI' WHEN '6' THEN 'SI' WHEN '7' THEN
'SI' WHEN '8' THEN 'SI' WHEN 'Mas de 8' THEN 'SI' ELSE 'NO' END as AsistioTutIndividual,
CASE [¿Trabajas?] WHEN 'Sí, permanentemente' THEN '1' WHEN 'Sí, eventualmente' THEN '1' WHEN
'No' THEN '2' END as Trabaja,
promedioBachillerato,
edad
from (
-- Esta nueva subconsulta es la que realiza el pivote
SELECT
r.idAlumno as idAlumno,
p.nombre as nombre,
i.opcion as opcion,
d.nombre as delegacion,
ap.promedio as promedioPrimerSemestre,
mc.asignaturasCursadas AS asignaturasCursadas,
CASE WHEN mr.asignaturasAprobadas IS NULL THEN '0' Else mr.asignaturasAprobadas END As
asignaturasAprobadas,
2009 - YEAR(fechaNac) as edad,
idCarrera as carrera,
ae.PromedioBachillerato as promedioBachillerato
FROM resultados r
inner join pregunta p on (r.idPregunta=p.idPregunta)
inner join incisos i on (r.idIncisos=i.idIncisos)
inner join alumno a on (r.idAlumno=a.idAlumno)
inner join delegacion d on (d.idDelegacion=a.idDelegacion)
inner join alumnopromedio ap on (ap.idAlumno=a.idAlumno)
inner join mc on (mc.idAlumno=a.idAlumno)
left join mr on mr.idAlumno=mc.idAlumno
inner join ae on (ae.idAlumno=a.idAlumno)
where ap.idSemestre = 1
) piv
PIVOT(
Max(piv.opcion)
FOR piv.nombre IN
(
[¿A cuántas sesiones de tutoría grupal asististe?],
[¿A cuántas sesiones de tutoría individual asististe?],
[¿Trabajas?]
)
) as chld) as tbl
GO

```

Nótese que a diferencia de la consulta del análisis anterior, esta vez se incluyen el número de sesiones a las que se asistió, lo anterior para obtener la influencia que tuvo el que un alumno trabajara sobre las asistencias a la tutoría.

4.2.4.3 Elección del algoritmo

Para este análisis se utilizará el algoritmo de redes neuronales, ya que este algoritmo permite dar una entrada en el modelo de manera que se pueda conocer el valor de las salidas seleccionadas, la entrada para este caso será la característica de que el alumno tiene trabajo (Trabaja = 1), y las salidas el promedio del primer semestre y los resultados (índice académico).

4.2.4.4 Minería de datos

Características principales del modelo:

Estructura a la que pertenece: Tabla Minable Res Variación

Modelo: Tabla Minable Res Variación –Neural Network

Objetivo del modelo:

Atributos a predecir: Promedio Primer Semestre, Resultados

Structure	Tabla Minable Res Variacion - Neural Network
	Microsoft_Neural_Network
Asistio Tut Grupal	Input
Asistio Tut Individual	Input
Id Alumno	Key
Promedio Primer Semestre	PredictOnly
Resultados	PredictOnly
Trabaja	Input

Descripción de los datos a procesar

Indica si el alumno asistió al menos a una tutoría grupal
 Indica si el alumno asistió al menos a una tutoría individual
 Id del alumno
 Promedio del alumno durante su primer semestre
 Clasificación del resultado del alumno (Índice académico)
 Indica si el alumno trabaja o no [1] = Si, [2] = No

4.2.4.5 Patrones

Colocando el atributo de “Trabaja” en 1 para indicar que el filtro sea sobre los alumnos que trabajan podemos obtener que el promedio del alumno fue:

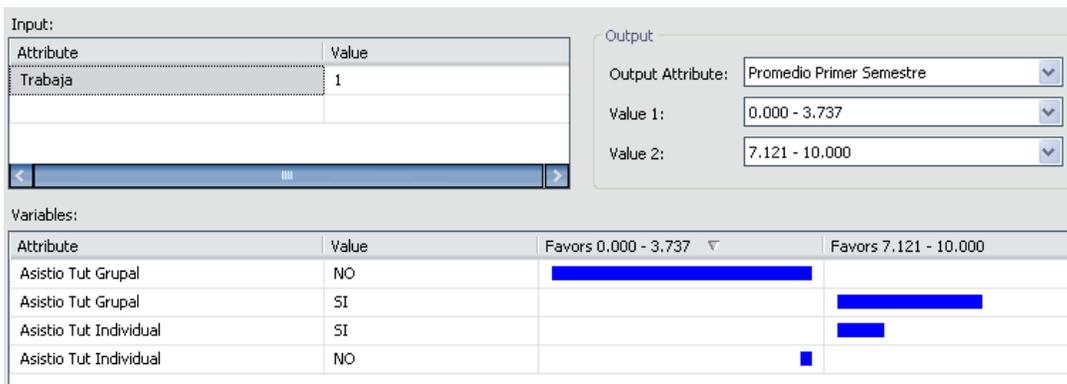


Figura 4.2.4.5.a Resultados para el promedio primer semestre

De la misma forma para los resultados del alumno se puede obtener que:

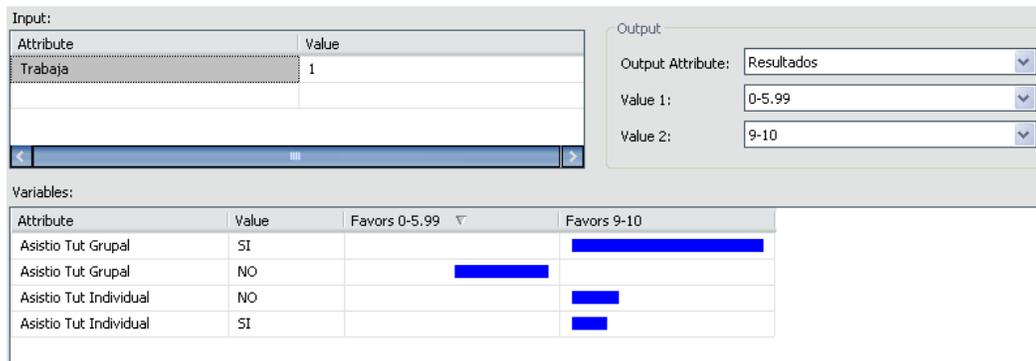


Figura 4.2.4.5.b Resultados para el índice académico

4.2.4.6 Evaluación e interpretación

Los patrones obtenidos permiten observar que tanto el promedio de los alumnos como sus índices académicos fueron mejores para aquellos alumnos que pese a estar trabajando asistieron a las tutorías grupales.

Con el grafico de elevación es posible observar que el análisis es bastante preciso:

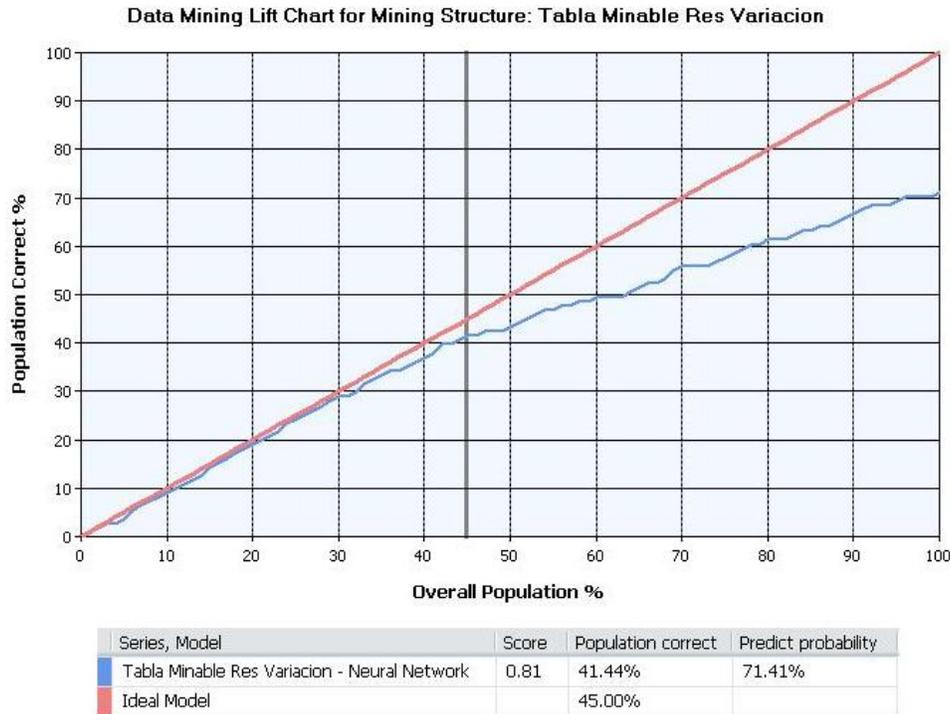


Figura 4.2.4.6.a Grafico de elevación

De igual forma podríamos generar una matriz de clasificación para un atributo específico como la que se muestra en la siguiente figura:

Counts for Tabla Minable Res Variacion - Neural Network on Resultados:

Predicted	0-5.99 (Actual)	6-6.99 (Actual)	7-7.99 (Actual)	8-8.99 (Actual)	9-10 (Actual)
0-5.99	79	11	6	9	6

Figura 4.2.4.6.b Matriz de clasificación

Lo anterior muestra que el porcentaje de alumnos que fueron seleccionados como prueba fueron en su mayor parte predichos correctamente.

4.2.4.7 Conocimiento

Con base en todos los modelos procesados para este análisis se pudieron obtener las características principales que poseen los alumnos que asisten a las tutorías, así como aquellas de los que no lo hacen, se conoció además el hecho de que dichas características son compartidas para el índice académico: trabajo y promedio del

bachillerato, finalmente se mostro que el impacto que tiene la tutoría sobre el índice académico de los estudiantes es muy alta, incluidos aquellos que trabajan y aun con esto se comprometen a asistir a dicha tutoría.

4.2.4.8 Toma de decisiones

Con todos los modelos procesados para este primer análisis y como se mencionó anteriormente la decisión que se ha de tomar estará enfocada en los alumnos que compartían características similares tanto para el segundo como tercer análisis (asistencia e índice académico), dichas características similares fueron el promedio de bachillerato y si el alumno trabaja o no, siendo para el primero que entre más bajo más posible es que el alumno no asista a la tutoría, de igual forma ocurre si el alumno trabaja, además el último análisis (los que trabajan) ha permitido asegurar que aquellos que asistieron a las tutorías aun cuando tenían algún tipo de trabajo tuvieron un índice académico mucho mejor que aquellos que no lo hicieron.

El promedio de bachillerato como factor en la asistencia a la tutoría puede deberse a que aquellos alumnos que lograron un mejor desempeño académico en el nivel bachillerato (reflejado en su promedio escolar) intentan mantenerlo, lo cual al enfrentarse a un nuevo nivel escolar y ante la posible “amenaza” de disminuirlo asisten a la tutoría con el objetivo de mejorar sus hábitos de estudio o bien adecuarlos a este nuevo nivel.

El factor trabajo no requiere mayor explicación que mencionar que aquellos alumnos que trabajan asisten menos que los que no lo hacen debido a que tienen el tiempo más limitado. Aunque existen otros factores obtenidos independientemente por cada modelo, los dos antes mencionados fueron referentes en cada uno de ellos, es por esta razón que la decisión a tomar solo ha de considerar estos dos aspectos, donde para un promedio menor a 8.5 (el promedio más alto en el que comienzan a haber resultados negativos) y siendo que el alumno trabaje es necesario aplicar una toma de decisión.

A continuación y con base en todo lo expuesto anteriormente se propondrán una adecuación y una alternativa que intentan aumentar el promedio de asistencia a la tutoría y con ello mejorar el índice académico:

Adecuar la asignación estudiante/tutor

Actualmente la asignación estudiante/tutor se lleva acabo de la siguiente manera:

1. Se integran grupos o bloques de alumnos en la División de Ciencias Básicas, con las diferentes materias que van a cursar en el primer semestre.
2. En esos bloques se les determina un horario para la tutoría.
3. La Coordinación de Programas de Atención Diferenciada para Alumnos (COPADI) solicita previamente, a los Coordinadores de Tutores, la lista y horarios de los profesores de su carrera que van a participar en el programa durante ese periodo.
4. Con estos datos la COPADI inicia la integración de los tutores en los bloques, tratando de que coincidan principalmente con la carrera de los alumnos.
5. Finalmente aparece el nombre del tutor, el salón y el horario de la tutoría, en la tira de materias de los alumnos, igual que todas sus materias a cursar en el semestre.

Como se puede observar la asignación de los tutores considera el aspecto académico del estudiante, pero en

ningún momento algún aspecto personal, y particularizando para el análisis realizado no considera si el alumno trabaja.

Dicho lo anterior la adecuación envolvería tanto el tercer como el cuarto paso:

3. La Coordinación de Programas de Atención Diferenciada para Alumnos (COPADI) solicita previamente, a los Coordinadores de Tutores, la lista y horarios de los profesores de su carrera que van a participar en el programa durante ese periodo, de igual forma solicita específicamente el dato de la encuesta socioeconómica para saber si cada estudiante de cada bloque trabaja o no lo hace.
4. Con esos datos la COPADI inicia la integración de los tutores en los bloques, tratando de que coincidan principalmente con la carrera de los alumnos y asignando a los tutores de tiempo completo³ a aquellos alumnos que trabajan.

Con esto se pretende que los alumnos puedan tener mayor flexibilidad para asistir a las tutorías ya que el tutor dispondrá de mayor tiempo para atender sus necesidades.

Aumentar la tutoría electrónica

La tutoría electrónica ofrece oportunidades únicas para comprometer al estudiante. Muchas de las estrategias provenientes de la enseñanza directa pueden utilizarse en un sistema on-line; por ejemplo: preguntas autoformuladas, uso de la información (links), formulación y presentación de situaciones adecuadas, favorecimiento de la comprensión, recomendaciones para la consulta y lectura posterior. [38]

El uso de los mediadores electrónicos cambia la naturaleza de la interacción tutor-estudiante. El manejo de los instrumentos tecnológicos y la confianza en su uso por parte de ambos miembros es un requisito frecuentemente señalado en la literatura especializada. El estilo conversacional de la interacción online provee un modo diferente de generación de conocimiento y habilidades. [38]

A continuación se muestra una tabla de beneficios potenciales contra dificultades potenciales que se pueden encontrar dentro de la tutoría electrónica.

Beneficios potenciales	Dificultades potenciales
El e-mail brinda al tutor y al estudiante tiempo para medir acerca de una pregunta o texto.	La naturaleza asincrónica del e-mail puede reducir la "negociación" entre los participantes..
Algunos estudiantes prefieren la comunicación escrita. Cuando se trata de opiniones, el "anonimato" es mayor en e-mail	Los estudiantes que prefieren la comunicación oral, pueden no sentirse cómodos con el e-mail
Las horas de trabajo pueden ser reguladas según las necesidades del tutor y del estudiante.	Los estudiantes pueden no sentirse obligados a responder cuando no obtienen la respuesta que esperaban.
Guía para búsqueda en el hiperespacio de información disponible.	Desplazamiento para controlar actividades y mensajes poco pertinentes.
Apoyo a la resolución interactiva de problemas.	Falta de familiaridad con el medio y la tarea.

Figura 4.2.4.8.a Beneficios vs Dificultades [38]

³ La tutoría cuenta con profesores tanto de tiempo como de tiempo parcial.

La tutoría electrónica permite utilizar la experiencia adquirida en la tutoría directa, centrar la acción en los modos de respuesta del estudiante, dirigir la atención hacia la información pertinente a través de links, hacer consiente la importancia del auto monitoreo y el desarrollo de competencias críticas sobre el propio progreso. Emplea tecnología simple y accesible, utiliza más efectivamente el tiempo del alumno y permite llegar a estudiantes geográficamente lejanos con limitado tiempo para asistir a clase, así como aquellos que por alguna razón se les dificulta estar presencialmente en ellas (como el caso de los alumnos que trabajan y no pueden desperdiciar tiempo estar viajando hacia las tutorías).

Aunque es indiscutible que la tutoría electrónica se ocupa hoy día, en muchas ocasiones no se ocupa como debería de ser, y es este aspecto en el que el tutor debe “educar” al estudiante mostrándole los beneficios que pueden conseguir de hacerlo adecuadamente. Si el tutor logra convencer al estudiante a través de la primera tutoría (presencial) a mantener una comunicación constante de forma electrónica, aquellos cuya dificultad para asistir presencialmente a las mismas podrán encontrar una manera de resolver sus necesidades.

4.3 Segundo análisis: Los tutores



Figura 4.3.a Los tutores

4.3.1 Definición del problema

Como se ha venido exponiendo con anterioridad un factor importante para el funcionamiento de la tutoría es sin duda alguna la confianza que logre establecer el tutor con el alumno en las sesiones que se lleven a cabo, para lograr dicha confianza el tutor debe poseer ciertas características, entra las que se encuentran:

- La disposición para atender a los alumnos.
- El interés en los problemas de los alumnos.
- La comunicación permanente con el alumno.
- Saber cómo establecer los objetivos con los alumnos.
- La forma de canalizar las dudas de los alumnos
- La planeación de las actividades con los alumnos.
- La confianza que inspire en los alumnos.
- La puntualidad del tutor en las tutorías.
- El seguimiento sistemático de las actividades

- La orientación necesario para que el alumno atienda sus necesidades de orden personal
- El estímulo para que el alumno desarrolle habilidades para estudiar

Las características antes mencionadas permitirán sin duda alguna que el alumno se sienta más identificado con el tutor, con lo cual se obtendrán mejores índices de aprovechamiento de la tutoría, esto es habrá una mayor asistencia grupal e individual, mejorara tanto el desempeño académico del alumno así como la integración del mismo a la Facultad aumentando su interés por estudiar la carrera de ingeniería y su satisfacción con el programa de tutoría.

El siguiente análisis a realizar será enfocado precisamente a encontrar cuales características de los tutores fueron las que más influyeron (de manera positiva) en los alumnos, el cual expondrá por si solo argumentos para la toma de decisiones enfocadas principalmente a los resultados.

4.3.2 Vista minable

Para este análisis se utilizara una vista minable que muestre los resultados de las preguntas enfocadas a la satisfacción, integración, comunicación etc. de la encuesta de tutoría.

```
create view TablaMinableTutores as

WITH amd AS (select * from
              (SELECT idAlumno,ROUND(avg(calificacion),2) as t1 FROM alumnomateriadiagnostico
              group by idAlumno) as t1),
ae AS (select * from
       (select i.idAlumno,e.nombre as escuela,r.descripcion as PromedioBachillerato from
        alumno i
        inner join alumnosescuela ae on (i.idAlumno=ae.idAlumno)
        inner join rango r on (ae.idRango=r.idRango)
        inner join escuela e on (ae.idEscuela=e.idEscuela)
        where nivel=3) as t2)

select idAlumno,
CASE [¿A cuántas sesiones de tutoría grupal asististe?] WHEN '1' THEN '1' WHEN '2' THEN '2'
WHEN '3' THEN '3' WHEN '4' THEN '4' WHEN '5' THEN '5' WHEN '6' THEN '6' WHEN '7' THEN '7' WHEN
'8' THEN '8' WHEN 'Mas de 8' THEN '9' ELSE '0' END as NumAsistenciasG,
CASE [¿A cuántas sesiones de tutoría individual asististe?] WHEN '1' THEN '1' WHEN '2' THEN
'2' WHEN '3' THEN '3' WHEN '4' THEN '4' WHEN '5' THEN '5' WHEN '6' THEN '6' WHEN '7' THEN '7'
WHEN '8' THEN '8' WHEN 'Mas de 8' THEN '9' ELSE '0' END as NumAsistenciasI,
CASE [Muestra el tutor buena disposición el tutor para atender a los alumnos] WHEN 'Poca' THEN
1 WHEN 'Mediana' THEN 2 WHEN 'Regular' THEN 3 WHEN 'Suficiente' THEN 4 WHEN 'mucha' THEN 5
END as disposicionTutor,
CASE [Muestra el tutor interés en los problemas académicos y personales que afectan el
rendimiento de los alumnos] WHEN 'Poca' THEN 1 WHEN 'Mediana' THEN 2 WHEN 'Regular' THEN 3
WHEN 'Suficiente' THEN 4 WHEN 'mucha' THEN 5 END as interesTutor,
CASE [Muestra el tutor disposición a mantener una comunicación permanente con el alumno] WHEN
'Poca' THEN 1 WHEN 'Mediana' THEN 2 WHEN 'Regular' THEN 3 WHEN 'Suficiente' THEN 4 WHEN
'mucha' THEN 5 END as comunicacionTutor,
CASE [Tu tutor establecio los objetivos de la tutoría] WHEN 'Poca' THEN 1 WHEN 'Mediana' THEN
2 WHEN 'Regular' THEN 3 WHEN 'Suficiente' THEN 4 WHEN 'mucha' THEN 5 END as objetivosTutor,
CASE [Las dudas sobre tu desempeño académico expresadas en las sesiones de tutoría, fueron
aclaradas o canalizadas acetadamente] WHEN 'Poca' THEN 1 WHEN 'Mediana' THEN 2 WHEN 'Regular'
THEN 3 WHEN 'Suficiente' THEN 4 WHEN 'mucha' THEN 5 END as dudasCanalizadas,
CASE [El trabajo que desarrollaste con el tutor evidenció que hubo una planeación de las
actividades y no una mera improvisacion] WHEN 'Poca' THEN 1 WHEN 'Mediana' THEN 2 WHEN
'Regular' THEN 3 WHEN 'Suficiente' THEN 4 WHEN 'mucha' THEN 5 END as planeacionTutor,
CASE [El clima propiciado por el tutor te dio la confianza para expresarle tus dudas] WHEN
'Poca' THEN 1 WHEN 'Mediana' THEN 2 WHEN 'Regular' THEN 3 WHEN 'Suficiente' THEN 4 WHEN
```

```

'mucha' THEN 5 END as ConfianzaTutor,
CASE [Asistió el tutor puntualmente a las sesiones] WHEN 'Poca' THEN 1 WHEN 'Mediana' THEN 2
WHEN 'Regular' THEN 3 WHEN 'Suficiente' THEN 4 WHEN 'mucha' THEN 5 END as puntualidadTutor,
CASE [Tu tutor mantuvo seguimiento sistemático de las actividades y acuerdos establecidos]
WHEN 'Poca' THEN 1 WHEN 'Mediana' THEN 2 WHEN 'Regular' THEN 3 WHEN 'Suficiente' THEN 4 WHEN
'mucha' THEN 5 END as seguimientoTutor,
CASE [Tu tutor te proporcionó la orientación necesaria para atender tus necesidades de orden
personal y de salud] WHEN 'Poca' THEN 1 WHEN 'Mediana' THEN 2 WHEN 'Regular' THEN 3 WHEN
'Suficiente' THEN 4 WHEN 'mucha' THEN 5 END as orientacionTutor,
CASE [Te estimuló para que desarrollaras habilidades para estudiar de manera independiente]
WHEN 'Poca' THEN 1 WHEN 'Mediana' THEN 2 WHEN 'Regular' THEN 3 WHEN 'Suficiente' THEN 4 WHEN
'mucha' THEN 5 END as estimuloTutor,
CASE [Tu participación en las sesiones de tutoría ha mejorado tu desempeño académico] WHEN
'Poca' THEN 1 WHEN 'Mediana' THEN 2 WHEN 'Regular' THEN 3 WHEN 'Suficiente' THEN 4 WHEN
'mucha' THEN 5 END as mejoroDesempeñoAlumno,
CASE [Tu integración a la Facultad ha mejorado con el programa de tutoría] WHEN 'Poca' THEN 1
WHEN 'Mediana' THEN 2 WHEN 'Regular' THEN 3 WHEN 'Suficiente' THEN 4 WHEN 'mucha' THEN 5 END
as integracion,
CASE [Te es satisfactorio el programa de tutoría] WHEN 'Poca' THEN 1 WHEN 'Mediana' THEN 2
WHEN 'Regular' THEN 3 WHEN 'Suficiente' THEN 4 WHEN 'mucha' THEN 5 END as satisfaccion,
CASE [En este momento tu interés por estudiar la carrera de ingeniería] WHEN 'Se ha
incrementado' THEN 3 WHEN 'Se ha mantenido' THEN 2 WHEN 'Se ha disminuido' THEN 1 END as
interesesEstudios
from (
    SELECT
    r.idAlumno as idAlumno,
    p.nombre as nombre,
    i.opcion as opcion
    FROM resultados r
    inner join pregunta p on (r.idPregunta=p.idPregunta)
    inner join incisos i on (r.idIncisos=i.idIncisos)
    inner join alumno a on (r.idAlumno=a.idAlumno)
) piv
PIVOT(
    Max(piv.opcion)
    FOR piv.nombre IN
    (
        [Muestra el tutor buena disposición el tutor para atender a los alumnos],
        [Muestra el tutor interés en los problemas académicos y personales que afectan el
rendimiento de los alumnos],
        [Muestra el tutor disposición a mantener una comunicación permanente con el alumno],
        [Tu tutor estableció los objetivos de la tutoría],
        [Las dudas sobre tu desempeño académico expresadas en las sesiones de tutoría, fueron
aclaradas o canalizadas acetadamente],
        [El trabajo que desarrollaste con el tutor evidenció que hubo una planeación de las
actividades y no una mera improvisacion],
        [El clima propiciado por el tutor te dio la confianza para expresarle tus dudas],
        [Asistió el tutor puntualmente a las sesiones],
        [Tu tutor mantuvo seguimiento sistemático de las actividades y acuerdos
establecidos],
        [Tu tutor te proporcionó la orientación necesaria para atender tus necesidades de
orden personal y de salud],
        [Te estimuló para que desarrollaras habilidades para estudiar de manera
independiente],
        [Tu participación en las sesiones de tutoría ha mejorado tu desempeño académico],
        [¿A cuántas sesiones de tutoría grupal asististe?],
        [¿A cuántas sesiones de tutoría individual asististe?],
        [Tu integración a la Facultad ha mejorado con el programa de tutoría],
        [Te es satisfactorio el programa de tutoría],
        [En este momento tu interés por estudiar la carrera de ingeniería]
    )
) as chld
where [¿A cuántas sesiones de tutoría individual asististe?]!='Ninguna'
or [¿A cuántas sesiones de tutoría grupal asististe?]!='Ninguna'

```

4.3.3 Elección del algoritmo

Para este análisis se utilizará el modelo de regresión lineal, el cual permite trabajar con valores continuos, los cuales han sido establecidos en la vista minable. La idea principal de este algoritmo es determinar una relación entre dos columnas continuas, de manera que conforme una aumenta la otra se comporta de manera similar, lo anterior es ideal para este análisis ya que se busca encontrar cuáles características que posee un tutor entre más grandes mayores resultados obtendrá y por consiguiente más efectivos.

4.3.4 Minería de datos

Características principales del modelo:

Estructura a la que pertenece: Tabla Minable Tutores

Modelo: Tabla Minable Tutores – Linear Regression

Objetivo del modelo: Obtener los factores que más influyen en las respuestas positivas de los alumnos hacia la tutoría.

Atributos a predecir: Integración, Interés Estudios, Mejoro Desempeño Alumno, Num Asistencias G, Num Asistencias I, Satisfacción

Structure	Tabla Minable Tutores - Linear Regression	Descripción de los datos a procesar
	Microsoft_Linear_Regression	
Comunicacion Tutor	Input	Comunicación permanente con el alumno
Confianza Tutor	Input	Nivel de confianza que inspiró el tutor sobre los alumnos
Disposicion Tutor	Input	Disposición de tutor para atender a los alumnos
Dudas Canalizadas	Input	Especifica si el tutor canalizó de forma correcta las dudas de los alumnos
Estimulo Tutor	Input	Estimulo del tutor para mejorar la forma de estudiar de los alumnos
Id Alumno	Key	Id del alumno
Integracion	PredictOnly	Nivel de integración del alumno a la Facultad de Ingeniería
Interes Estudios	PredictOnly	Nivel de interés en la carrera generado por el programa de tutoría
Interes Tutor	Input	Interés del tutor en los asuntos académicos y personales del alumno
Mejoro Desempeño Alumno	PredictOnly	Nivel de mejoría en el desempeño académico del alumno
Num Asistencias G	PredictOnly	Indica el número de asistencias que tuvo el alumno en las tutorías grupales
Num Asistencias I	PredictOnly	Indica el número de asistencias que tuvo el alumno en las tutorías individuales
Objetivos Tutor	Input	Establecimiento de objetivos por parte del tutor
Orientacion Tutor	Input	Orientación proporcionada por el tutor
Planeacion Tutor	Input	Capacidad de planeación del tutor
Puntualidad Tutor	Input	Puntualidad que tuvo el tutor en las sesiones de tutoría
Satisfaccion	PredictOnly	Nivel de satisfacción de los alumnos respecto al programa de tutoría
Seguimiento Tutor	Input	Seguimiento sistemático del tutor en las actividades del alumno

Como se puede observar en la figura los datos de entrada son aquellas características que genera un tutor sobre los tutorados de forma que produzcan algún efecto positivo en el alumno (salidas).

4.3.5 Patrones

La salida posterior a procesar el modelo se muestra a continuación:

$$Integracion = 3.087 + 0.292 * (Planeacion Tutor - 4.211) + 0.287 * (Orientacion Tutor - 4.063)$$

$$Interes Estudios = 2.499 + 0.065 * (Estimulo Tutor - 4.035)$$

$$\begin{aligned} \text{Mejoro Desempeño Alumno} &= 3.194 + 0.356 * (\text{Estimulo Tutor} - 4.044) + 0.261 * (\text{Planeacion Tutor} - 4.140) \\ \text{Num Asistencias G} &= 6.304 + 0.390 * (\text{Disposicion Tutor} - 4.533) + 0.341 * (\text{Seguimiento Tutor} - 4.156) \\ \text{Num Asistencias I} &= 0.709 + 0.184 * (\text{Orientacion Tutor} - 3.979) \\ \text{Satisfaccion} &= 3.528 + 0.283 * (\text{Planeacion Tutor} - 4.194) + 0.236 * (\text{Orientacion Tutor} - 4.036) + 0.262 * (\text{Interes Tutor} - 4.367) \end{aligned}$$

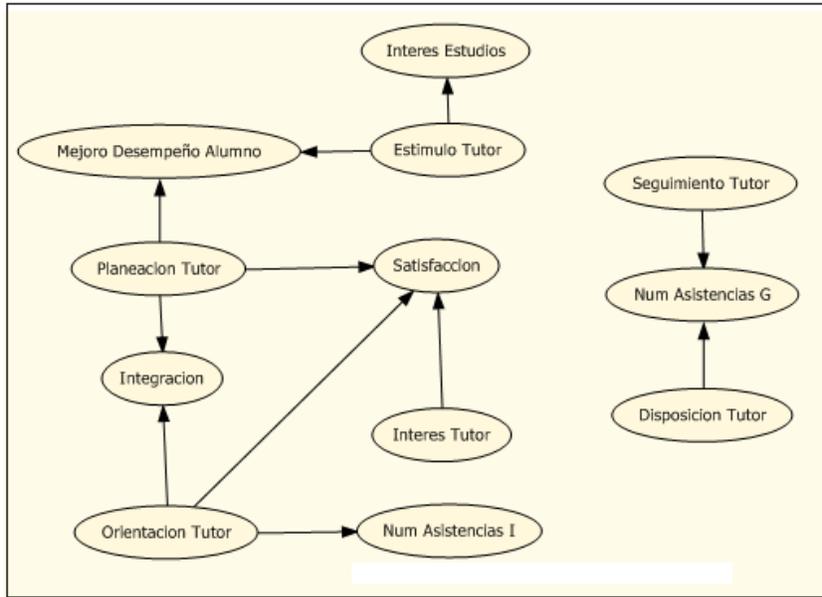


Figura 4.3.5.a Red de dependencias

Lo anterior es llamado Red de dependencias y en ella se muestran las relaciones entre los atributos que contribuyen a la capacidad de predicción del modelo de minería de datos.

4.3.6 Evaluación e interpretación

A diferencia de las gráficas de elevación mostradas anteriormente, el gráfico de elevación para una regresión lineal es diferente debido a que se utilizan solo datos continuos. Por la misma razón la matriz de clasificación no está disponible.

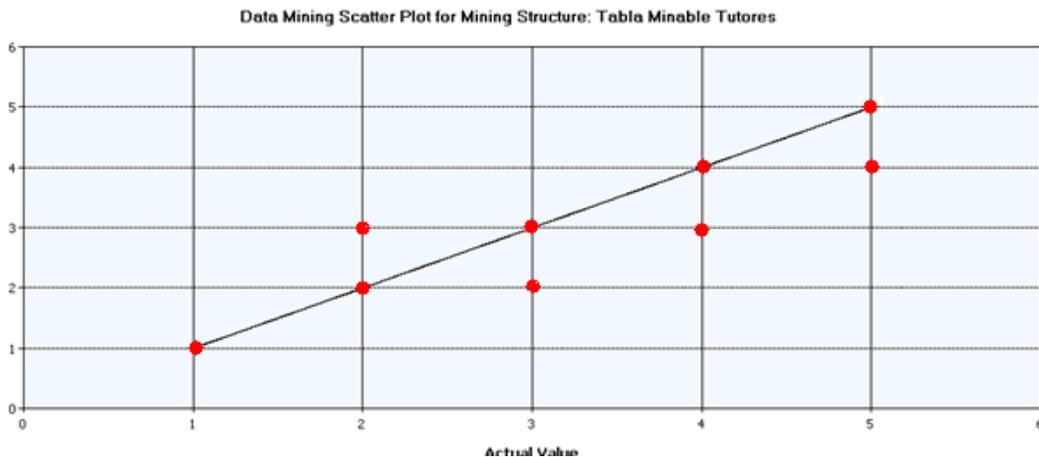


Figura 4.3.6.a Gráfico de elevación para 'Desempeño del Alumno'

Como se observa en la figura a mayor estímulo planeación del tutor mayor fue el desempeño del alumno

Las reglas de dependencia del algoritmo de regresión lineal proporciona la información necesaria (conocimiento) para la toma de decisiones que resuelva el problema inicialmente planteado para este análisis, por lo cual no es necesario otro análisis distinto para este problema.

4.3.7 Conocimiento

El presente análisis permitió conocer las principales características que un alumno considera para dar respuestas positivas respecto a la tutoría, así como para asistir con mayor regularidad a la misma.

4.3.8 Decisiones

Una toma de decisión enfocada a los tutores debe ser cautelosa de tal manera que nunca se pierda el punto de vista humano, esto es, un tutor es una persona con virtudes y defectos, una persona que día a día busca hacer la diferencia en la vida de aquellos a quienes imparte su conocimiento y quienes a su vez impactan en su vida de una forma u otra.

La complejidad de tomar una decisión respecto a los tutores se basa principalmente en que muchas veces al igual que un alumno que se encuentra laborando, un tutor no deja de tener una profesión, una familia y demás actividades las cuales limiten su tiempo y lo cual “mermaría” de cierto modo actividades extras que tenga. Sin embargo, también hay que considerar que el ser un tutor es de forma voluntaria, y que al hacerlo se adquieren grandes compromisos no solo con sus superiores, o con los alumnos si no con ellos mismos, y que al final de cuentas el aprendizaje será mutuo, citando al Tutor Octavio Estrada Castillo después de haber ayudado a un estudiante con sus problemas académicos:

Los dos aprendimos, él será un gran ingeniero y yo redimensioné lo que significa ser tutor. Concluyendo que se trata de un servicio prioritario, trascendente y al que debieran acudir todos los estudiantes.

A mis colegas tutores sólo les digo que la oportunidad de hablar con los estudiantes fuera del salón de clases constituye una aventura académica que, si la emprendemos con sencillez y apertura, nos hace aprender y crecer más de lo que imaginamos.

Lo anterior es muestra del impacto que la tutoría tiene sobre las personas involucradas, en una y otra dirección. Así es como el ser un mejor tutor te convierte en una mejor persona y ser una mejor persona te convierte en un mejor tutor.

Dicho lo anterior y regresando al análisis, se pudo observar que los factores que influyeron de mayor manera (simultáneamente) en la obtención de resultados positivos fueron la orientación y la planeación que un tutor lleva durante las sesiones, la primera refiriéndose a la orientación necesaria para que un alumno pudiera resolver sus problemas de orden personal y de salud que estaban fuera del ámbito de competencia del tutor y la segunda refiriéndose a la planeación de actividades que el tutor realizo con el alumno, lo cual evidenció el interés que tenía el tutor en llevar dichas actividades y no solo improvisar con el paso de la sesiones. Como puede observarse la línea que separa un atributo de otro es muy delgada, una buena planeación muestra interés y produce confianza, a la par que una buena comunicación permite establecer de mejor manera los objetivos y realizar una mejor orientación y una mejor resolución de dudas logrando de igual forma una mayor confianza. Todos los

atributos, por tanto, forman parte de un todo y para mejorar uno de ellos, es necesario mejorar los demás paralelamente. Sin embargo, y con base en los resultados del análisis realizado, la toma de decisión estará enfocada en la orientación y planeación, los cuales al ser incrementados de forma positiva habrán de impactar en las demás características.

Una característica importante para que el tutor pueda ayudar al alumno a atender sus necesidades es que primeramente el tutorado las exprese, es decir que el alumno tenga la confianza de comentárselas a su tutor, lo cual difícilmente se lograra en la primer sesión, sin embargo esta es indiscutiblemente el pilar para que se den las sesiones posteriores. De esta forma, y sabiendo que la "primera impresión" es la que realmente cuenta en las relaciones interpersonales [39], el tutor debe ganar la confianza del alumno o su mayor parte en la primera sesión, para lo cual un tutor debe estar preparado desde la primer sesión, es decir, existe una relación directa con la planeación. Lo anterior parte de una de las características más importante en cualquier relación interpersonal que es la comunicación, ante esto y con el único objetivo de mejorar este aspecto es importante considerar la programación neurolingüística como una herramienta para mejorar dicha comunicación.

La programación neurolingüística (PNL) es un conjunto de técnicas diseñadas para producir en la persona cambios permanentes a corto plazo. Proporciona un marco de referencia sistemático para dirigir el cerebro y además, enseña cómo manejar los estados y comportamientos propios y de los demás. La PNL describe cómo los individuos se comunican consigo mismo y con los demás, y propone que la comunicación interior puede originar estados de óptima disponibilidad de recursos y crear una amplia gama de comportamientos. [40]

La PNL se puede utilizar para desarrollar de manera rápida y eficaz un proceso de aprendizaje y así superar una situación de estrés, de conflicto, negociar con mayor ventaja frente a nuestros adversarios, etc. Es un complemento en el desarrollo de la Inteligencia emocional. Entre otras cosas, la PNL:

- Aumenta de manera notable y rápida la autoconfianza.
- Mejora las relaciones interpersonales.
- Desarrolla el crecimiento personal y profesional hacia el éxito.
- Nos permite convertirnos en quien deseamos y queremos ser.
- Sirve para reducir el estrés.
- Negociar y solucionar conflictos de manera positiva.

Enfocado en la educación la PNL ayuda a solucionar problemas de aprendizaje, aumentando la creatividad, el aprendizaje de las matemáticas, el aprendizaje de la física y la química, etc. De igual forma mejora las relaciones en el aula, ayudando a la solución de conflictos y mejora de la eficacia docente; entre otros.

Con todo lo anteriormente mencionado referente a la PNL no se está proponiendo que el tutor se vuelva un experto en este tema, sin embargo si se está planteando la posibilidad de que el tutor conozca un poco más de este tema con la finalidad de que el tutor logre interpretar de mejor manera las señales directas o indirectas del tutorado, esto es que el tutor pueda identificar aquellos alumnos que necesitan una mayor ayuda y de esta manera el tutor pueda enfocar mayores esfuerzos en ellos.

4.4 Tercer análisis: La herramienta

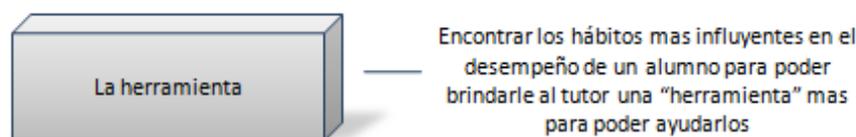


Figura 4.4.a La herramienta

4.4.1 Definición del Problema

Hasta el momento se han abordado dos asuntos importantes buscando encontrar ideas que permitan mejorar la tutoría: las características que diferencian a los alumnos que asisten de los que no (buscando tomar acciones en este último grupo), y qué elementos de un tutor hacen que un alumno tenga respuestas afirmativas acerca del programa de tutoría, pero es justamente después de este análisis que surge una nueva interrogante: ¿De qué forma se puede ayudar al tutor para trabajar de mejor manera con los tutorados?, dicha cuestión arroja una nueva idea, la cual consiste en analizar los hábitos de estudio de los alumnos, ya que si se lograran encontrar aquellos que influyen más en su desempeño el tutor sabría en que aspectos de sus tutorados enfocarse.

Así es como esta sección está destinada a analizar los hábitos de estudio de los alumnos, buscando con ello determinar cuáles son los que más afectan en el buen o mal desempeño de dichos alumnos, dándole así al tutor una “herramienta” más para ayudar a sus tutorados.

4.4.2 Vista Minable

La vista minable de este análisis consiste en mostrar las opciones escogidas por los alumnos en la encuesta socioeconómica referentes a los hábitos de estudio (90 preguntas⁴), además dicha vista muestra el resultado del alumno (rango del índice académico obtenido) durante el primer semestre.

```
create view TablaMinableHabitos as
WITH
  mc AS (SELECT idAlumno, count(*) as asignaturasCursadas FROM alumnomateria a
  group by idAlumno),
  mr AS (SELECT idAlumno, count(*) as asignaturasAprobadas FROM alumnomateria a
  where calificacion>=6
  group by idAlumno)
select *,
CASE WHEN indiceAcademico>=9 THEN '9-10'
  WHEN indiceAcademico>=8 AND indiceAcademico<9 THEN '8-8.99'
  WHEN indiceAcademico>=7 AND indiceAcademico<8 THEN '7-7.99'
  WHEN indiceAcademico>=6 AND indiceAcademico<7 THEN '6-6.99'
  ELSE '0-5.99'
END as resultados
from (
select idAlumno,
promedioPrimerSemestre,
```

⁴ Ver Anexo II Encuesta Socioeconómica – Hábitos de estudio

```
asignaturasAprobadas,
asignaturasCursadas,
cast(asignaturasAprobadas as float) / cast(asignaturasCursadas as float) as avanceAcademico,
round(promedioPrimerSemestre * (cast(asignaturasAprobadas as float) / cast(asignaturasCursadas
as float)),2) as indiceAcademico,
[Al leer lo hago de manera crítica y obteniendo mis propias conclusiones.],
[Mantengo buenas relaciones con mis compañeros y maestros.],
[Falto a clases.],
[Tomo apuntes en clase.],
[Aspectos emocionales o personales me impiden concentrarme en el estudio.],
[Copio a mis compañeros las respuestas de los ejercicios que pide el profesor.],
[Cuando estudio, repito mentalmente o en voz alta, y con libro cerrado, el material más
relevante.],
[Intento aprovechar al máximo mis estudios. ],
[Aprovecho adecuadamente las horas libres entre clases. ],
[Al leer mis apuntes, fácilmente distingo lo más importante para estudiar.],
[Me resulta difícil tomar decisiones respecto a mis estudios.],
[Al trabajar en equipo, me distraigo fácilmente en temas diferentes al estudio.],
[Cuando estudio un texto, me hago preguntas para cerciorarme que lo he comprendido. ],
[Encuentro agradable el ambiente de mi escuela. ],
[Estoy consciente del tiempo que debo dedicar al estudio cada día. ],
[Pido prestados apuntes a mis compañeros porque los míos son malos. ],
[Me quedo con dudas por temor a preguntar al profesor. ],
[En la biblioteca, en lugar de estudiar, me distraigo fácilmente.],
[Investigo o leo por mi propia iniciativa aspectos relacionados con mis materias.],
[Avanzo en mis estudios me convenzo que hice una buena elección profesional.],
[Dejo para el último momento la realización de mis tareas escolares.],
[Mis apuntes de clase están ordenados, limpios y legibles. ],
[Mis profesores tienen una buena opinión de mi como estudiante.],
[En la clase pregunto por la hora.],
[Al estudiar en un libro de texto, realizo los ejercicios que ahí se sugieren.],
[Estoy convencido(a) de que me gusta ir a la escuela y estudiar. ],
[Tomo en cuenta todas mis materias cuando organizo mi tiempo. ],
[Tengo mis apuntes de clase limpios y ordenados.],
[Me desvelo estudiando para los exámenes.],
[En el lugar donde estudio hay ruidos que me distraen.],
[Estudio con suficiente anticipación para los exámenes.],
[El trabajo que voy a realizar como profesional me parece interesante y creativo.],
[Planeo mis actividades escolares.],
[Antes de entregar un examen, verifico ordenadamente mis respuestas. ],
[Para aclarar mis dudas, pregunto al profesor. ],
[Cuando estudio me distraigo fácilmente con otras cosas. ],
[Practico lo aprendido.],
[En los exámenes sucede que me preguntan temas que no revisé.],
[A la hora de estudiar para los exámenes, lamento no tener los apuntes de clase.],
[Entrego puntualmente las tareas o trabajos que dejan los profesores. ],
[Me siento seguro (a) de mis conocimientos antes de iniciar un examen.],
[Tengo presente mi horario de manera que sé cual es la actividad planeada para determinada
hora.],
[Realizó más ejercicios que los que se me asignan],
[Antes de ponerme a estudiar, me aseguro de tener lo necesario para realizar la actividad
escolar (ca),
[Me presento a clase puntualmente el primer día de actividades.],
[Escribo de una manera legible y ordenada, mis respuestas en los exámenes.],
[Me siento frustrado (a) como estudiante.],
[ Realizo con mis compañeros sesiones de estudio en equipo.],
[Reviso mis apuntes oportunamente antes de entrar a clase.],
[Me intereso por las materias que llevo este semestre.],
[Dedico suficiente tiempo fuera de clase para estudiar mis materias.],
[Al iniciar el semestre defino claramente mis objetivos escolares.],
[Sólo estudio cuando tengo la presión de un examen. ],
[El lugar donde estudio tiene objetos que me distraen (fotografías, televisión, etc.). ],
[Para estudiar reviso libros adicionales a los que indica el profesor],
[Considero que mis clases son interesantes.],
```

[Siento que el día no me rinde para terminar mis actividades escolares.],
 [Contesto de manera clara y precisa las preguntas de los exámenes.],
 [Estudio sólo para pasar los exámenes.],
 [Escucho con atención lo que dice el maestro en clase],
 [Cuando me pongo a estudiar, hago caso omiso de lo que ocurre a mi alrededor.],
 [Copio los esquemas, ejemplos y anotaciones importantes que escribe el profesor.],
 [Utilizo una agenda y un reloj para organizar mis actividades.],
 [Mis apuntes son breves y concisos .],
 [Mis compañeros tienen una buena opinión de mi como estudiante.],
 [Mi escritorio o lugar de estudio está desorganizado.],
 [En un examen planeo mentalmente la respuesta antes de escribirla.],
 [Relaciono el tema estudiado con la vida diaria.],
 [Todos los días anticipo y programo las actividades que voy a realizar.],
 [Consigo oportunamente los libros y material que el maestro solicita en su clase.],
 [Participo en clase.],
 [Cuando estudio me concentro desde el principio.],
 [Elaboro cuadros sinópticos o resúmenes para sintetizar lo que leo.],
 [Me agrada ayudar a estudiar a mis compañeros.],
 [Tengo presente las fechas de inicio y terminación del semestre.],
 [Tengo un lugar organizado para guardar mis libros y útiles escolares.],
 [Mi antipatía por un profesor me impide aprender su materia.],
 [Mientras tomo notas pierdo puntos importantes de la explicación del maestro.],
 [Cuando estudio, elaboro resúmenes, utilizando mis propias palabras, sobre los temas expuestos en el],
 [Evito inscribirme con los maestros que gozan fama de exigentes.],
 [Con anticipación al examen, reviso que tenga todo el material necesario para estudiar.] ,
 [Al estudiar distingo claramente el qué, cómo y por qué de las cosas.],
 [Me pongo muy nervioso (a) al presentar un examen.],
 [En clase descubro con facilidad las ideas principales del tema.],
 [Cuando tomo notas, hago cuadros y tablas para hacerlos más comprensibles.],
 [Me gusta participar en actividades escolares complementarias],
 [El tiempo me alcanza para estudiar y tener actividades recreativas.],
 [Tengo una carpeta o cuaderno para cada una de las asignaturas.],
 [Cuando tengo que estudiar me siento cansado (a) y con sueño.] ,
 [Me concentro plenamente cuando estudio]

```

from (
    SELECT
        r.idAlumno as idAlumno,
        p.nombre as nombre,
        i.opcion as opcion,
        d.nombre as delegacion,
        ap.promedio as promedioPrimerSemestre,
        mc.asignaturasCursadas AS asignaturasCursadas,
        CASE WHEN mr.asignaturasAprobadas IS NULL THEN '0' Else mr.asignaturasAprobadas END As
asignaturasAprobadas,
        2009 - YEAR(fechaNac) as edad,
        idCarrera as carrera
    FROM resultados r
    inner join pregunta p on (r.idPregunta=p.idPregunta)
    inner join incisos i on (r.idIncisos=i.idIncisos)
    inner join alumno a on (r.idAlumno=a.idAlumno)
    inner join delegacion d on (d.idDelegacion=a.idDelegacion)
    inner join alumnopromedio ap on (ap.idAlumno=a.idAlumno)
    inner join mc on (mc.idAlumno=a.idAlumno)
    left join mr on mr.idAlumno=mc.idAlumno
    where ap.idSemestre = 1
) piv
PIVOT(
    Max(piv.opcion)
    FOR piv.nombre IN
    (
        [Al leer lo hago de manera crítica y obteniendo mis propias conclusiones.],
        [Mantengo buenas relaciones con mis compañeros y maestros.],
        [Falto a clases.],
    
```

[Tomo apuntes en clase.],
 [Aspectos emocionales o personales me impiden concentrarme en el estudio.],
 [Copio a mis compañeros las respuestas de los ejercicios que pide el profesor.],
 [Cuando estudio, repito mentalmente o en voz alta, y con libro cerrado, el material más relevante.],
 [Intento aprovechar al máximo mis estudios.],
 [Aprovecho adecuadamente las horas libres entre clases.],
 [Al leer mis apuntes, fácilmente distingo lo más importante para estudiar.],
 [Me resulta difícil tomar decisiones respecto a mis estudios.],
 [Al trabajar en equipo, me distraigo fácilmente en temas diferentes al estudio.],
 [Cuando estudio un texto, me hago preguntas para cerciorarme que lo he comprendido.],
 [Encuentro agradable el ambiente de mi escuela.],
 [Estoy consciente del tiempo que debo dedicar al estudio cada día.],
 [Pido prestados apuntes a mis compañeros porque los míos son malos.],
 [Me quedo con dudas por temor a preguntar al profesor.],
 [En la biblioteca, en lugar de estudiar, me distraigo fácilmente.],
 [Investigo o leo por mi propia iniciativa aspectos relacionados con mis materias.],
 [Avanzo en mis estudios me convenzo que hice una buena elección profesional.],
 [Dejo para el último momento la realización de mis tareas escolares.],
 [Mis apuntes de clase están ordenados, limpios y legibles.],
 [Mis profesores tienen una buena opinión de mí como estudiante.],
 [En la clase pregunto por la hora.],
 [Al estudiar en un libro de texto, realizo los ejercicios que ahí se sugieren.],
 [Estoy convencido(a) de que me gusta ir a la escuela y estudiar.],
 [Tomo en cuenta todas mis materias cuando organizo mi tiempo.],
 [Tengo mis apuntes de clase limpios y ordenados.],
 [Me desvelo estudiando para los exámenes.],
 [En el lugar donde estudio hay ruidos que me distraen.],
 [Estudio con suficiente anticipación para los exámenes.],
 [El trabajo que voy a realizar como profesional me parece interesante y creativo.],
 [Planeo mis actividades escolares.],
 [Antes de entregar un examen, verifico ordenadamente mis respuestas.],
 [Para aclarar mis dudas, pregunto al profesor.],
 [Cuando estudio me distraigo fácilmente con otras cosas.],
 [Practico lo aprendido.],
 [En los exámenes sucede que me preguntan temas que no revisé.],
 [A la hora de estudiar para los exámenes, lamento no tener los apuntes de clase.],
 [Entrego puntualmente las tareas o trabajos que dejan los profesores.],
 [Me siento seguro (a) de mis conocimientos antes de iniciar un examen.],
 [Tengo presente mi horario de manera que sé cual es la actividad planeada para determinada hora.],
 [Realizó más ejercicios que los que se me asignan],
 [Antes de ponerme a estudiar, me aseguro de tener lo necesario para realizar la actividad escolar (ca),
 [Me presento a clase puntualmente el primer día de actividades.],
 [Escribo de una manera legible y ordenada, mis respuestas en los exámenes.],
 [Me siento frustrado (a) como estudiante.],
 [Realizo con mis compañeros sesiones de estudio en equipo.],
 [Reviso mis apuntes oportunamente antes de entrar a clase.],
 [Me intereso por las materias que llevo este semestre.],
 [Dedico suficiente tiempo fuera de clase para estudiar mis materias.],
 [Al iniciar el semestre defino claramente mis objetivos escolares.],
 [Sólo estudio cuando tengo la presión de un examen.],
 [El lugar donde estudio tiene objetos que me distraen (fotografías, televisión, etc.).],
 [Para estudiar reviso libros adicionales a los que indica el profesor],
 [Considero que mis clases son interesantes.],
 [Siento que el día no me rinde para terminar mis actividades escolares.],
 [Contesto de manera clara y precisa las preguntas de los exámenes.],
 [Estudio sólo para pasar los exámenes.],

```

    [Escucho con atención lo que dice el maestro en clase],
    [Cuando me pongo a estudiar, hago caso omiso de lo que ocurre a mi alrededor.],
    [Copio los esquemas, ejemplos y anotaciones importantes que escribe el profesor.
],
    [Utilizo una agenda y un reloj para organizar mis actividades.],
    [Mis apuntes son breves y concisos .],
    [Mis compañeros tienen una buena opinión de mi como estudiante.],
    [Mi escritorio o lugar de estudio está desorganizado.],
    [En un examen planeo mentalmente la respuesta antes de escribirla.],
    [Relaciono el tema estudiado con la vida diaria.],
    [Todos los días anticipo y programo las actividades que voy a realizar. ],
    [Consigo oportunamente los libros y material que el maestro solicita en su
clase.],
    [Participo en clase.],
    [Cuando estudio me concentro desde el principio.],
    [Elaboro cuadros sinópticos o resúmenes para sintetizar lo que leo. ],
    [Me agrada ayudar a estudiar a mis compañeros.],
    [Tengo presente las fechas de inicio y terminación del semestre.],
    [Tengo un lugar organizado para guardar mis libros y útiles escolares.],
    [Mi antipatía por un profesor me impide aprender su materia.],
    [Mientras tomo notas pierdo puntos importantes de la explicación del maestro.],
    [Cuando estudio, elaboro resúmenes, utilizando mis propias palabras, sobre los
temas expuestos en el ],
    [Evito inscribirme con los maestros que gozan fama de exigentes.],
    [Con anticipación al examen, reviso que tenga todo el material necesario para
estudiar.],
    [Al estudiar distingo claramente el qué, cómo y por qué de las cosas.],
    [Me pongo muy nervioso (a) al presentar un examen.],
    [En clase descubro con facilidad las ideas principales del tema.],
    [Cuando tomo notas, hago cuadros y tablas para hacerlos más comprensibles.],
    [Me gusta participar en actividades escolares complementarias],
    [El tiempo me alcanza para estudiar y tener actividades recreativas.],
    [Tengo una carpeta o cuaderno para cada una de las asignaturas. ],
    [Cuando tengo que estudiar me siento cansado (a) y con sueño. ],
    [Me concentro plenamente cuando estudio]
)
) as chld) as tbl

```

4.4.3 Elección del algoritmo

Debido a la gran cantidad de datos así como a su naturaleza, algoritmos como el de red neuronal o el de regresión lineal serían inadecuados y poco eficientes, y con base en estos mismos argumentos es que la elección del algoritmo será el de *Clústeres*, con el cual se pretende lograr agrupaciones de alumnos que hayan tenido buenos resultados y aquellos que no lo hayan hecho así, para posteriormente analizar los grupos más significativos, lo cual permitirá encontrar los principales hábitos que los diferencian.

4.4.4 Minería de datos

Fundamentado en lo expuesto en el apartado anterior (La gran cantidad de datos: 90 hábitos + identificador del alumno + resultados) y a que el nombre del atributo corresponde a la pregunta exacta, la tabla que se muestra a continuación no contiene una columna de descripción de datos como en los anteriores análisis.

Características principales del modelo:

Estructura a la que pertenece: Tabla Minable Hábitos

Modelo: Tabla Minable Hábitos – Clústeres

Objetivo del modelo:

Atributos a predecir: Resultados

Structure ▲	Tabla Minable Habitos - Clusters
	🚨 Microsoft_Clustering
📄 A La Hora De Estudiar Para Los Exámenes Lamento No Tener Los Apuntes De Clase	📄 Input
📄 Al Estudiar Distingo Claramente El Qué Cómo y Por Qué De Las Cosas	📄 Input
📄 Al Estudiar En Un Libro De Texto Realizo Los Ejercicios Que Ahí Se Sugieren	📄 Input
📄 Al Iniciar El Semestre Defino Claramente Mis Objetivos Escolares	📄 Input
📄 Al Leer Lo Hago De Manera Crítica y Obteniendo Mis Propias Conclusiones	📄 Input
📄 Al Leer Mis Apuntes Fácilmente Distingo Lo Más Importante Para Estudiar	📄 Input
📄 Al Trabajar En Equipo Me Distraigo Fácilmente En Temas Diferentes Al Estudio	📄 Input
📄 Antes De Entregar Un Examen Verifico Ordenadamente Mis Respuestas	📄 Input
📄 Antes De Ponerme a Estudiar Me Aseguro De Tener Lo Necesario Para Realizar La Actividad Escolar ca	📄 Input
📄 Aprovecho Adecuadamente Las Horas Libres Entre Clases	📄 Input
📄 Aspectos Emocionales o Personales Me Impiden Concentrarme En El Estudio	📄 Input
📄 Avanzo En Mis Estudios Me Convenzo Que Hice Una Buena Elección Profesional	📄 Input
📄 Con Anticipación Al Examen Reviso Que Tenga Todo El Material Necesario Para Estudiar	📄 Input
📄 Considero Que Mis Clases Son Interesantes	📄 Input
📄 Consigo Oportunamente Los Libros y Material Que El Maestro Solicita En Su Clase	📄 Input
📄 Contesto De Manera Clara y Precisa Las Preguntas De Los Exámenes	📄 Input
📄 Copio a Mis Compañeros Las Respuestas De Los Ejercicios Que Pide El Profesor	📄 Input
📄 Copio Los Esquemas Ejemplos y Anotaciones Importantes Que Escribe El Profesor	📄 Input
📄 Cuando Estudio Elaboro Resúmenes Utilizando Mis Propias Palabras Sobre Los Temas Expuestos En El	📄 Input
📄 Cuando Estudio Me Concentro Desde El Principio	📄 Input
📄 Cuando Estudio Me Distraigo Fácilmente Con Otras Cosas	📄 Input
📄 Cuando Estudio Repito Mentalmente o En Voz Alta y Con Libro Cerrado El Material Más Relevante	📄 Input
📄 Cuando Estudio Un Texto Me Hago Preguntas Para Cerciorarme Que Lo He Comprendido	📄 Input
📄 Cuando Me Pongo a Estudiar Hago Caso Omiso De Lo Que Ocurre a Mi Alrededor	📄 Input
📄 Cuando Tengo Que Estudiar Me Siento Cansado a y Con Sueño	📄 Input
📄 Cuando Tomo Notas Hago Cuadros y Tablas Para Hacerlos Más Comprensibles	📄 Input
📄 Dedico Suficiente Tiempo Fuera De Clase Para Estudiar Mis Materias	📄 Input
📄 Dejo Para El Último Momento La Realización De Mis Tareas Escolares	📄 Input
📄 El Lugar Donde Estudio Tiene Objetos Que Me Distraen fotografías Televisión Etc	📄 Input
📄 El Tiempo Me Alcanza Para Estudiar y Tener Actividades Recreativas	📄 Input
📄 El Trabajo Que Voy a Realizar Como Profesional Me Parece Interesante y Creativo	📄 Input
📄 Elaboro Cuadros Sinópticos o Resúmenes Para Sintetizar Lo Que Leo	📄 Input
📄 En Clase Descubro Con Facilidad Las Ideas Principales Del Tema	📄 Input
📄 En El Lugar Donde Estudio Hay Ruidos Que Me Distraen	📄 Input
📄 En La Biblioteca En Lugar De Estudiar Me Distraigo Fácilmente	📄 Input
📄 En La Clase Pregunto Por La Hora	📄 Input
📄 En Los Exámenes Sucede Que Me Preguntan Temas Que No Revisé	📄 Input

	En Un Examen Planeo Mentalmente La Respuesta Antes De Escribirla		Input
	Encuentro Agradable El Ambiente De Mi Escuela		Input
	Entrego Puntualmente Las Tareas o Trabajos Que Dejan Los Profesores		Input
	Escribo De Una Manera Legible y Ordenada Mis Respuestas En Los Exámenes		Input
	Escucho Con Atención Lo Que Dice El Maestro En Clase		Input
	Estoy Consciente Del Tiempo Que Debo Dedicar Al Estudio Cada Día		Input
	Estoy Convencidoa De Que Me Gusta Ir a La Escuela y Estudiar		Input
	Estudio Con Suficiente Anticipación Para Los Exámenes		Input
	Estudio Sólo Para Pasar Los Exámenes		Input
	Evito Inscribirme Con Los Maestros Que Gozan Fama De Exigentes		Input
	Falto a Clases		Input
	Id Alumno		Key
	Intento Aprovechar Al Máximo Mis Estudios		Input
	Investigo o Leo Por Mi Propia Iniciativa Aspectos Relacionados Con Mis Materias		Input
	Mantengo Buenas Relaciones Con Mis Compañeros y Maestros		Input
	Me Resulta Difícil Tomar Decisiones Respecto a Mis Estudios		Input
	Me Agrade Ayudar a Estudiar a Mis Compañeros		Input
	Me Concentro Plenamente Cuando Estudio		Input
	Me Desvelo Estudiando Para Los Exámenes		Input
	Me Gusta Participar En Actividades Escolares Complementarias		Input
	Me Intereso Por Las Materias Que Llevo Este Semestre		Input
	Me Pongo Muy Nervioso a Al Presentar Un Examen		Input
	Me Presento a Clase Puntualmente El Primer Día De Actividades		Input
	Me Quedo Con Dudas Por Temor a Preguntar Al Profesor		Input
	Me Siento Frustrado a Como Estudiante		Input
	Me Siento Seguro a De Mis Conocimientos Antes De Iniciar Un Examen		Input
	Mi Antipatía Por Un Profesor Me Impide Aprender Su Materia		Input
	Mi Escritorio o Lugar De Estudio Está Desorganizado		Input
	Mientras Tomo Notas Pierdo Puntos Importantes De La Explicación Del Maestro		Input
	Mis Apuntes De Clase Están Ordenados Limpios y Legibles		Input
	Mis Apuntes Son Breves y Concisos		Input
	Mis Compañeros Tienen Una Buena Opinión De Mi Como Estudiante		Input
	Mis Profesores Tienen Una Buena Opinión De Mi Como Estudiante		Input
	Para Aclarar Mis Dudas Pregunto Al Profesor		Input
	Para Estudiar Reviso Libros Adicionales a Los Que Indica El Profesor		Input
	Participo En Clase		Input
	Pido Prestados Apuntes a Mis Compañeros Porque Los Míos Son Malos		Input
	Planeo Mis Actividades Escolares		Input
	Practico Lo Aprendido		Input
	Realizo Con Mis Compañeros Sesiones De Estudio En Equipo		Input
	Realizó Más Ejercicios Que Los Que Se Me Asignan		Input
	Relaciono El Tema Estudiado Con La Vida Diaria		Input
	Resultados		PredictOnly

Reviso Mis Apuntes Oportunamente Antes De Entrar a Clase	Input
Siento Que El Día No Me Rinde Para Terminar Mis Actividades Escolares	Input
Sólo Estudio Cuando Tengo La Presión De Un Examen	Input
Tengo Mis Apuntes De Clase Limpios y Ordenados	Input
Tengo Presente Las Fechas De Inicio y Terminación Del Semestre	Input
Tengo Presente Mi Horario De Manera Que Sé Cual Es La Actividad Planeada Para Determinada Hora	Input
Tengo Un Lugar Organizado Para Guardar Mis Libros y Útiles Escolares	Input
Tengo Una Carpeta o Cuaderno Para Cada Una De Las Asignaturas	Input
Todos Los Días Anticipo y Programo Las Actividades Que Voy a Realizar	Input
Tomo Apuntes En Clase	Input
Tomo En Cuenta Todas Mis Materias Cuando Organizo Mi Tiempo	Input
Utilizo Una Agenda y Un Reloj Para Organizar Mis Actividades	Input

4.4.5 Patrones

Los *Clústeres* generados para este análisis se muestran a continuación:

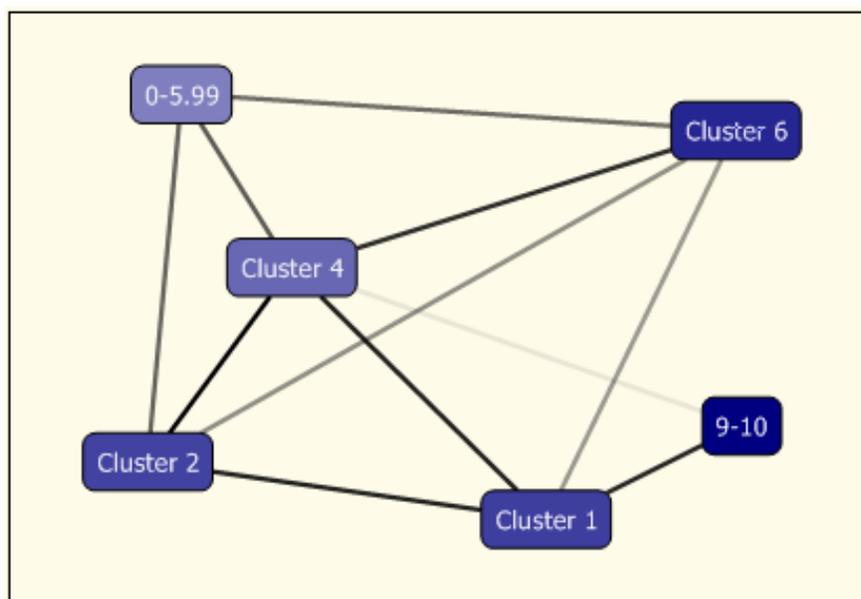


Figura 4.4.5.a Clústeres

En la figura se indican cuales *Clústeres* tuvieron una mayor densidad para las clasificaciones extremas, es decir de 0-5.99 y el de 9-10.

Características del *Clúster* "9-10" el cual contiene las principales características de los alumnos que obtuvieron un índice académico dentro de ese rango.

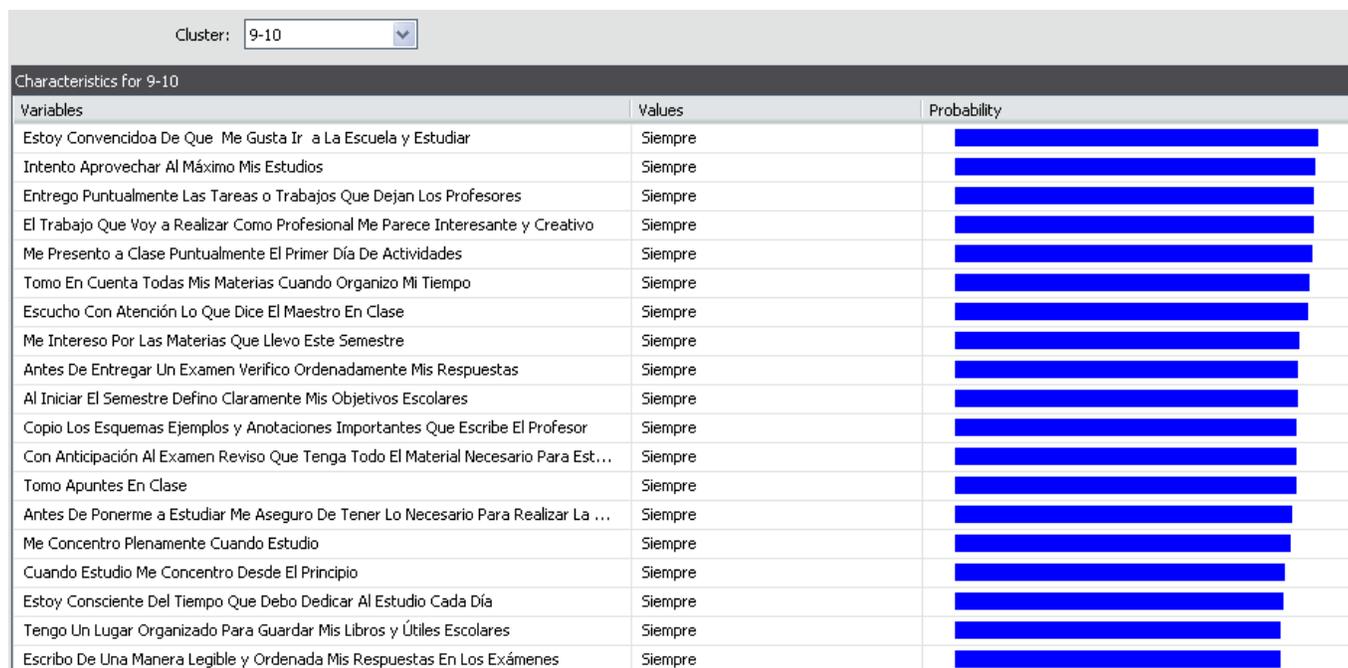


Figura 4.4.5.b Características del clúster 9-10

Características del Clúster “0-5.99” el cual contiene las principales características de los alumnos que obtuvieron un índice académico dentro de ese rango.

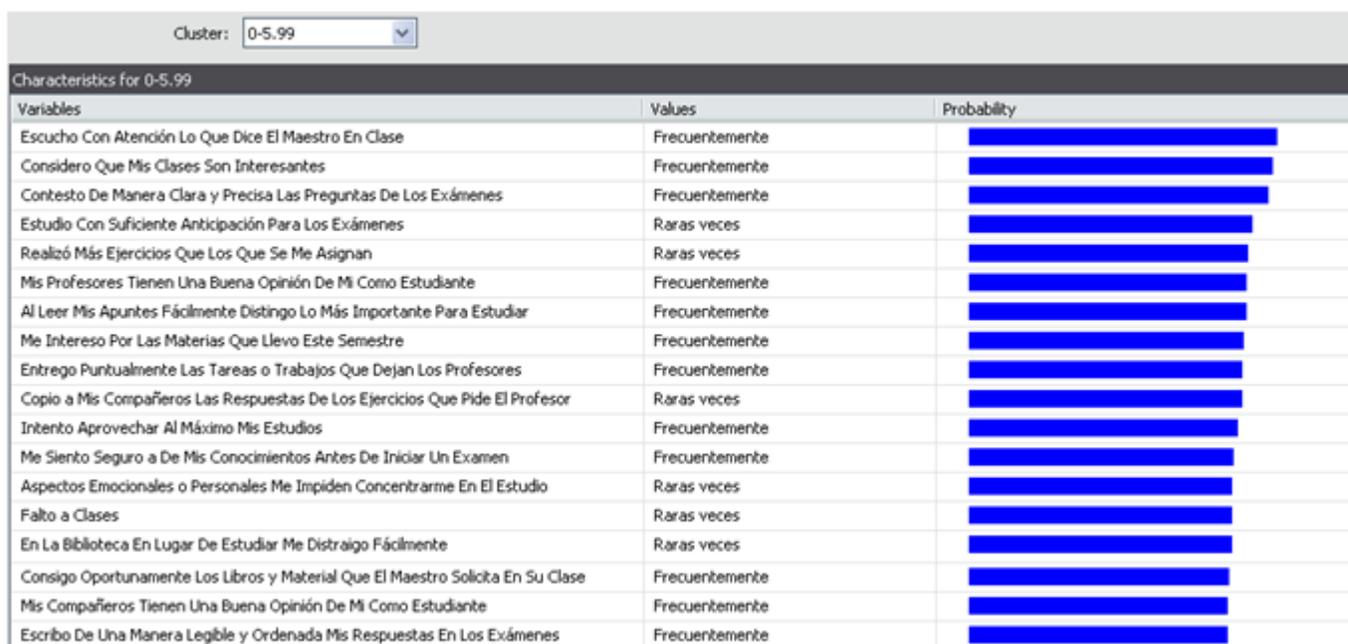


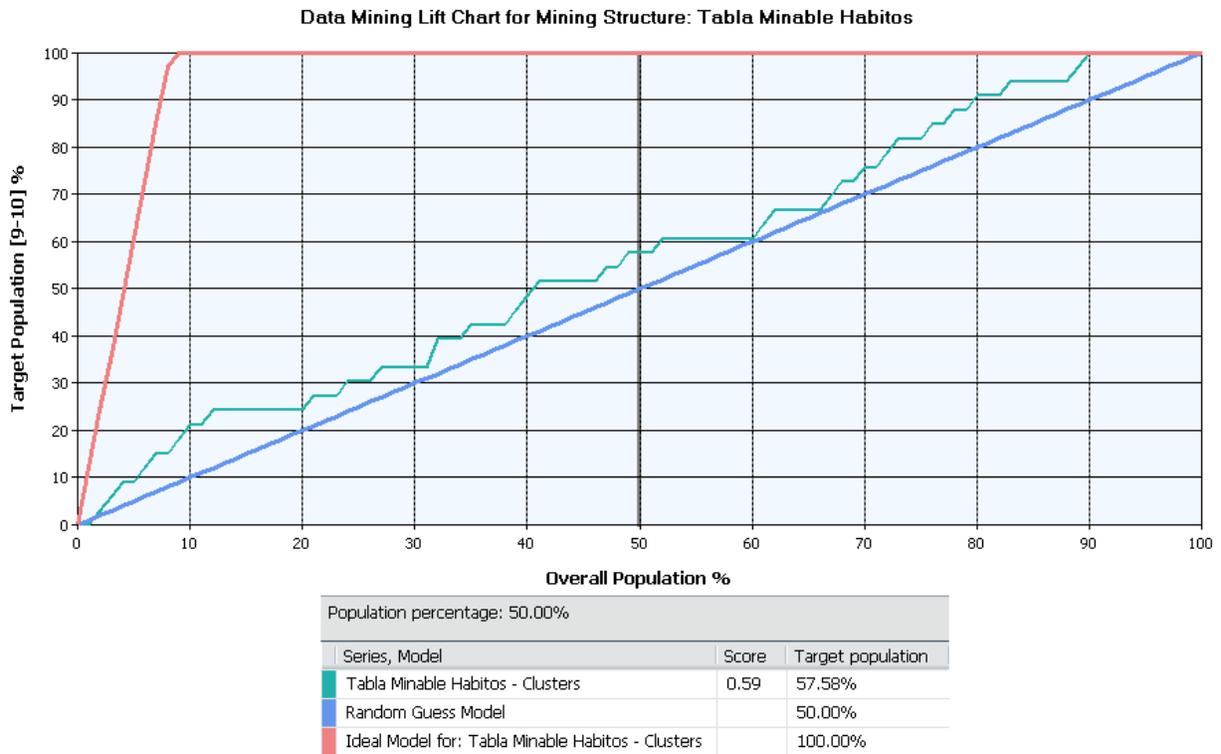
Figura 4.4.5.c Características del clúster 0-5.99

Principales diferencias entre los Clústeres de “9-10” y “0-5.99”:

Variables	States	9-10 ▾	0-5.99
Escucho Con Atención Lo Que Dice El Maestro En Clase	<ul style="list-style-type: none"> ● Siempre ● Frecuentemente ● Raras veces ● Nunca ● Other ... 		
Estoy Convencidoa De Que Me Gusta Ir a La Escuela y Estudiar	<ul style="list-style-type: none"> ● Siempre ● Frecuentemente ● Raras veces ● Nunca ● Other ... 		
Dedico Suficiente Tiempo Fuera De Clase Para Estudiar Mis Materias	<ul style="list-style-type: none"> ● Frecuentemente ● Raras veces ● Siempre ● Nunca ● Other ... 		
A La Hora De Estudiar Para Los Exámenes Lamento No Tener Los Apuntes De Clase	<ul style="list-style-type: none"> ● Nunca ● Raras veces ● Frecuentemente ● Siempre ● Other ... 		
Copio Los Esquemas Ejemplos y Anotaciones Importantes Que Escribe El Profesor	<ul style="list-style-type: none"> ● Siempre ● Frecuentemente ● Raras veces ● Nunca ● Other ... 		
Me Intereso Por Las Materias Que Llevo Este Semestre	<ul style="list-style-type: none"> ● Siempre ● Frecuentemente ● Raras veces ● Nunca ● Other ... 		
El Trabajo Que Voy a Realizar Como Profesional Me Parece Interesante y Creativo	<ul style="list-style-type: none"> ● Siempre ● Frecuentemente ● Raras veces ● missing ● Other ... 		
Antes De Ponerme a Estudiar Me Aseguro De Tener Lo Necesario Para Realizar L...	<ul style="list-style-type: none"> ● Siempre ● Frecuentemente ● Raras veces ● Nunca ● Other ... 		
Al Estudiar En Un Libro De Texto Realizo Los Ejercicios Que Ahí Se Sugieren	<ul style="list-style-type: none"> ● Frecuentemente ● Raras veces ● Siempre ● Nunca ● Other ... 		
Escribo De Una Manera Legible y Ordenada Mis Respuestas En Los Exámenes	<ul style="list-style-type: none"> ● Siempre ● Frecuentemente ● Raras veces ● Nunca ● Other ... 		

Figura 4.4.5.d Principales diferencias entre clústeres

4.4.6 Evaluación e interpretación



4.4.6.a Gráfico de elevación

La matriz de clasificación nos muestra que:

Counts for Tabla Minable Habitos - Clusters on Resultados:					
Predicted	0-5.99 (Actual)	6-6.99 (Actual)	7-7.99 (Actual)	8-8.99 (Actual)	9-10 (Actual)
0-5.99	248	34	29	56	33

4.4.6.b Matriz de clasificación

El algoritmo utilizado es capaz por sí solo de dar los datos necesarios para adquirir el conocimiento y con base en el tomar la decisión, por lo cual no es necesario ningún análisis más.

4.4.7 Conocimiento

El análisis ha proporcionado satisfactoriamente los hábitos de estudio más importantes en el desempeño del alumno de una forma coherente, esto es, los resultados muestran un patrón claro respecto a la opción seleccionada, siendo que las respuestas del grupo con mayores índices académicos estuvieron en extremos (Siempre o Nunca), no así como en el caso del grupo con menor índice académico donde las respuestas de dicho grupo nunca estuvieron en extremos si no que se mantuvieron en el intermedio.

4.4.8 Decisiones

Como se planteo en la definición del problema que corresponde a este análisis, lo que se pretende es encontrar los hábitos más importantes tanto para un buen y mal desempeño, de manera que un tutor pudiera enfocarse en ellos, es decir, que el tutor ayudará a cambiar los hábitos de estudio más perjudiciales en el desempeño de los alumnos y de igual forma fomentará los hábitos más benéficos en los mismos.

Dicho lo anterior, una manera de ayudar al tutor sería principalmente brindarle al mismo la siguiente tabla, en la cual se muestran los hábitos (P) en los que un tutor tendría que enfocarse si un tutorado contesta las opciones mostradas (R) bajo la pregunta.

1	P	Escucho con atención lo que dice el maestro en clase
	R	Frecuente, Raras veces, Nunca
2	P	Estoy convencido de que me gusta ir a la escuela y estudiar
	R	Frecuente, Raras veces, Nunca
3	P	Dedico suficiente tiempo fuera de clase para estudiar mis materias
	R	Raras veces, Nunca
4	P	A la hora de estudiar para los exámenes lamento no tener los apuntes de clase
	R	Raras veces, Nunca
5	P	Copio los esquemas, ejemplos y anotaciones importantes que escribe el profesor
	R	Frecuente, Raras veces, Nunca
6	P	Me intereso por las materias que llevo en este semestre
	R	Frecuente, Raras veces, Nunca
7	P	El trabajo que voy a realizar como profesional me parece interesante y creativo
	R	Frecuente, Raras veces, Nunca
8	P	Antes de ponerme a estudiar me aseguro de tener lo necesario para realizar la actividad escolar (calculadora, goma, etc.)
	R	Frecuente, Raras veces, Nunca
9	P	Al estudiar en un libro de texto realizo los ejercicios que ahí se sugieren
	R	Raras veces, Nunca
10	P	Escribo de una manera legible y ordenada mis respuestas en los exámenes
	R	Frecuente, Raras veces, Nunca

Así mismo se le debe de proporcionar al tutor las respuestas respecto a los hábitos de los tutorados que le sean asignados, para que el pueda saber sobre quienes enfocarse (sin descuidar a los demás por supuesto). Inicialmente un alumno podría no contestar las 10 preguntas anteriores como se indica, pero un tutor debería considerar tomar medidas si uno de sus tutorados contesta así más del 50%.

Se puede observar además que en algunas veces no se considera la respuesta “frecuente”, aunque esta haya

aparecido en el hábito de estudio para un mal índice académico, esto es porque dicha respuesta también aparece en el grupo que tuvo un buen índice académico con igual o mayor frecuencia.

Para que todo lo anteriormente mencionado funcione, es necesario que dicha información se le proporcione al tutor previo al inicio del semestre.

Con todo lo anterior el tutor con base en su experiencia sabrá cómo dirigir a sus tutorados de la mejor manera, por ejemplo si un tutorado selecciona una respuesta negativa para el hábito “El trabajo que voy a realizar como profesional me parece interesante y creativo” el tutor podría buscar la forma de que el estudiante se sumerja más en la carrera que eligió, es posible que el tutor decida recomendarle exposiciones en donde se muestren las actividades que se desarrollan en su carrera o bien el tutor decida ayudar al alumno a encontrar la carrera que mejor se adapte a sus necesidades ya que siendo el primer semestre el alumno aún está a tiempo de encontrar su verdadera vocación ya sea dentro o fuera de la Facultad.

4.5 Conclusiones Finales

La tutoría es una de las herramientas de integración, adaptación y superación tanto laboral como personal más importantes con las que se cuenta actualmente, y no sólo en el ámbito estudiantil sino también en el ámbito laboral, con ella se logra que el desempeño de los integrantes de un grupo determinado sea de una mejor manera y de una forma más rápida. La historia nos muestra que la tutoría ha sido un recurso utilizado largamente durante el pasar del tiempo, buscando siempre que mediante la tutela el estudiante logre concretar todo de lo que es capaz. Así es como buscando lo anterior las diversas instituciones educativas y/o gubernamentales han desarrollado estrategias para llevarlo a cabo, en algunas empresas se suelen brindar a sus empleados lo que es llamado “Tutoría de prácticas en empresas” en donde los tutores son las personas encargadas del control y el seguimiento de las prácticas laborales que sus tutorados están realizando.

Ahora si nos enfocamos a la tutoría educativa es posible observar cómo es que en varios países se toman medidas para ayudar en la preparación académica y humana del estudiante. En las universidades anglosajonas (USA, Gran Bretaña) habitualmente se designa un tutor común en los estudios de postgrado, al que se le asignan varios estudiantes, su misión consiste en coordinar cursos con el alumno, así como actividades académicas y de investigación y de igual forma se encargan de supervisar su tránsito. En la Universidad Nacional de La Plata en Argentina, se suele utilizar la tutoría electrónica como servicio de apoyo al estudiante en seminarios de postgrado, las cuales son apoyadas con tutorías presenciales.

Sin embargo aún cuando en muchos de los casos existe una planificación de los programas a llevarse a cabo para el funcionamiento ideal de la tutoría, dichas herramientas no consiguen obtener el impacto deseado, ya sea en la cobertura de las mismas o en el cumplimiento de las metas para las cuales fueron establecidas. Lo expresado anteriormente se debe en gran medida a la falta de estudio de los individuos, es decir en muchas ocasiones se considera sólo la parte académica o laboral de la persona, y es que cuando la tutoría está enfocada solo en mejorar estos aspectos puede descuidar el aspecto más importante: el humano. Aún con lo anterior, es indiscutible que conforme el paso del tiempo los programas de tutoría van “pulíéndose”, la forma de hacerlo es corrigiendo o mejorando características de los mismos, así como aplicando nuevos mecanismos que permitan satisfacer en gran medida sus objetivos, y aunque es claro que difícilmente un programa de este tipo alcanzara un 100% de satisfacción, también es claro que un programa bien analizado, estructurado y llevado a cabo estará cerca de alcanzarlo. Y es en este punto en donde entra este trabajo.

El presente trabajo tenía como objetivo obtener conocimiento útil y novedoso acerca de todas las fuentes de información recolectadas las cuales permitieran lograr los siguientes objetivos específicos:

- Demostrar la utilidad del programa de tutoría como herramienta en la mejoría del desempeño académico de los alumnos.
- Propuestas para incrementar la asistencia a la tutoría.
- Propuestas para mejorar la tutoría enfocado al tutor.
- Propuestas para ayudar al tutor con su labor.

Sin embargo antes de abordar a detalle cada uno de estos puntos, lo ideal es darle el merecimiento adecuado a la forma de obtención de los datos que permitieron alcanzar dichos objetivos.

Anteriormente la forma de convertir los datos en conocimiento consistía en analizar e interpretar la información histórica de forma manual, es decir si un grupo de médicos necesitaba analizar la evolución de enfermedades infecto-contagiosas en una población para determinar ciertas características como el rango de edad más frecuente de las personas afectadas, necesitaba revisar los datos de persona en persona, esta forma era o quizás es (de seguir siendo usada en algunas instituciones) lenta, costosa y altamente subjetiva, esto último debido a que muchas decisiones importantes se realizan siguiendo la propia intuición del usuario al no disponer de las herramientas necesarias, además de que el análisis manual es impracticable en dominios donde el volumen de los datos crece exponencialmente, desbordando la capacidad humana de comprenderlos sin la ayuda de herramientas potentes. Así es como surge la necesidad de herramientas y técnicas que soportarán la extracción de conocimiento útil desde la información disponible. La minería de datos se distingue porque no obtiene información extensional (datos) sino intencional (conocimiento) y, este no es, un modelo preestablecido o intuido por el usuario, es un modelo novedoso y original, extraído por la herramienta. Para la realización de la minería de datos es importante seguir un proceso bien definido, el cual permita la correcta consecución del conocimiento. Fases de descubrimiento de conocimiento en bases de datos (KDD), es el proceso mediante el cual la minería de datos debe realizarse, estas incluyen desde la fase de análisis hasta la fase de toma de decisiones, siendo además un proceso iterativo, permitiendo que la salida de alguna de las fases puede hacer volver a pasos anteriores para mejorar la calidad del proceso.

En este trabajo se llevó a cabo el proceso del KDD siguiendo todas sus fases, eligiendo en cada caso el algoritmo más adecuado para la realización de la minería de datos y evaluándolos mediante gráficos de elevación y matrices de clasificación. Con esto fue posible obtener el conocimiento necesario que permitiera cumplir con los objetivos expresados anteriormente.

Para el presente trabajo se plantearon tres análisis que abarcaron a los tutorados, los tutores y la herramienta, entendiéndose por esta última al apoyo extra que se le debe dar al tutor en su labor, dichos análisis buscaban alcanzar los objetivos planteados de la mejor manera, además de que permitirían confirmar o negar las hipótesis formuladas en el capítulo 1:

- La tutoría alcanza a un 70% de la población de alumnos de licenciatura en ingeniería.
- Aquellos alumnos que no obtuvieron una buena calificación en su examen diagnóstico son aquellos que asisten con mayor regularidad a la tutoría.
- La tutoría influye de manera considerable en el promedio del alumno.

- La distancia hogar-universidad es uno de los principales factores en la inasistencia del alumno a la tutoría, entre mas distancia mayor inasistencia.
- La asistencia del alumno a la tutoría va ligada directamente a la disponibilidad del alumno, esto es, aquellos que trabajan tienen una asistencia menor que aquellos que no lo hacen.
- La clase social de un alumno no influye en la asistencia del alumno en la tutoría ni su promedio del semestre.
- El desempeño de un alumno va de la mano con la efectividad que tenga un tutor en las sesiones de tutoría.
- Los hábitos de estudio de un alumno son parte fundamental en un buen desempeño académico.

Las siguientes figuras muestran los datos utilizados para realizar cada uno de los análisis. En ellas se pueden observar las características referentes a la minería de datos, desde la vistas minables hasta los patrones obtenidos una vez procesados los modelos.

Análisis	Objetivo	Minería					
		Vista	Algoritmo	Estructura	Modelo	Patrones	
Analizar el impacto de la tutoría	¿Cuánto influye en el índice académico del alumno asistir a la tutoría?	TablaMinableRes	Red Neuronal	Tabla Minable Res	Tabla Minable Res – Neuronal Network	Relación elevada entre la asistencia y el índice académico.	
Analizar la asistencia a la tutoría	¿Qué características tienen en común los alumnos que asisten a la tutoría de los que no lo hacen?	TablaMinable	Clústeres	TablaMinable	Tabla Minable – Clúster	Edad Trabajo Promedio bachillerato	✓
			Árboles de decisión		Tabla Minable – Trees	Edad Promedio bachillerato	
			Naive Bayes		Tabla Minable – Bayes	Edad Trabajo Promedio bachillerato	
Analizar el índice académico del alumno	¿Qué características tienen en común los alumnos que obtienen un mejor desempeño académico reflejado en su índice académico?	TablaMinableRes 2	Naive Bayes	Tabla Minable Res2	Tabla Minable Res2 – Naive Bayes	Edad Trabajo Promedio bachillerato Escuela Promedio ED	✓
Análisis Extra: Los que trabajan	Demostrar que aquellos alumnos que trabajan y aun así asisten a la tutoría tienen un mejor desempeño académico.	TablaMinableRes Variación	Red Neuronal	Tabla Minable Res Variación	Tabla Minable Res Variación – Neural Network	Se demostró que los que trabajan y asisten tienen un mejor desempeño académico	✓

Figura 4.5.a Tabla de análisis – Los tutorados

Análisis	Objetivo	Minería					
		Vista	Algoritmo	Estructura	Modelo	Patrones	
Analizar las cualidades del tutor	Encontrar las cualidades de un tutor que mas satisfacen a un alumno	TablaMinable Tutores	Regresión Lineal	Tabla Minable Tutores	Tabla Minable Tutores – Linear Regression	Obtención de 6 diferentes ecuaciones para cada aspecto de interés	✓
		$Integración = 3.087+0.292*(Planeación\ Tutor-4.211)+0.287*(Orientación\ Tutor-4.063)$ $Interés\ Estudios = 2.499+0.065*(Estimulo\ Tutor-4.035)$ $Mejoro\ Desempeño\ Alumno = 3.194+0.356*(Estimulo\ Tutor-4.044)+0.261*(Planeación\ Tutor-4.140)$ $Núm.\ Asistencias\ G = 6.304+0.390*(Disposición\ Tutor-4.533)+0.341*(Seguimiento\ Tutor-4.156)$ $Núm.\ Asistencias\ I = 0.709+0.184*(Orientación\ Tutor-3.979)$ $Satisfacción = 3.528+0.283*(Planeación\ Tutor-4.194)+0.236*(Orientación\ Tutor-4.036)+0.262*(Interés\ Tutor-4.367)$					

Figura 4.5.b Tabla de análisis – Los tutores

Análisis	Objetivo	Minería					
		Vista	Algoritmo	Estructura	Modelo	Patrones	
Analizar los hábitos de estudio	Encontrar los hábitos mas influyentes en el desempeño de un alumno para poder brindarle al tutor una "herramienta" mas para poder ayudarlos	TablaMinable Hábitos	Clústeres	Tabla Minable Hábitos	Tabla Minable Hábitos – Clústeres	Obtención de una lista de hábitos mas importantes	✓
		<p>Escucho con atención lo que dice el maestro en clase.</p> <ul style="list-style-type: none"> • Estoy convencido de que me gusta ir a la escuela y estudiar. • Dedico suficiente tiempo fuera de clase para estudiar mis materias. • A la hora de estudiar para los exámenes lamento no tener los apuntes de clase. • Copio los esquemas, ejemplos y anotaciones importantes que escribe el profesor. • Me intereso por las materias que llevo en este semestre. • El trabajo que voy a realizar como profesional me parece interesante y creativo. • Antes de ponerme a estudiar me aseguro de tener lo necesario para realizar la actividad escolar (calculadora, goma, etc.). • Al estudiar en un libro de texto realizo los ejercicios que ahí se sugieren. • Escribo de una manera legible y ordenada mis respuestas en los exámenes. 					

Figura 4.5.c Tabla de análisis – La herramienta

Los análisis mostrados permitieron demostrar que un estudiante que asiste a las tutorías sin importar que esté trabajando, tiene un mejor desempeño académico (reflejado en su índice académico) que aquellos que no lo hacen, además se pudieron realizar la toma decisiones, las cuales pueden enlistarse como sigue:

- Asignar tutores de tiempo completo a los estudiantes que se encuentren trabajando.
- Aumentar la tutoría electrónica sin perder de vista lo esencial de las tutorías presenciales.
- Utilizar PNL en la capacitación de los tutores.
- Brindar al tutor una lista de los hábitos más influyentes en el mejor desempeño académico, así como una relación de los alumnos asignados al tutor que no cumplan con los mismos.

Cuando nos basamos en la gráfica obtenida en el primer análisis:

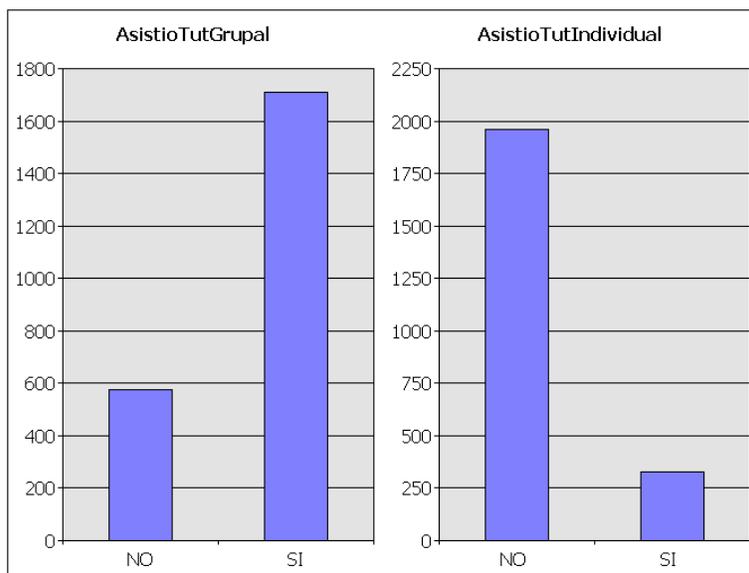


Figura 4.5.d Asistencia tutorías

Es posible observar que la tutoría grupal alcanza más de un 70% de la población total, lo cual corrobora el primer punto de las hipótesis.

El primer análisis demostró que el examen diagnóstico no influye en la asistencia de la tutoría, y si el promedio en bachillerato.

La distancia hogar-universidad no es un factor principal en la inasistencia del alumno a la tutoría, siendo el factor más importante si el alumno trabaja o no.

La clase social no influye en la asistencia del alumno a la tutoría.

La planeación que tenga el tutor en las tutorías así como el estímulo que logre en el alumno son los factores principales para un buen desempeño académico.

Los hábitos de estudio juegan un papel muy importante en el desempeño académico del estudiante, siendo los extremos de las opciones planteadas para cada hábito las que más arrojaron resultados, esto es, aquellos que seleccionaron opciones como 'Siempre' o 'Nunca' fueron quienes tuvieron los mejores y peores desempeños.

La cercanía a la perfección de cualquier programa de tutoría está fundamentado en el estudio del mismo así como en el de los individuos que lo conforman, cuando los resultados y el conocimiento han sido obtenidos es necesario la aplicación de decisiones que corrijan los baches encontrados, una vez transcurrido un tiempo adecuado (el suficiente para recopilar una cantidad de información igual o mayor que la primera analizada) es necesario reanalizar la información e incluso pueden existir nuevas fuentes de datos disponibles que arrojen mayor cantidad de conocimiento, el cual nuevamente permita tomar decisiones, es decir, que para mejorar un programa de tutoría es indispensable realizar este proceso tantas veces sea necesario, tantas veces como se quiera seguir mejorando. Este trabajo ha estado enfocado en la primera parte: se analizó la información, se obtuvo conocimiento y se propusieron decisiones, la responsabilidad respecto a su aplicación o rechazo, no recae

en un servidor, sin embargo, es deseable que las decisiones propuestas sean estudiadas como es debido por parte de aquellos cuyo compromiso es con los estudiantes. Y es que, en ocasiones lo más fácil es continuar apegado a un sistema que ha dado resultados, pero porqué detenerse ahí, porqué no pensar en tener un mayor impacto en la comunidad, mejorando su cobertura y sus beneficios, haciendo que el paso del tiempo enriquezca más aún el programa, le establezca nuevas metas o les exija más a las ya existentes. Con lo anterior, y como propuesta personal, es necesario que se realice un mayor seguimiento de estudiante el cual permita recopilar mayor información, si bien es cierto que la cantidad de datos proporcionada fue razonablemente buena, también es cierto que la minería de datos para ser ideal necesita ser sobre volúmenes considerablemente grandes, es necesario hacer un mayor seguimiento de los alumnos que son constantes durante el paso de sus semestres académicos, seguir recopilando opiniones y asistencias como si de calificaciones se tratara, de esta manera en el futuro la minería de datos podría ser mejor, más útil y arrojando nuevo conocimiento.

La realidad es que el programa de Tutoría “Nueva Era” tiene una buena cobertura, así como un impacto grande en el desempeño de los alumnos, generando buenas opiniones y satisfaciendo no solo a los estudiantes, sino también a los tutores, quienes ven en este programa una oportunidad más de seguir aprendiendo junto al alumno. Pero es precisamente cuando un programa está funcionando tan bien, que el siguiente objetivo no debe ser otro que el de mejorarlo, porque mientras no se deje de trabajar en él, la cantidad de gente beneficiada nunca dejara de crecer.

APÉNDICE I Cuestionario de evaluación del programa Tutoría Nueva Era

1.- Muestra el tutor buena disposición el tutor para atender a los alumnos,

Poca Mediana Regular Suficiente Mucha No aplica

2.- Muestra el tutor interés en los problemas académicos y personales que afectan el rendimiento de los alumnos

Poca Mediana Regular Suficiente Mucha No aplica

3.- Muestra el tutor disposición a mantener una comunicación permanente con el alumno

Poca Mediana Regular Suficiente Mucha No aplica

4.- Tu tutor estableció los objetivos de la tutoría

Poca Mediana Regular Suficiente Mucha No aplica

5.- Las dudas sobre tu desempeño académico expresadas en las sesiones de tutoría, fueron aclaradas o canalizadas acertadamente

Poca Mediana Regular Suficiente Mucha No aplica

6.- El trabajo que desarrollaste con el tutor evidenció que hubo una planeación de las actividades y no una mera improvisación

Poca Mediana Regular Suficiente Mucha No aplica

7.- El clima propiciado por el tutor te dio la confianza para expresarle tus dudas

Poca Mediana Regular Suficiente Mucha No aplica

8.- Asistió el tutor puntualmente a las sesiones

Poca Mediana Regular Suficiente Mucha No aplica

9.- Tu tutor mantuvo seguimiento sistemático de las actividades y acuerdos establecidos

Poca Mediana Regular Suficiente Mucha No aplica

10.- Tu tutor te proporcionó la orientación necesaria para atender tus necesidades de orden personal y de salud

Poca Mediana Regular Suficiente Mucha No aplica

11.- Te estimuló para que desarrollaras habilidades para estudiar de manera independiente

Poca Mediana Regular Suficiente Mucha No aplica

12.- Mostraste buena disposición para asistir a la tutoría

Poca Mediana Regular Suficiente Mucha No aplica

13.- Asististe puntualmente a las sesiones de tutoría

Poca Mediana Regular Suficiente Mucha No aplica

14.- Mostraste disposición para mantener una comunicación permanente con el tutor

Poca Mediana Regular Suficiente Mucha No aplica

15.- Cuando tenías dudas se las expresabas al tutor

Poca Mediana Regular Suficiente Mucha No aplica

16.- Mantuviste un seguimiento sistemático con las actividades realizadas en las sesiones de tutoría

Poca Mediana Regular Suficiente Mucha No aplica

- 17.- Tu participación en las sesiones de tutoría ha mejorado tu desempeño académico
 Poca Mediana Regular Suficiente Mucha No aplica
- 18.- ¿A cuántas sesiones de tutoría grupal asististe?
 Ninguna 1 2 3 4 5 6 7 8 Mas de 8
- 19.- ¿A cuántas sesiones de tutoría grupal asististe?
 Ninguna 1 2 3 4 5 6 7 8 Mas de 8
- 20.- Tu integración a la Facultad ha mejorado con el programa de tutoría
 Nada Poca Regular Mucha No recibiste atención
- 21.- Te es satisfactorio el programa de tutoría
 Nada Poca Regular Mucha No recibiste atención
- 24.- ¿Asististe a la plática de bienvenida que brindó la Facultad de Ingeniería?
 Si No
- 24a.- En caso de haber asistido a dicha plática te pareció:
 Muy provechosa Provechosa Poco provechosa Nada provechosa No aplica
- 25.- ¿Asistió tu padre, madre o tutor a la plática de bienvenida a la Facultad de Ingeniería?
 Si No
- 25a.- En caso de haber asistido les pareció:
 Muy provechosa Provechosa Poco provechosa Nada provechosa No aplica
- 26.- ¿Asististe a la plática de bienvenida que brindaron los funcionarios de tu carrera?
 Si No
- 26a.- En caso de haber asistido a dicha plática te pareció:
 Muy provechosa Provechosa Poco provechosa Nada provechosa No aplica
- 27.- El folleto o cuadernillo sobre el Procedimiento de Inscripción para el Primer Ingreso te pareció
 Muy útil Útil Poco útil Nada útil
- 28.- ¿Consultaste el CD Multimedia que se te proporciono al ingresar?
 Si No
- 28a.- El contenido de dicho CD Multimedia te pareció:
 Regular Excelente Muy bueno Bueno Malo No aplica
- 29.- El examen diagnóstico que respondiste vía red antes de iniciar el semestre te pareció:
 Muy fácil Fácil Difícil Muy difícil
- 30.- Las sugerencias que recibiste por tus resultados en dicho examen te parecieron:
 Muy útiles Útiles Poco útiles Nada útiles No aplica
- 31.- ¿Asististe a la plática sobre los servicios que brinda la biblioteca que se ofreció como parte de las actividades de tutoría?,
 Si No
- 31a.- Dicha plática te pareció:
 Muy provechosa Provechosa Poco provechosa Nada provechosa No aplica

32.- ¿Conociste tus resultados del Sistema de Valoración de Conductas Orientadas al Estudio (SIVACORE)?

Si No

32a.- En caso afirmativo, lo comentarios recibidos te parecieron

Muy útiles Útiles Poco útiles Nada útiles No aplica

33.- El cambio de bachillerato a la universidad te ha resultado

Muy fácil Fácil Difícil Muy difícil

34.- En este momento tu interés por estudiar la carrera de ingeniería

Se ha incrementado Se ha mantenido Se ha disminuido

35.- ¿Has modificado la administración de tu tiempo para que se adapte a las exigencias de tu carrera?

Sí de manera definitiva Sí de manera reducida No porque no ha sido necesario

No porque no te lo han propuesto No porque te lo has propuesto para más adelante

APÉNDICE II Encuesta socioeconómica – Hábitos de estudio

1.- Al leer lo hago de manera crítica y obteniendo mis propias conclusiones.

Siempre Frecuentemente Raras veces Nunca

2.- Mantengo buenas relaciones con mis compañeros y maestros.

Siempre Frecuentemente Raras veces Nunca

3.- Falto a clases.

Siempre Frecuentemente Raras veces Nunca

4.- Tomo apuntes en clase.

Siempre Frecuentemente Raras veces Nunca

5.- Aspectos emocionales o personales me impiden concentrarme en el estudio.

Siempre Frecuentemente Raras veces Nunca

6.- Copio a mis compañeros las respuestas de los ejercicios que pide el profesor.

Siempre Frecuentemente Raras veces Nunca

7.- Cuando estudio, repito mentalmente o en voz alta, y con libro cerrado, el material más relevante.

Siempre Frecuentemente Raras veces Nunca

8.- Intento aprovechar al máximo mis estudios.

Siempre Frecuentemente Raras veces Nunca

9.- Aprovecho adecuadamente las horas libres entre clases.

Siempre Frecuentemente Raras veces Nunca

10.- Al leer mis apuntes, fácilmente distingo lo más importante para estudiar.

Siempre Frecuentemente Raras veces Nunca

11.- Me resulta difícil tomar decisiones respecto a mis estudios.

Siempre Frecuentemente Raras veces Nunca

12.- Al trabajar en equipo, me distraigo fácilmente en temas diferentes al estudio.

Siempre Frecuentemente Raras veces Nunca

13.- Cuando estudio un texto, me hago preguntas para cerciorarme que lo he comprendido.

Siempre Frecuentemente Raras veces Nunca

14.- Encuentro agradable el ambiente de mi escuela.

Siempre Frecuentemente Raras veces Nunca

15.- Estoy consciente del tiempo que debo dedicar al estudio cada día.

Siempre Frecuentemente Raras veces Nunca

16.- Pido prestados apuntes a mis compañeros porque los míos son malos.

Siempre Frecuentemente Raras veces Nunca

17.- Me quedo con dudas por temor a preguntar al profesor.

Siempre Frecuentemente Raras veces Nunca

18.- En la biblioteca, en lugar de estudiar, me distraigo fácilmente.

Siempre Frecuentemente Raras veces Nunca

- 19.- Investigo o leo por mi propia iniciativa aspectos relacionados con mis materias.
Siempre Frecuentemente Raras veces Nunca
- 20.- avanzo en mis estudios me convenzo que hice una buena elección profesional.
Siempre Frecuentemente Raras veces Nunca
- 21.- Dejo para el último momento la realización de mis tareas escolares.
Siempre Frecuentemente Raras veces Nunca
- 22.- Mis apuntes de clase están ordenados, limpios y legibles.
Siempre Frecuentemente Raras veces Nunca
- 23.- Mis profesores tienen una buena opinión de mi como estudiante.
Siempre Frecuentemente Raras veces Nunca
- 24.- En la clase pregunto por la hora.
Siempre Frecuentemente Raras veces Nunca
- 25.- Al estudiar en un libro de texto, realizo los ejercicios que ahí se sugieren.
Siempre Frecuentemente Raras veces Nunca
- 26.- Estoy convencido(a) de que me gusta ir a la escuela y estudiar.
Siempre Frecuentemente Raras veces Nunca
- 27.- Tomo en cuenta todas mis materias cuando organizo mi tiempo.
Siempre Frecuentemente Raras veces Nunca
- 28.- Tengo mis apuntes de clase limpios y ordenados.
Siempre Frecuentemente Raras veces Nunca
- 29.- Me desvelo estudiando para los exámenes.
Siempre Frecuentemente Raras veces Nunca
- 30.- En el lugar donde estudio hay ruidos que me distraen.
Siempre Frecuentemente Raras veces Nunca
- 31.- Estudio con suficiente anticipación para los exámenes.
Siempre Frecuentemente Raras veces Nunca
- 32.- El trabajo que voy a realizar como profesional me parece interesante y creativo.
Siempre Frecuentemente Raras veces Nunca
- 33.- Planeo mis actividades escolares.
Siempre Frecuentemente Raras veces Nunca
- 34.- Antes de entregar un examen, verifico ordenadamente mis respuestas.
Siempre Frecuentemente Raras veces Nunca
- 35.- Para aclarar mis dudas, pregunto al profesor.
Siempre Frecuentemente Raras veces Nunca
- 36.- Cuando estudio me distraigo fácilmente con otras cosas.
Siempre Frecuentemente Raras veces Nunca
- 37.- Practico lo aprendido.
Siempre Frecuentemente Raras veces Nunca

- 38.- En los exámenes sucede que me preguntan temas que no revisé.
Siempre Frecuentemente Raras veces Nunca
- 39.- A la hora de estudiar para los exámenes, lamento no tener los apuntes de clase.
Siempre Frecuentemente Raras veces Nunca
- 40.- Entrego puntualmente las tareas o trabajos que dejan los profesores.
Siempre Frecuentemente Raras veces Nunca
- 41.- Me siento seguro (a) de mis conocimientos antes de iniciar un examen.
Siempre Frecuentemente Raras veces Nunca
- 42.- Tengo presente mi horario de manera que sé cual es la actividad planeada para determinada hora.
Siempre Frecuentemente Raras veces Nunca
- 43.- Realizó más ejercicios que los que se me asignan
Siempre Frecuentemente Raras veces Nunca
- 44.- Antes de ponerme a estudiar, me aseguro de tener lo necesario para realizar la actividad escolar (calculadora, goma, etc.)
Siempre Frecuentemente Raras veces Nunca
- 45.- Me presento a clase puntualmente el primer día de actividades.
Siempre Frecuentemente Raras veces Nunca
- 46.- Escribo de una manera legible y ordenada, mis respuestas en los exámenes.
Siempre Frecuentemente Raras veces Nunca
- 47.- Me siento frustrado (a) como estudiante.
Siempre Frecuentemente Raras veces Nunca
- 48.- Realizo con mis compañeros sesiones de estudio en equipo.
Siempre Frecuentemente Raras veces Nunca
- 49.- Reviso mis apuntes oportunamente antes de entrar a clase.
Siempre Frecuentemente Raras veces Nunca
- 50.- Me intereso por las materias que llevo este semestre.
Siempre Frecuentemente Raras veces Nunca
- 51.- Dedico suficiente tiempo fuera de clase para estudiar mis materias.
Siempre Frecuentemente Raras veces Nunca
- 52.- Al iniciar el semestre defino claramente mis objetivos escolares.
Siempre Frecuentemente Raras veces Nunca
- 53.- Sólo estudio cuando tengo la presión de un examen.
Siempre Frecuentemente Raras veces Nunca
- 54.- El lugar donde estudio tiene objetos que me distraen (fotografías, televisión, etc.).
Siempre Frecuentemente Raras veces Nunca
- 55.- Para estudiar reviso libros adicionales a los que indica el profesor
Siempre Frecuentemente Raras veces Nunca
- 56.- Considero que mis clases son interesantes.

- | | Siempre | Frecuentemente | Raras veces | Nunca |
|---|---------|----------------|-------------|-------|
| 57.- Siento que el día no me rinde para terminar mis actividades escolares. | | | | |
| 58.- Contesto de manera clara y precisa las preguntas de los exámenes. | | | | |
| 59.- Estudio sólo para pasar los exámenes. | | | | |
| 60.- Escucho con atención lo que dice el maestro en clase | | | | |
| 61.- Cuando me pongo a estudiar, hago caso omiso de lo que ocurre a mí alrededor. | | | | |
| 62.- Copio los esquemas, ejemplos y anotaciones importantes que escribe el profesor. | | | | |
| 63.- Utilizo una agenda y un reloj para organizar mis actividades. | | | | |
| 64.- Mis apuntes son breves y concisos. | | | | |
| 65.- Mis compañeros tienen una buena opinión de mí como estudiante. | | | | |
| 66.- Mi escritorio o lugar de estudio está desorganizado. | | | | |
| 67.- En un examen planeo mentalmente la respuesta antes de escribirla. | | | | |
| 68.- Relaciono el tema estudiado con la vida diaria. | | | | |
| 69.- Todos los días anticipo y programo las actividades que voy a realizar. | | | | |
| 70.- Consigo oportunamente los libros y material que el maestro solicita en su clase. | | | | |
| 71.- Participo en clase. | | | | |
| 72.- Cuando estudio me concentro desde el principio. | | | | |
| 73.- Elaboro cuadros sinópticos o resúmenes para sintetizar lo que leo. | | | | |
| 74.- Me agrada ayudar a estudiar a mis compañeros. | | | | |
| 75.- Tengo presente las fechas de inicio y terminación del semestre. | | | | |

- | | Siempre | Frecuentemente | Raras veces | Nunca |
|---|---------|----------------|-------------|-------|
| 76.- Tengo un lugar organizado para guardar mis libros y útiles escolares. | | | | |
| | Siempre | Frecuentemente | Raras veces | Nunca |
| 77.- Mi antipatía por un profesor me impide aprender su materia. | | | | |
| | Siempre | Frecuentemente | Raras veces | Nunca |
| 78.- Mientras tomo notas pierdo puntos importantes de la explicación del maestro. | | | | |
| | Siempre | Frecuentemente | Raras veces | Nunca |
| 79.- Cuando estudio, elaboro resúmenes, utilizando mis propias palabras, sobre los temas expuestos en el libro. | | | | |
| | Siempre | Frecuentemente | Raras veces | Nunca |
| 80.- Evito inscribirme con los maestros que gozan fama de exigentes. | | | | |
| | Siempre | Frecuentemente | Raras veces | Nunca |
| 81.- Con anticipación al examen, reviso que tenga todo el material necesario para estudiar. | | | | |
| | Siempre | Frecuentemente | Raras veces | Nunca |
| 82.- Al estudiar distingo claramente el qué, cómo y por qué de las cosas. | | | | |
| | Siempre | Frecuentemente | Raras veces | Nunca |
| 83.- Me pongo muy nervioso (a) al presentar un examen. | | | | |
| | Siempre | Frecuentemente | Raras veces | Nunca |
| 84.- En clase descubro con facilidad las ideas principales del tema. | | | | |
| | Siempre | Frecuentemente | Raras veces | Nunca |
| 85.- Cuando tomo notas, hago cuadros y tablas para hacerlos más comprensibles. | | | | |
| | Siempre | Frecuentemente | Raras veces | Nunca |
| 86.- Me gusta participar en actividades escolares complementarias | | | | |
| | Siempre | Frecuentemente | Raras veces | Nunca |
| 87.- El tiempo me alcanza para estudiar y tener actividades recreativas. | | | | |
| | Siempre | Frecuentemente | Raras veces | Nunca |
| 88.- Tengo una carpeta o cuaderno para cada una de las asignaturas. | | | | |
| | Siempre | Frecuentemente | Raras veces | Nunca |
| 89.- Cuando tengo que estudiar me siento cansado (a) y con sueño. | | | | |
| | Siempre | Frecuentemente | Raras veces | Nunca |
| 90.- Me concentro plenamente cuando estudio | | | | |
| | Siempre | Frecuentemente | Raras veces | Nunca |

APÉNDICE III Creando Predicciones

SQL Server provee un Generador de consultas de predicción de Extensiones de minería de datos, la cual permite generar predicciones basadas en los datos de entrada y condiciones indicadas.

Para el siguiente ejemplo se utilizara el modelo 'Tabla Minable - Bayes' y la vista 'TablaMinable' (donde al ir agregando nuevos alumnos a la base de datos es posible realizar el análisis específico para el mismo), lo que habrá de predecirse es si un alumno asistirá o no a las tutorías grupales.

Primeramente es necesario seleccionar el modelo y la tabla con lo cual es posible obtener algo como lo siguiente:

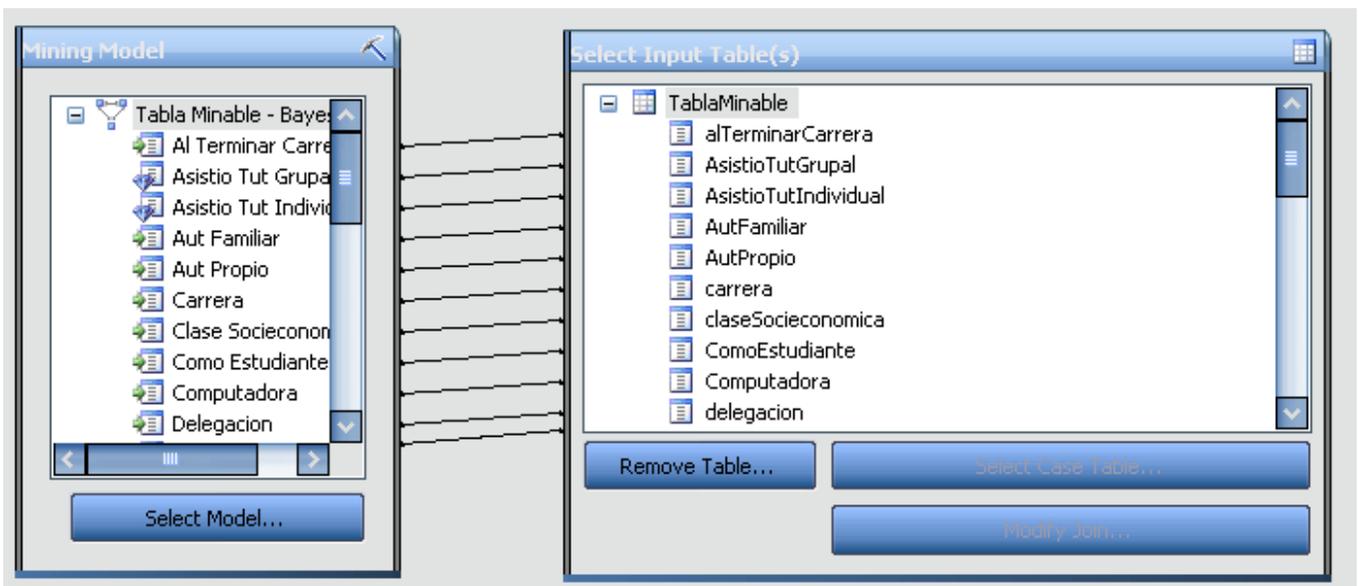


Figura Anexo 1.1

Lo siguiente es modificar las relaciones creadas automáticamente, en caso de que estén correctas, solo es necesario confirmarlas:

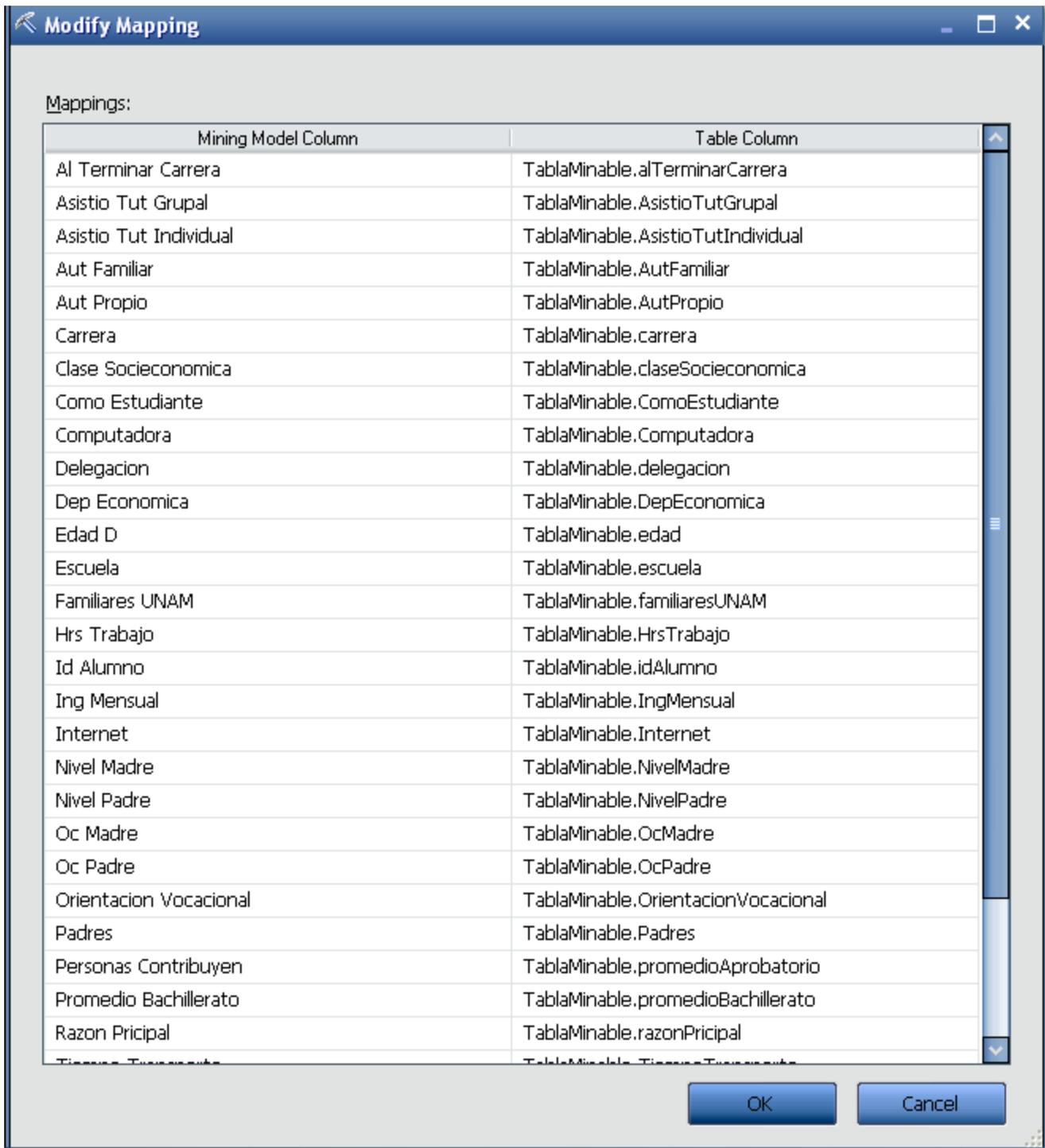


Figura Anexo 1.2

Es importante que el valor a predecir (Asistió Tut Grupal) tenga una columna asignada, en este caso esa columna ya tiene un valor, pero será sobrescrito en el resultado final por el valor predicho.

Una vez hecho lo anterior es necesario colocar algunas características de la predicción a realizar:

Source	Field	Alias	Show	Group	And/Or	Criteria/Argument
 Prediction Function	 PredictProbability	Probabilidad	<input checked="" type="checkbox"/>			[Tabla Minable - Bayes].[Asistio Tut Grupal]
 Tabla Minable - Bayes	 Asistio Tut Grupal	Asistira	<input checked="" type="checkbox"/>			= 'SI'
 TablaMinable	 idAlumno		<input checked="" type="checkbox"/>			

Figura Anexo 1.3

La primera fila indica que lo que se desea realizar es una predicción y que el campo deseado es la probabilidad de que dicha predicción ocurra y que su alias será *Probabilidad*, además se indica que el valor debe ser mostrado y que el argumento a predecir es *Asistio Tut Grupál*

La segunda columna indica el atributo que valor debe tener *Asistio Tut Grupál*, en este caso el valor es 'SI'.

La tercera indica que también se desea ver en el resultado el id del alumno.

Lo anterior es equivalente a la consulta:

```

SELECT
  (PredictProbability([Tabla Minable - Bayes].[Asistio Tut Grupal])) as [Probabilidad],
  ([Tabla Minable - Bayes].[Asistio Tut Grupal]) as [Asistira],
  t.[idAlumno]
From
  [Tabla Minable - Bayes]
PREDICTION JOIN
  OPENQUERY ([Generacion2010],
    'SELECT
      [idAlumno],
      [delegacion],
      [escuela],
      [promedioBachillerato],
      [AsistioTutGrupal],
      [AsistioTutIndividual],
      [razonPricipal],
      [familiaresUNAM],
      [alTerminarCarrera],
      [Computadora],
      [Internet],
      [AutPropio],
      [AutFamiliar],
      [Trabaja],
      [HrsTrabajo],
      [Padres],
      [NivelPadre],
      [NivelMadre],
      [OrientacionVocacional],
      [ComoEstudiante],
      [UsoComputadora],
      [OcPadre],
      [OcMadre],
      [DepEconomica],
      [promedioAprobatorio],
      [IngMensual],
      [Transporte],
      [TiempoTransporte],
      [claseSocieconomica],
      [edad]
    ')

```

```

FROM
  [dbo].[TablaMinable]
') AS t
ON
[Tabla Minable - Bayes].[Delegacion] = t.[delegacion] AND
[Tabla Minable - Bayes].[Escuela] = t.[escuela] AND
[Tabla Minable - Bayes].[Promedio Bachillerato] = t.[promedioBachillerato] AND
[Tabla Minable - Bayes].[Asistio Tut Grupal] = t.[AsistioTutGrupal] AND
[Tabla Minable - Bayes].[Asistio Tut Individual] = t.[AsistioTutIndividual] AND
[Tabla Minable - Bayes].[Razon Pricipal] = t.[razonPricipal] AND
[Tabla Minable - Bayes].[Familiares UNAM] = t.[familiaresUNAM] AND
[Tabla Minable - Bayes].[Al Terminar Carrera] = t.[alTerminarCarrera] AND
[Tabla Minable - Bayes].[Computadora] = t.[Computadora] AND
[Tabla Minable - Bayes].[Internet] = t.[Internet] AND
[Tabla Minable - Bayes].[Aut Propio] = t.[AutPropio] AND
[Tabla Minable - Bayes].[Aut Familiar] = t.[AutFamiliar] AND
[Tabla Minable - Bayes].[Trabaja] = t.[Trabaja] AND
[Tabla Minable - Bayes].[Hrs Trabajo] = t.[HrsTrabajo] AND
[Tabla Minable - Bayes].[Padres] = t.[Padres] AND
[Tabla Minable - Bayes].[Nivel Padre] = t.[NivelPadre] AND
[Tabla Minable - Bayes].[Nivel Madre] = t.[NivelMadre] AND
[Tabla Minable - Bayes].[Orientacion Vocacional] = t.[OrientacionVocacional] AND
[Tabla Minable - Bayes].[Como Estudiante] = t.[ComoEstudiante] AND
[Tabla Minable - Bayes].[Uso Computadora] = t.[UsoComputadora] AND
[Tabla Minable - Bayes].[Uso Intenet] = t.[UsoComputadora] AND
[Tabla Minable - Bayes].[Oc Padre] = t.[OcPadre] AND
[Tabla Minable - Bayes].[Oc Madre] = t.[OcMadre] AND
[Tabla Minable - Bayes].[Dep Economica] = t.[DepEconomica] AND
[Tabla Minable - Bayes].[Personas Contribuyen] = t.[promedioAprobatorio] AND
[Tabla Minable - Bayes].[Ing Mensual] = t.[IngMensual] AND
[Tabla Minable - Bayes].[Transporte] = t.[Transporte] AND
[Tabla Minable - Bayes].[Tiempo Transporte] = t.[TiempoTransporte] AND
[Tabla Minable - Bayes].[Clase Socieconomica] = t.[claseSocieconomica] AND
[Tabla Minable - Bayes].[Id Alumno] = t.[idAlumno] AND
[Tabla Minable - Bayes].[Edad D] = t.[edad]
WHERE
[Tabla Minable - Bayes].[Asistio Tut Grupal] = 'Si'

```

Al procesar lo anterior se obtiene:

Probabilidad	Asistira	idAlumno
0.601575440516844	SI	17
0.778329348777693	SI	25
0.82062764406391	SI	27
0.983472446726411	SI	30
0.798314183254225	SI	31
0.736725344906767	SI	32
0.918566808888997	SI	33

Figura Anexo 1.4 Muestra de la predicción

	idAlumno	AsistioTutGrupal
1	17	SI
2	25	NO
3	27	SI
4	30	SI
5	31	SI
6	32	SI
7	33	SI

Figura Anexo 1.5 Valores reales

La figura Anexo 1.4 es solo una muestra de los valores obtenidos y en ella podemos observar los valores deseados e indicados previamente. La figura Anexo 1.5 por su parte muestra los valores reales almacenados en la base, es posible observar entonces que la única equivocación ocurrió con el alumno del id 25, ya que el no asistió a la tutoría y el modelo hubiera predicho que si

Recursos bibliográficos y mesográficos

- [1] María Ángela Cárdenas López, Francisco J. Trigo Tavera y María de Jesús Tron Fierros. La experiencia del programa de tutoría para la licenciatura, México, 2004
- [2] Brandenstein Mendez, Angelica. El sistema de tutoría como propuesta pedagógica en la Facultad de Ingeniería, análisis comparativo, México: El autor, 1993
- [3] Resumen COPADI. Mtra. Ana G. García y Colomé, México, 2010
- [4] Ing. Pablo García y Colomé, Lic. María Elena Cano Salazar, Lic. Ana G. García y Colomé y Lic. Anda Mará Vieyra Ávila. La tutoría en la Facultad de Ingeniería de la UNAM, México, Agosto de 2001
- [5] COPADI. <http://copadi.fi-c.unam.mx/copadi.jsp>
- [6] Secretaría de apoyo a la docencia. Primer Encuentro Nueva Era, México Agosto 2007
- [7] Página del programa Tutoría Nueva Era <http://www.ingenieria.unam.mx/~copadi/tutoria.html>
- [8] Silberschatz Abraham, Korth Henry. Fundamentos de bases de datos, Mc Graw Hill 4ª edición
- [9] José Hernández Orallo, Ma. José Ramírez Quintana, César Ferri Ramírez. Introducción a la minería de datos, Prentice Hall
- [10] Bellini Saibene, Yanina. Los Sistemas de Información y Bases de Datos en el proyecto RIAP, 2004
- [11] Apuntes de la materia Bases de Datos
- [12] Conceptos de minería de datos. <http://msdn.microsoft.com/es-es/library/ms174949%28v=sql.90%29.aspx>
- [13] Javier García de Jalón, José Ignacio Rodríguez, Jesús Vidal. Aprenda Matlab 7.0 como si estuviera en primero, Universidad Politécnica de Madrid, 2005
- [14] Diego García Morate. Manual de Weka, <http://www.dia.fi.uom.es/concha/SPAM/morate.pdf>
- [15] Liseth Abreu, Mónica. Análisis Exploratorio de Datos Multivariantes de Vigilantes Universitarios, Universidad de Los Andes Mérida, Venezuela
- [16] N. Guarín. Estadística Aplicada, Colombia 2002
- [17] CASTAÑER, OLGA. ¿Por qué no logro ser asertivo?. DESCLE. 2001.
- [18] Jorge Domínguez Menéndez y Edgar F. Timor Sánchez. La tutoría en el modelo pedagógico de continuidad de estudio: una propuesta instrumental, Cuba, 2005
- [19] Algoritmos de minería de datos <http://technet.microsoft.com/es-es/library/ms175595.aspx>
- [20] Microsoft SQL Server. <http://www.bing-vs-google.com/?q=sql+serverC>
- [21] Algoritmos de minería de datos. <http://msdn.microsoft.com/es-es/library/ms175595%28v=sql.90%29.aspx>
- [22] SQL Server Analysis Services - Minería de datos <http://technet.microsoft.com/es-es/library/bb510517.aspx>
- [23] SQL Server Management Studio y Business Intelligence Development Studio <http://msdn.microsoft.com/es-es/library/ms174170%28v=sql.90%29.aspx>
- [24] Herramientas de minería de datos <http://technet.microsoft.com/es-es/library/ms174467.aspx>
- [25] Árboles de decisión. <http://msdn.microsoft.com/es-es/library/ms175312.aspx7>
- [26] Selección de características. <http://technet.microsoft.com/es-es/library/ms175382.aspx>
- [27] Clústeres. <http://msdn.microsoft.com/es-es/library/ms174879.aspx>
- [28] Referencia técnica Clústeres. <http://msdn.microsoft.com/es-es/library/cc280445.aspx>
- [29] Asociación <http://msdn.microsoft.com/es-es/library/ms174916.aspx>
- [30] Referencia técnica Asociación. <http://msdn.microsoft.com/es-es/library/cc280428.aspx>
- [31] Bayes naive <http://msdn.microsoft.com/es-es/library/ms174806.aspx>
- [32] Redes Neuronales. <http://msdn.microsoft.com/es-es/library/ms174941.aspx>

- [33] Referencia técnica Redes Neuronales. <http://msdn.microsoft.com/es-es/library/cc645901.aspx>
- [34] Regresión Lineal. <http://msdn.microsoft.com/es-es/library/ms174824.aspx>
- [35] Andrew J. Brust. Programación avanzada con SQL Server 2005. Mc Graw Hill, 2005
- [36] Origen de datos <http://msdn.microsoft.com/es-es/library/ms170409.aspx>
- [37] Gráfico de elevación <http://msdn.microsoft.com/es-es/library/ms175428.aspx>
- [38] Revista de Informática Educativa y Medios Audiovisuales Vol. 1. Argentina, 2004
- [39] Publicación Journal of Social and Personal Relationships
- [40] Desarrollo Estratégico Integral, p. 1