



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

ANÁLISIS DEL CONSUMO DE
ESTUPEFACIENTES UTILIZANDO
CÓMPUTO NEURONAL

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Actuario

PRESENTA:

Julio César Ríos Dolores

TUTOR

Dr. Pedro Eduardo Miramontes Vidal

Ciudad Universitaria, CD. MX., 2017





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.

1. Datos del alumno

Ríos

Dolores

Julio

César

5576387473

Universidad Nacional Autónoma de
México

Facultad de Ciencias

Actuaría

413009094

2. Datos del tutor

Dr.

Pedro Eduardo

Miramontes

Vidal

3. Datos del sinodal 1

Dra.

Ruth Selene

Fuentes

García

4. Datos del sinodal 2

Dr.

José de Jesús

Galaviz

Casas

5. Datos del sinodal 3

Mat.

Margarita Elvira

Chávez

Cano

6. Datos del sinodal 4

L. en C.C.

Sergio

Hernández

López

7. Datos del trabajo escrito

Análisis del consumo de estupefacientes
utilizando cómputo neuronal

115

2017

Índice general

1. Consumo de drogas	6
1.1. Conceptos generales	6
1.1.1. Uso	7
1.1.2. Abuso	8
1.1.3. Adicción	9
1.1.4. Clasificación de las drogas	10
1.2. Impacto del consumo de drogas en la sociedad	11
1.2.1. Impactos en la salud	11
1.2.2. Impactos económicos	15
1.2.3. Violencia y crimen	16
1.2.4. Medio ambiente	17
1.3. Factores relacionados con el consumo	18
1.4. Prevención	20
1.5. Redes de neuronas artificiales y consumo de drogas	22
2. Redes de neuronas artificiales	24
2.1. Antecedentes	25
2.2. Neurona biológica	26
2.3. Neurona artificial	27
2.4. Tipos de redes de neuronas artificiales	38
2.4.1. Arquitectura	38
2.4.2. Tipo de entrenamiento	40

2.4.3. Tipo de conexión	43
2.5. Perceptrón Multicapa	44
3. Análisis de cúmulos	49
3.1. Conceptos generales	50
3.2. Medidas de proximidad	51
3.2.1. Distancia	52
3.2.2. Similitud	53
3.3. Métodos jerárquicos	54
3.3.1. Métodos jerárquicos aglomerativos	54
3.3.2. Métodos jerárquicos divisivos	58
3.3.3. Dendrograma	58
3.4. Métodos de partición	59
3.4.1. Algoritmo de las k -medias	60
4. Análisis de datos y ajuste de la red neuronal artificial	62
4.1. Base de datos	62
4.1.1. Codificación y recodificación de los datos	63
4.2. Reducción de dimensionalidad	71
4.2.1. Variables no relacionadas con el consumo	76
4.2.2. Análisis de cúmulos	78
4.3. Redes de neuronas artificiales utilizadas	86
4.3.1. Descripción de las redes	86
4.3.2. Entrenamiento	87
4.3.3. Resultados	92
5. Conclusiones	94
A. Cuestionario	97
Bibliografía	110

Agradecimientos

Quiero agradecer a mis padres Benito Alejandro Ríos Colín y Blanca Estela Dolores Hernández, a mi hermana Brenda Alejandra Ríos Dolores por su cariño y apoyo incondicional.

A mi tutor Pedro Eduardo Miramontes Vidal, a mis sinodales Ruth Selene Fuentes García, Margarita Elvira Chávez Cano, José de Jesús Galaviz Casas y Sergio Hernández López por brindarme su tiempo y darme los consejos necesarios para realizar este trabajo de la mejor manera posible, algo que aprecio infinitamente.

A la Universidad Nacional Autónoma de México por convertirse en mi segunda casa y ser parte fundamental dentro de mi formación.

Introducción

El consumo de estupefacientes representa una problemática importante a escala global debido a todos los sectores en los que este fenómeno impacta, es por esto que el buscar nuevas herramientas o enfoques para enfrentar este problema resulta necesario. Solo en el año de 2011 el número estimado de muertes relacionadas con el consumo de drogas era de 211,000, en donde la mayoría de éstas se asociaba a la población más joven de consumidores, pérdidas humanas que en gran medida pudieron haber sido evitadas. [1]

Esta cuestión y las problemáticas específicas que genera han sido ampliamente estudiadas desde distintos enfoques, es por esto que las herramientas que han ayudado para el entendimiento de este fenómeno han sido muy variadas. En el campo de la inteligencia artificial, se han utilizado redes de neuronas artificiales (RNA) para lograr identificar a los consumidores de ciertos estupefacientes. [2, 3] Este trabajo pretende, al igual que los antes mencionados, mostrar que las RNA son capaces de encontrar patrones en las características psicosociales y de personalidad de los consumidores, esto con el fin de identificarlos y ofrecer tratamiento de manera oportuna. Es importante destacar que el utilizar este tipo de herramientas ofrece la capacidad de trabajar con una gran cantidad de datos, lo que facilita la aplicación en sitios de interés como por ejemplo: escuelas, hospitales, etc.

En el capítulo 1 se aborda a grandes rasgos el tema referente a los drogas, se da una definición de los términos utilizados en este trabajo, además de una descripción general de las problemáticas que el consumo de estas sustancias ocasiona en la sociedad, también se abordan de manera muy general los trabajos referentes al tema en donde se utilizaron RNA.

Posteriormente, el capítulo 2 se refiere al análisis de cúmulos, una técnica estadística utilizada en este trabajo para el tratamiento preliminar de la base de datos, se abordan los conceptos necesarios para que el lector comprenda como es que este análisis fue aplicado.

Dado que la herramienta principal para realizar este estudio son las redes de neuronas artificiales, se dedica el capítulo 3 a las mismas, se da una breve introducción con respecto al desarrollo dentro de este campo, para después proceder a la descripción de los distintos tipos de redes de neuronas artificiales, su clasificación y los algoritmos utilizados para su entrenamiento, esto con el fin de familiarizar al lector con la herramienta.

Por último, el capítulo 4 recaba los resultados obtenidos en este trabajo, así como la interpretación de los mismos, para posteriormente dar las conclusiones finales.

Capítulo 1

Consumo de drogas

Históricamente, el hombre siempre ha consumido sustancias que alteran el funcionamiento normal del sistema nervioso central [4], sin embargo, este hecho ha generado un alto costo para todos los miembros de la sociedad.

El uso incorrecto de estas sustancias, además de ocasionar problemas a los usuarios, también repercute en sectores de la población que no están directamente relacionados con el consumo de los mismos, como veremos posteriormente, esta cuestión tiene impactos económicos, de salud, ambientales, relacionados con el crimen y muchos otros que hacen de este tema una prioridad.

1.1. Conceptos generales

Debido a los cambios políticos, ambientes sociales y el estado legal de ciertas drogas, el desarrollo de los conceptos en el tema de adicción ha sido complicado. [5]

Según la organización mundial de la salud (OMS) el término droga en farmacología se refiere a toda sustancia química que modifica los procesos fisiológicos y bioquímicos de los tejidos o los organismos [6]. Dada su ambigüedad para este trabajo también se definirán los términos, estupefaciente y sustancia psicoactiva.

- Estupefaciente: según la real academia española un estupefaciente es una

sustancia que altera la sensibilidad y puede producir efectos estimulantes, deprimentes, narcóticos o alucinógenos, y cuyo uso continuado crea adicción. En la convención única de 1961 sobre estupefacientes se plantea un listado el cual nos brinda un marco de referencia para saber con exactitud si una sustancia puede o no ser clasificada dentro de esta categoría. [7]

- Sustancia psicoactiva: Es toda sustancia que actúa sobre el sistema nervioso, alterando las funciones psíquicas (RAE). Al igual que en los estupefacientes, existe una lista de las sustancias que pueden clasificarse como sustancias psicoactivas.[8]

Otros términos clave dentro de este tema son el uso, el abuso y la adicción, muchas veces éstos son confundidos entre sí o mal empleados, es por esto que serán abordados con mayor detalle a continuación.

1.1.1. Uso

En este caso, el consumo de la sustancia es esporádico y debido a la baja frecuencia con la que se realiza no genera problemas notables de salud (sin tomar en cuenta los casos de sobredosis) y tampoco dependencia. [9]

Dentro de esta categoría pueden ser clasificados el consumo experimental, recreativo y habitual.

Consumo experimental: El individuo prueba una o más sustancias, una o dos veces, y decide no volverlo a hacer. No tiene intenciones de hacerlo de nuevo porque tal vez los efectos no fueron placenteros, no le ve un sentido a continuar haciéndolo, o tal vez porque no va de acuerdo a su estilo de vida o forma de pensar. En este caso el individuo ha saciado su curiosidad y tiene otras prioridades para su vida. [10]

Consumo recreativo: Se empieza a presentar el consumo con cierta regularidad entre círculos de amigos, en contextos de ocio. No compra ni guarda, lo hace con gente que conoce y confía solo cuando siente deseos de hacerlo. Se da como un consumo pasajero asociado a situaciones concretas, y en algunos casos a épocas o

etapas de la vida sin dejar consecuencias evidentes. Rara vez se pierde el control, no consume cuando está solo, ni se tiende a experimentar con sustancias o vías de administración de alto riesgo. [10]

Consumo habitual: Para llegar aquí se tuvo que haber pasado por alguna fase previa. Se tiene una droga de preferencia (quizás después de probar con muchas otras), un refuerzo permanente y un gusto por lo que siente. Por eso la frecuencia y la cantidad aumentan. La droga juega un papel importante en la vida y cumple funciones cada vez que la consume, por lo que es muy probable que lo haga repetidamente. Dependiendo de la droga puede desarrollar una dependencia física o psicológica, y tarde o temprano empezará a tener problemas cada vez más evidentes. [10]

El uso representa las primeras etapas dentro del consumo de sustancias psicoactivas, en este caso el consumidor, como ya vimos, puede apartarse del problema o adentrarse más lo cual lleva al abuso y probablemente a la adicción.

1.1.2. Abuso

Se define como a un modelo desadaptativo de uso de una sustancia psicoactiva caracterizado por un consumo continuado, a pesar de que el sujeto sabe que tiene un problema social, laboral, psicológico o físico, persistente o recurrente, provocado o estimulado por el consumo recurrente en situaciones en las que es físicamente peligroso. [6]

En este caso pueden ser enmarcados el consumo compulsivo y el consumo problemático.

Consumo compulsivo: Se depende de la droga pero no sólo de ella sino de los amigos y del estilo de vida que se desarrolla en torno a ella. Es probable que ese “estilo de vida” no sea sólo el consumo sino el círculo de amigos, el reconocimiento y aceptación dentro del grupo, pero puede desarrollar también un consumo solitario y aislado, sin amigos, y sin mayores actividades. Aquí el uso de la sustancia se da por necesidad y muchas veces para evitar sentirse mal física o

psicológicamente. [10]

Consumo problemático. Está determinado por el impacto que éste tiene sobre la salud (consecuencias visibles, enfermedades crónicas, etc.) y la aparición de consecuencias de tipo social con una connotación negativa (deterioro, aislamiento, conductas problemáticas, etc.). Se caracteriza porque se presenta el uso de una o varias drogas en situaciones que pueden significar peligro: pérdida de velocidad de respuesta y reflejos, dificultades o incapacidad para cumplir con las responsabilidades laborales, académicas o familiares. [10]

En general, podemos decir que en esta etapa ya se tienen consecuencias notables en la vida del consumidor y la de las personas que lo rodean, aunque aún no está en la etapa de adicción, las repercusiones sociales y físicas empezaran a generar un cambio drástico en el ritmo de vida.

1.1.3. Adicción

Según la Organización Mundial de la Salud (OMS), la adicción puede ser definida como el consumo repetido de una o varias sustancias psicoactivas, hasta el punto de que el consumidor (denominado adicto) se intoxica periódicamente o de forma continua, muestra un deseo compulsivo de consumir la sustancia (o las sustancias) preferida, tiene una enorme dificultad para interrumpir voluntariamente o modificar el consumo de la sustancia y se muestra decidido a obtener sustancias psicoactivas por cualquier medio. Por lo general, hay una tolerancia acusada y un síndrome de abstinencia que ocurre frecuentemente cuando se interrumpe el consumo. La vida del adicto está dominada por el consumo de la sustancia, hasta llegar a excluir prácticamente todas las demás actividades y responsabilidades. Es importante resaltar dos términos en esta definición: tolerancia y síndrome de abstinencia.

Tolerancia: se define como una disminución de la respuesta a una dosis concreta de una droga o un medicamento que se produce con el uso continuado [6], ésta puede clasificarse dentro de las siguientes tres categorías tolerancia directa, tolerancia

inversa y tolerancia cruzada:

Tolerancia directa: estado de adaptación que se caracteriza por la disminución de la respuesta a una misma cantidad de droga o al requerimiento de dosis mayores para obtener el mismo efecto. [11]

Tolerancia inversa: estado en el que se producen los mismos efectos en el organismo con dosis menores. Típicamente se da en el abuso del alcohol. [11]

Tolerancia cruzada: forma de tolerancia que aparece con una droga y con otras del mismo tipo o en ocasiones con otras diferentes. Típicamente aparece entre la heroína y la morfina y entre el alcohol y los barbitúricos.[11]

Síndrome de abstinencia: es un conjunto de síntomas con grado de intensidad que aparecen al suspender o reducir el consumo de una sustancia psicoactiva que se ha consumido de forma repetida, habitualmente durante un período prolongado o en dosis altas. [6]

En este caso existe una clara dependencia a la sustancia la cual puede ser clasificada en dependencia física y dependencia psicológica.

Dependencia física implica un cambio permanente en el funcionamiento del cuerpo y del cerebro, se da cuando ya se ha generado una tolerancia hacia la sustancia, por lo tanto el cuerpo cada vez necesita una dosis mayor para sentir los efectos deseados.

Dependencia psicológica: la producen todas las sustancias a excepción de los hongos alucinógenos. La privación de la sustancia produce malestar, angustia, irritabilidad y depresión, por lo tanto para evitar estos malestares se busca la manera de consumir permanentemente. Se piensa que “no se puede vivir” sin consumir y todo lo que esto implica. [10]

1.1.4. Clasificación de las drogas

Para clasificar las drogas se pueden seguir diferentes criterios, por ejemplo, por su origen (naturales, sintéticas, semisintéticas), por su estructura química, por su acción farmacológica, por el medio socio-cultural (legales-ilegales, institucionalizadas-

no institucionalizadas, duras-blandas, más peligrosas-menos peligrosas) o con respecto al sujeto, utilizándose en este caso clasificaciones clínicas. [12]

Dados los múltiples criterios de clasificación, en este trabajo solo se abordará aquella que se basa en los efectos producidos por la sustancia, ésta agrupa a las drogas en tres grandes grupos:

- Estimulantes: sustancias que activan el Sistema Nervioso Central (SNC) en esta categoría se encuentran drogas como la cocaína, anfetaminas, cafeína, etc.
- Depresores: drogas que disminuyen el grado de actividades del SNC, ejemplos de este tipo de droga son: alcohol, barbitúricos, benzodiacepinas, etc.
- Psicodélicos: sustancias que producen alteraciones perceptivas, en este caso se encuentra la dietilamida de ácido lisérgico (LSD), psilocibina, ketamina, etc.

1.2. Impacto del consumo de drogas en la sociedad

1.2.1. Impactos en la salud

El consumo de sustancias psicoactivas puede producir un gran número de trastornos del tipo físico y mental [12]. En general, esta práctica no solo representa un problema para los consumidores, las consecuencias del uso de drogas también se extienden a la familia del usuario y a la comunidad.

Los efectos negativos producidos en el consumidor varían de acuerdo al tipo de sustancia, es por eso que a continuación se dará una descripción muy general de los efectos producidos por cada una en específico.

- Nicotina: el humo del tabaco aumenta el riesgo de un consumidor de contraer cáncer, enfisema, trastornos bronquiales y enfermedades cardiovasculares. La tasa de mortalidad asociada con la adicción al tabaco es asombrosa.

El consumo de tabaco acabó con la vida de aproximadamente 100 millones de personas durante el siglo XX y, si las tendencias actuales de tabaquismo continúan, se ha proyectado que el total de muertes en este siglo alcanzará los mil millones. [13]

- Alcohol: su consumo puede dañar el cerebro y la mayoría de los órganos del cuerpo. Las áreas del cerebro que son especialmente vulnerables a los daños relacionados con el alcohol son la corteza cerebral, el hipocampo y el cerebelo, también puede acarrear un gran número de alteraciones pancreáticas [14], patologías graves como la miocardiopatía alcohólica [15], hipertensión arterial [16], arritmias [17], etc.
- Marihuana: es la sustancia ilegal que más se consume. Esta droga deteriora la memoria a corto plazo y el aprendizaje, la capacidad de concentración y la coordinación. También aumenta la frecuencia cardíaca, puede dañar los pulmones y puede aumentar el riesgo de psicosis en quienes padecen una vulnerabilidad subyacente. [13]
- Medicamentos por prescripción médica (en este grupo se encuentran los analgésicos opiáceos, los sedantes ansiolíticos y los estimulantes): el mal uso o el abuso de estos medicamentos puede conducir a la adicción, e incluso, en algunos casos, a la muerte. Por ejemplo, con frecuencia se abusa de los analgésicos opiáceos al aplastarlos e inyectarlos o inhalarlos, aumentando así en gran medida el riesgo de sobredosis.
- Inhalantes: son sustancias volátiles que se encuentran en muchos productos para el hogar, tales como limpiadores para hornos, gasolina, pinturas en aerosol y otros aerosoles, que inducen efectos que alteran la mente. Los inhalantes son sumamente tóxicos y pueden dañar el corazón, los riñones, los pulmones y el cerebro. Incluso una persona sana puede sufrir insuficiencia cardíaca y la muerte en cuestión de minutos con una sola sesión de inhalación prolongada de un inhalante. [13]

- **Cocaína:** es un estimulante de acción corta, que puede llevar a los usuarios a tomar la droga muchas veces en una sola sesión. El consumo de cocaína puede conducir a graves consecuencias médicas relacionadas con el corazón y los sistemas respiratorio, nervioso, digestivo etc. [13]
- **Anfetaminas** (en este grupo se incluye a la metanfetamina): son estimulantes poderosos que producen sensaciones de euforia y estado de alerta. Los efectos de la metanfetamina son especialmente duraderos y perjudiciales para el cerebro. Las anfetaminas pueden elevar la temperatura corporal y pueden producir problemas graves en el corazón y convulsiones. [13]
- **MDMA:** produce tanto efectos estimulantes como alucinógenos. Puede aumentar la temperatura corporal, la frecuencia cardíaca, la presión arterial y el estrés de la pared cardíaca, también puede ser tóxica para las células nerviosas. [13]
- **LSD:** es una de las drogas alucinógenas, o que alteran la percepción, más potentes. Sus efectos son impredecibles y los que la consumen pueden ver imágenes y colores vívidos, oír sonidos y percibir sensaciones que parecen reales, pero que no existen. Los consumidores también pueden tener experiencias traumáticas y emociones que pueden durar muchas horas. [13]
- **Heroína:** es un medicamento opiáceo potente que produce euforia y sentimientos de relajación. Enlentece la respiración, y su uso está ligado a un aumento del riesgo de enfermedades infecciosas graves, en especial cuando se consume por vía intravenosa, lo cual aumenta en gran medida la probabilidad de contraer hepatitis B y C, VIH o una variedad de otros virus transmitidos por la sangre. [13]
- **Esteroides:** si bien se los puede recetar para determinadas afecciones médicas, se abusa de ellos para aumentar la masa muscular y mejorar el rendimiento atlético o la apariencia física. Entre las graves consecuencias del

abuso podemos mencionar acné severo, enfermedades del corazón, problemas en el hígado, accidentes cerebrovasculares, enfermedades infecciosas, depresión y suicidio. [13]

Como ya habíamos mencionado además del impacto en la salud del usuario, el consumo de este tipo de sustancias repercute directamente en su entorno, algunas de las consecuencias más notables son:

- Efectos negativos de la exposición prenatal a las drogas en bebés y niños: el abuso de heroína u opiáceos recetados por parte de la madre durante el embarazo puede provocar un síndrome de abstinencia (llamado síndrome de abstinencia neonatal, o NAS) en su bebé. También es probable que algunos niños expuestos a las drogas vayan a necesitar apoyo educativo en la escuela que les permita superar lo que pueden ser déficits sutiles en las áreas de desarrollo como el comportamiento, la atención y el pensamiento. [13]
- Efectos negativos del humo de tabaco de segunda mano: el humo del tabaco de segunda mano, también llamado humo de tabaco en el ambiente (ETS), es una fuente importante de exposición a un gran número de sustancias nocivas para la salud humana, en particular para los niños. De acuerdo con el *Surgeon General's 2006 Report, The Health Consequences of Involuntary Exposure to Tobacco Smoke*, la exposición involuntaria al humo de tabaco en el ambiente aumenta el riesgo de padecer enfermedades cardíacas en personas que nunca han fumado entre un 25 % y un 30 %, y cáncer de pulmón entre un 20 % y un 30 %. [13]
- Aumento de la propagación de las enfermedades infecciosas: la inyección de drogas como la heroína, la cocaína y la metanfetamina contribuye actualmente a alrededor del 12 por ciento de los nuevos casos de SIDA. El uso de drogas inyectables también es un factor importante en la propagación de la hepatitis C, una enfermedad grave del hígado, potencialmente mortal. El consumo de drogas inyectables no es la única manera en que el abuso

de drogas contribuye a la propagación de enfermedades infecciosas. Todas las drogas adictivas causan alguna forma de intoxicación, que interfiere con el juicio y aumenta la probabilidad de comportamientos sexuales de riesgo. Esto a su vez contribuye a la propagación del VIH/SIDA, la hepatitis B y C y otras enfermedades de transmisión sexual. [13]

1.2.2. Impactos económicos

El consumo y producción de drogas ilícitas tiene un claro impacto en la economía. El gobierno destina una parte de los recursos públicos a áreas de salud y justicia relacionadas directa o indirectamente con este fenómeno, sin embargo, los recursos invertidos para estos fines, bien podrían ser ocupados en otros sectores como educación, infraestructura, etc.

Por otro lado, los individuos que son ingresados a algún centro de tratamiento o a prisión por un crimen relacionado con drogas representan una pérdida desde el punto de vista productivo, esto ya que son un sector de la población que es retirado de la fuerza laboral. Cabe destacar que los convictos implican aún más gastos ya que se invierten recursos para mantenerlos tras las rejas. [18]

Existen estudios que tratan de estimar los costos que el uso, abuso y adicción de drogas representa para la sociedad, un ejemplo es el estudio realizado por el Centro de Adicción y Abuso de Sustancias en la Universidad de Columbia, en donde solo en Nueva York el costo en 1994 rebasa los 20 mil millones de dólares. Según los cálculos se invierten \$5.1 mil millones en atención de salud; \$4.9 mil millones en la productividad perdida de la economía de la ciudad; \$4.1 mil millones en policías, tribunales, prisiones y el resto del sistema de justicia penal; \$3,5 mil millones en servicios sociales públicos y privados incluyendo bienestar; \$1.6 millones en aumento de los costos de seguros, seguridad y compensación a los trabajadores; Y \$835 millones en costos de propiedad. Es importante notar que de esta cantidad impresionante de recursos sólo se gastaron \$735 millones (3.7 %) en el tratamiento del abuso de drogas y la adicción, y \$80 millones (0.4 %) para prevenirlo. [5]

Como vemos un problema importante es la baja inversión en investigación, tratamiento y prevención, cuestiones que se ha demostrado tienden a obtener mejores resultados y un mejor rendimiento con respecto a los costos que el problema ocasiona. [5]

1.2.3. Violencia y crimen

Las drogas tienen una relación muy estrecha con el crimen y la violencia [19], este fenómeno durante mucho tiempo ha sido ampliamente documentado y representa una problemática que debe ser atendida a la brevedad.

Según Goldstein los crímenes relacionados con las drogas pueden ser clasificados en tres grupos: crímenes psicofarmacológicos, crímenes económico-compulsivos y crímenes sistemáticos. [19]

Crímenes psicofarmacológicos

Éstos son cometidos por individuos que se encuentran bajo la influencia de sustancias psicoactivas, las cuales pueden ocasionar en los usuarios estados de excitabilidad e irracionalidad. Las sustancias más destacadas en esta cuestión son el alcohol, estimulantes, barbitúricos y la fenciclidina (PCP). [19]

Dentro de esta categoría se engloban crímenes no solo relacionados con el agresor, sino también con la víctima, ya que ésta puede ser más vulnerable en un estado de intoxicación dado que su nivel de respuesta se ve disminuido. Existen estudios que evidencian este hecho, por ejemplo, investigaciones anteriores indican frecuencias relativamente altas del consumo de alcohol en la violación [20, 21] y las víctimas de homicidio. [22]

Los datos estadísticos sobre este tema no suelen ser tan precisos debido a que muchas veces no se reportan, o son mal registrados por las autoridades [19], esto dificulta el estudio del problema dado que no tenemos una percepción de sus dimensiones reales.

Crímenes económico-compulsivos

Estos son crímenes cometidos con el fin de obtener drogas o recursos para poder solventar la adicción. La heroína y la cocaína, dado su precio y los patrones de consumo son las sustancias más relevantes dentro de esta categoría. [19]

La violencia en estos crímenes está caracterizada por factores relacionados con el contexto del mismo, por ejemplo: el nivel de nerviosismo del agresor, algún tipo de arma que lleve consigo tanto el delincuente como la víctima en el caso de asalto, la intervención de algún tercero, etc. [19]

En general, la víctima de estos crímenes podría ser cualquier individuo, sin embargo, existe evidencia de que personas que viven en la misma zona que los agresores suelen ser más propensas a este tipo de ataques. [23]

Crímenes sistemáticos

El Observatorio Europeo de las Drogas y las Toxicomanías define este tipo de crímenes como aquellos que son cometidos dentro del funcionamiento de los mercados de drogas ilícitas, esto como parte del suministro, distribución y uso de las mismas. Algunos ejemplos de este tipo de crímenes son las disputas por los territorios de venta de droga, robos a los traficantes de drogas y las represalias violentas que esto genera, venganza por venta de droga adulterada o de mala calidad, corrupción relacionada al tráfico de drogas ilícitas etc.

Dada la ilegalidad de este negocio y la intensa competencia por beneficiarse de las exorbitantes cantidades de dinero que éste produce, los traficantes se valen del terror, la violencia, la extorsión, el soborno, o cualquier otra estrategia que les resulte eficaz. [24]

1.2.4. Medio ambiente

Un tema poco tratado con respecto al impacto del consumo de estupefacientes es el que se refiere a los daños ambientales que éste produce.

Como es de suponerse la fabricación de estupefacientes produce desechos, los

cuales terminan en lugares clandestinos, y dada su toxicidad afectan de manera clara el ecosistema, un ejemplo de este proceso de contaminación puede verse con la fabricación de metanfetamina en donde por cada libra que se fabrica se producen de 5 a 7 libras de desechos tóxicos que usualmente van a parar a arroyos, campos, alcantarillas, etc. [18]

Por otro lado, además de la fabricación sintética de estupefacientes, el cultivo ilegal de ciertas plantas representa una actividad altamente perjudicial para el ecosistema, ya que el uso de múltiples insecticidas, repelentes químicos y venenos llegan a contaminar mantos acuíferos matando peces y otro tipo de vida silvestre. No solo la producción representa un problema, también lo es la erradicación, una sustancia muy utilizada para eliminar los plantíos ilegales es el glifosato, un herbicida de amplio espectro, que no solo ataca la plantación ilegal sino a toda la vegetación de la zona alterando el equilibrio del ecosistema. [25]

Los cuerpos de agua son considerados como los más susceptibles a la contaminación por drogas ilícitas y sus desechos, la presencia de estas sustancias incluso en bajas concentraciones puede causar efectos tóxicos en organismos acuáticos, además de una muy amplia variedad de problemas ambientales y de salud. [26]

1.3. Factores relacionados con el consumo

El consumo de sustancias psicoactivas no es un fenómeno que pueda ser explicado por la ausencia o presencia de un determinado factor en la vida del consumidor, en este caso interviene una gran cantidad de variables, que deben ser estudiadas no solo de manera individual, sino también cómo es que éstas se relacionan.

A las variables que incrementan la probabilidad de consumo o abuso de drogas se le conoce como factores de riesgo [27], éstos pueden ser clasificados de acuerdo a muchos criterios, el que se abordará en este trabajo se divide en dos categorías: factores individuales y factores sociales. [28]

Los factores individuales se refieren a las características del individuo (edad, sexo, personalidad, etc.) que determinan un cierto grado de vulnerabilidad ante las

influencias sociales que motivan al uso de este tipo de sustancias. [29]

Aunque el consumo de sustancias psicoactivas puede ocurrir a cualquier edad, por lo general se empieza a experimentar con éstas en la adolescencia [28], una etapa muy importante para el desarrollo de la persona, tanto a nivel físico, emocional, social, académico, etc. [12]

El sexo es otro factor que puede influir en este fenómeno. Existen estudios en los que se muestra una diferencia con respecto al consumo de ciertas sustancias entre hombres y mujeres [30], sin embargo, también se ha visto que las desigualdades han ido disminuyendo e incluso en algunas sustancias el sexo se ha convertido en una variable poco significativa para explicar el consumo. [31]

La personalidad también juega un papel importante al momento de establecer una postura con respecto al uso de drogas. La búsqueda de nuevas sensaciones, baja autoestima, rebeldía, inseguridad, creencias, conductas problemáticas y el conocimiento con respecto a los daños ocasionados influyen al momento de decidir si se usaran o no estas sustancias. [32]

Los factores sociales se refieren a aquellas variables relacionadas con el entorno en el que se desenvuelve el sujeto y que influyen en su decisión con respecto al uso de sustancias psicoactivas, aquí se engloban los factores de riesgo externos al individuo los cuales pueden ser clasificados dentro de dos niveles, microsocioal y macrosocioal. [29]

El nivel microsocioal se refiere a los grupos en donde el sujeto tiene una participación directa (familia, amigos, escuela etc.), aquí se engloban las relaciones interpersonales y los modelos de comportamiento a los que debe ajustarse el individuo. [29]

La familia ejerce una gran influencia sobre el consumo de drogas en sus miembros, ésta puede incitar, neutralizar o inhibir este tipo de conductas. [33]. El uso de alcohol y drogas por parte de los padres, actitudes positivas hacia el uso de drogas, prácticas pobres e inconsistentes de manejo familiar y conflictos familiares son factores que pueden influir de manera negativa en los miembros de la familia favoreciendo a la aparición de problemas como por ejemplo el uso o abuso de

drogas. [12]

Los amigos son una parte muy importante en la vida de un adolescente, es por esto que en algunos casos su influencia puede conducir al consumo de drogas. Muchas veces el consumo de drogas se inicia en el grupo de amigos, por presión o por la necesidad de aceptación [29], en general, una persona que mantiene una relación estrecha con un grupo que muestra actitudes favorables hacia el uso de drogas tiene más probabilidades de iniciarse en el consumo. [12]

La insatisfacción escolar, un nivel más bajo de compromiso con las actividades académicas, un mayor nivel de absentismo escolar y una peor ejecución académica son factores que se relacionan con un conjunto de problemáticas entre las que destaca el consumo de drogas. [12]

En el nivel macrosocial se agrupan los factores de carácter socio-estructural, socio-económico, socio-cultural, que condicionan la calidad de vida del individuo. La influencia de los medios de comunicación, la oferta y disponibilidad de las sustancias, el estilo de vida, y la percepción general con respecto al uso de drogas, son factores ubicados en este nivel que influyen en el desarrollo de este fenómeno. [29]

1.4. Prevención

Uno de los métodos más eficaces para atacar las problemáticas que el consumo de sustancias psicoactivas ocasiona es la prevención [12]. Ésta podría definirse como el conjunto de acciones dirigidas a identificar, evitar, reducir, regular o eliminar el consumo no terapéutico de sustancias psicoactivas, así como sus consecuencias físicas, psíquicas, económicas, familiares y sociales. [34]

La prevención refiriéndonos al tema de consumo de sustancias psicoactivas puede ser clasificada en 3 categorías:

- Prevención indicada: es la intervención que se dirige a grupos de población con sospecha de consumo, usuarios experimentales, o de quienes exhiben

factores de alto riesgo que incrementan la posibilidad de desarrollar consumo perjudicial o adicción. [34]

- **Prevención selectiva:** es la que se enfoca a desarrollar intervenciones para grupos de riesgo específico, por ejemplo: hijos de alcohólicos, reclusos, menores infractores, víctimas de violencia doméstica y abuso sexual, niños con problemas de aprendizaje o de conducta, etc. Estos grupos se asocian al consumo de drogas ya que se han identificado factores biológicos, psicológicos y ambientales que sustentan la vulnerabilidad. Asimismo contribuye a la atención de grupos específicos de niños en situación de calle, indígenas y adultos mayores, entre otros. [34]
- **Prevención universal:** es la que está dirigida a la población en general y se lleva a cabo mediante la promoción de la salud para crear conocimiento y orientar sobre la problemática del consumo de sustancias y las formas de prevención. [34]

En general, lo que se pretende con estas medidas es: [12]

- Retrasar la edad de inicio del consumo de drogas.
- Limitar el número y tipo de sustancias utilizadas.
- Evitar la transición de la prueba de sustancias al abuso y dependencia de las mismas.
- Disminuir las consecuencias negativas del consumo en aquellos individuos que consumen o tienen problemas de abuso o dependencia de las drogas.
- Educar a los individuos para que sean capaces de mantener una relación madura y responsable con las drogas.
- Potenciar los factores de protección y disminuir los de riesgo para el consumo de drogas.

- Modificar las condiciones del entorno socio-cultural y proporcionar alternativas de vida saludables.

La inversión en el campo de la prevención es muy pequeña, por ejemplo, en México por cada dólar invertido en reducción de la demanda, se gastan 16 en el control de la oferta [35], sin embargo, es primordial que aumente la cantidad de recursos destinada a este campo ya que como mencionamos antes es uno de los métodos más eficaces para atacar las problemáticas que el uso, abuso, y adicción ocasionan.

1.5. Redes de neuronas artificiales y consumo de drogas

Como ya vimos el consumo de drogas representa un problema muy complejo, en el que muchos factores están involucrados y cuyas consecuencias impactan de manera considerable a la sociedad, es por esto que muchas áreas han tratado de contribuir en la búsqueda de una solución abordando el problema desde distintos enfoques y con múltiples herramientas, las redes de neuronas artificiales (RNA) son una herramienta desarrollada dentro del campo de la inteligencia artificial que ha sido utilizada para el análisis de este fenómeno.

Existen estudios que muestran como las RNA son capaces de identificar (con cierto grado de error) al consumidor de alguna sustancia (tabaco, éxtasis, heroína etc.) basándose en un conjunto de variables del tipo psicosocial y de personalidad [36, 2, 3], esto es de gran utilidad ya que el detectar el consumo a tiempo nos brinda la capacidad de ofrecer tratamiento de manera oportuna.

Esta herramienta no solo se ha limitado al estudio de los consumidores, también ha tratado de atacar el problema de venta o tráfico de drogas, por ejemplo, se ha utiliza una red de neuronas artificiales para predecir el aumento en la venta de drogas en alguna zona específica [37]. También se ha utilizado esta herramienta para identificar aquellos condados vulnerables a un aumento en la producción de

metanfetamina [38]. Como vemos, estos trabajos están orientados a detectar el problema en etapas tempranas, lo cual es primordial para poder enfrentarlo. El trabajar con este tipo de herramientas nos da la capacidad de manejar una gran cantidad de datos de manera automatizada lo cual disminuye los costos y el tiempo invertido, sin embargo no hay que olvidar que los resultados obtenidos por las RNA tienen un grado de error a considerar.

Capítulo 2

Redes de neuronas artificiales

El cerebro humano está compuesto por alrededor de 100 mil millones (10^{11}) de neuronas altamente interconectadas y su capacidad para procesar información resulta sorprendente, por mucho tiempo se ha tratado de entender su funcionamiento y cómo es que éste es capaz de realizar tareas tan complejas, al conjunto de disciplinas que se dedican al estudio del cerebro y en general del sistema nervioso se les denominan neurociencias.

El cerebro actúa como un sistema de procesamiento de información en paralelo, en el cual muchos elementos simples (neuronas) comparten el trabajo. Debido al gran número de neuronas involucradas en los procesos cerebrales, la contribución hecha por un sólo elemento no es tan importante, así que si alguna neurona comete algún error es poco probable que esto afecte de manera significativa a las demás, lo cual implica tolerancia a errores, a este tipo de distribución de trabajo se le conoce como proceso distribuido. [39]

La estructura del cerebro y cómo es que éste funciona ha inspirado al desarrollo de herramientas que buscan emular su comportamiento con el fin de solucionar problemas que resultan difíciles de abordar con las herramientas existentes hasta el momento, pero que el cerebro resuelve con facilidad, una de estas herramientas son las redes de neuronas artificiales.

Las redes de neuronas artificiales tratan de extraer las características más genera-

les del funcionamiento del cerebro y basan su funcionamiento en cuatro premisas: [40]

1. El procesamiento de información ocurre en muchos elementos simples llamados neuronas.
2. Se pasan señales entre neuronas a través de sus enlaces de conexión.
3. Cada enlace de conexión tiene un peso asociado, el cual, en las redes de neuronas típicas, es multiplicado por la señal transmitida.
4. Cada neurona aplica una función de activación a la señal de entrada total de la red, la cual es la suma de las señales de entrada multiplicada por el peso asociado, para determinar la señal de salida.

2.1. Antecedentes

Las RNA son una herramienta clasificada dentro de la inteligencia artificial que comenzó a desarrollarse en los años 40 donde destaca el trabajo realizado por Warren McCulloch y Walter Pitts los cuales en el texto *A logical calculus of the ideas immanent in nervous activity* [41] publicado en 1943, lanzan una teoría acerca del funcionamiento de las neuronas y la relación entre ellas, éste da lugar a la creación de un modelo que hoy conocemos como la neurona McCulloch-Pitts el cual es considerado como el padre del cómputo neuronal moderno [39]. Posteriormente, en 1949 es publicado por Donald O. Hebb *The organization of behavior* [42], el cual, además de muchos otros temas, trata de explicar cómo es que ocurren los procesos de aprendizaje, esto último representó un paso en el avance de las redes neuronales artificiales ya que con base en esas ideas se logró estructurar una forma de aprender para las RNA.

Otro personaje destacado dentro del desarrollo de este campo fue Frank Rosenblatt el cual en 1962 con su libro *Principles of neurodynamics* [43] presenta el perceptrón, el cual resulta ser un modelo matemático de una neurona biológi-

ca que a pesar de su simplicidad, mediante un algoritmo de aprendizaje lograr comportarse como un clasificador lineal.

Sin embargo, todo el desarrollo en la investigación de las redes de neuronas artificiales se vio frenado con la aparición del libro *Perceptrons* [44] en el año de 1969, en este libro se demuestran las grandes limitaciones del perceptrón, lo cual desalentó mucho la investigación en el campo. En 1982 John Hopfield propuso un modelo, el cual consistía en un conjunto de neuronas, las cuales estaban altamente interconectadas, es decir, cada neurona se conectaba con las otras, conocido hoy en día como la red de Hopfield, sin embargo, no fue este modelo el que despertó nuevamente el interés de los investigadores en las RNA, sino el preceptrón multicapa propuesto por Rumelhart y McClelland en 1986.[39]

Los avances dentro de este campo en los últimos años se han potenciado debido al gran desarrollo computacional que se ha logrado, esto ha permitido diseñar redes cada vez más sofisticadas y ambiciosas que logran abordar una extensa y variada cantidad de problemas.

2.2. Neurona biológica

Las neuronas son células clasificadas dentro del sistema nervioso las cuales se componen de un cuerpo celular llamado soma, dendritas y un axón el cual cuenta con una o dos ramas llamadas colaterales que a su vez cuentan con ramificaciones denominadas telodendron (figura 2.1). Sus funciones consisten en transmitir impulsos de los receptores periféricos (la piel y otros órganos de los sentidos) al sistema nervioso central (SNC), transmitir impulsos del SNC a los efectores (músculos, glándulas u otros órganos) y transmitir impulsos entre ellas mismas. [45]

Al proceso en donde la prolongación del axón de una neurona entra en contacto con las dendritas de otra se le denomina sinapsis. Este proceso comienza con una descarga química que produce un impulso nervioso en la neurona denominado potencial de acción, éste es transmitido a través del axón que al llegar a las

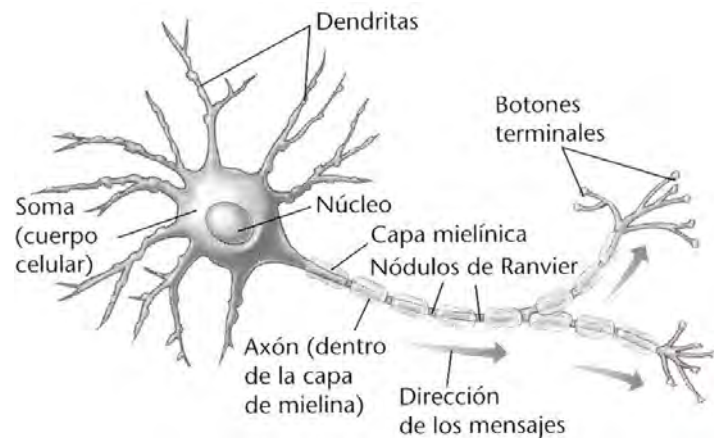


Figura 2.1: Características básicas de una neurona biológica.

prolongaciones del telodendron provoca la liberación de neurotransmisores, los cuales tendrán un efecto excitatorio o inhibitorio en la neurona receptora (post sináptica).

Las neuronas cuentan también con un umbral el cual representa la condición que debe cumplir el estímulo recibido por la neurona para que ésta logre generar un nuevo impulso nervioso. Dada la interconectividad de éstas, el estímulo recibido puede provenir de una o más neuronas.

En términos generales, la neurona recibe estímulos a través de las dendritas y si el estímulo total recibido es suficientemente grande como para rebasar su umbral, ésta producirá un nuevo potencial de acción que será transmitido por el axón, de lo contrario permanecerá inactiva.

Esta descripción tan general de la estructura de las neuronas biológicas y su funcionamiento será suficiente para comenzar a abordar su versión artificial.

2.3. Neurona artificial

Como ya mencionamos las neuronas artificiales tratan de extraer las características más generales de las neuronas biológicas, en general, su funcionamiento consiste en procesar un conjunto de valores de entrada (*inputs*) con el fin de producir un resultado (*outputs*).

El proceso de sinapsis ocurrido en las neuronas biológicas se incluye a este modelo agregando un conjunto de factores multiplicativos denominados pesos sinápticos, usualmente denotado como $W = (w_1, w_2, \dots, w_n)$, donde la entrada w_j representa el peso sináptico j , ($1 \leq j \leq n$), y n es el número de pesos sinápticos. Entre más eficiente sea la sinapsis, la señal transmitida al cuerpo de la neurona será más potente, es por eso que valores mayores en términos absolutos se asocian a los pesos correspondientes a sinapsis más eficientes, mientras que en señales débiles el peso sináptico será cercano a cero. Estos valores son muy importantes debido a que en ellos recae la capacidad de aprendizaje de las neuronas, esto último será abordado con mayor detalle posteriormente.

Al igual que en las neuronas biológicas, las neuronas artificiales cuentan con un umbral b . Al recibir los valores de entrada, la neurona procederá a multiplicar cada uno por su correspondiente peso sináptico para posteriormente sumarlos, el resultado de esta operación dará como resultado el estímulo total producido en la neurona el cual será comparado con el umbral para decidir si ésta pasara de un estado inactivo a uno activo, dicho de otra manera, la neurona se activará si y sólo si el estímulo total rebasa el umbral. A la función encargada de arrojar el resultado de la red se le conoce como función de activación.

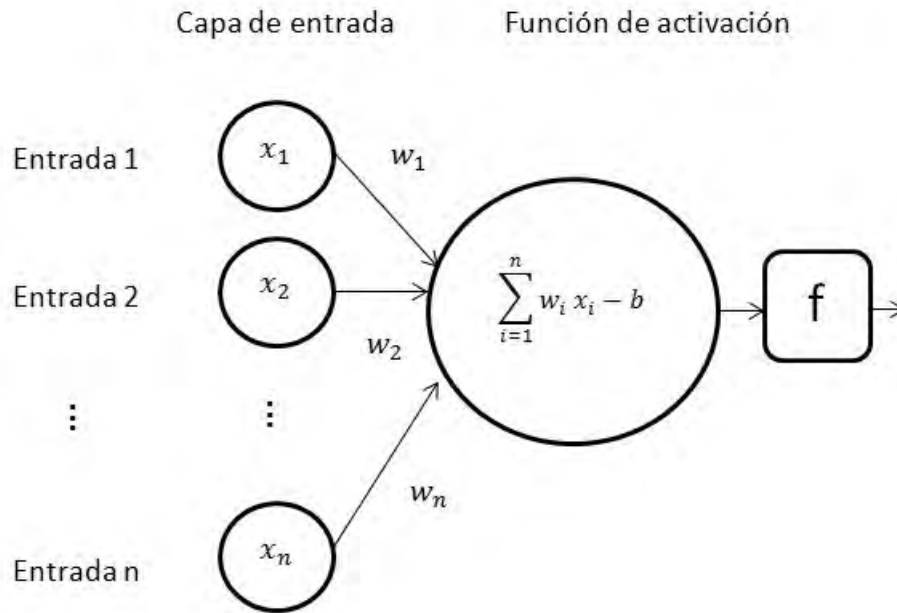


Figura 2.2: Modelo básico de una neurona artificial.

Para dar una mayor claridad al proceso antes mencionado se describirá paso a paso matemáticamente.

Definimos: $X = (x_1, x_2, \dots, x_n)$ el vector de entradas, $W = (w_1, w_2, \dots, w_n)$ el vector de pesos sinápticos, b el umbral asociado a la neurona y f la función de activación.

En la primera etapa de este proceso la neurona recibe al vector de entradas X y procede a realizar el producto escalar de éste con el vector W , a este resultado se le conoce como entrada total.

$$\text{entrada total} = \sum_{i=1}^n x_i w_i \quad (2.1)$$

Después se realiza la comparación de la entrada total con el umbral.

$$y = \text{entrada total} - b$$

$$y = \sum_{i=1}^n x_i w_i - b \quad (2.2)$$

Por último la función de activación es evaluada en y , lo cual nos dirá si la neurona

logro o no activarse.

$$f(y) = \begin{cases} 1 & y > 0 \\ 0 & e.o.c. \end{cases} \quad (2.3)$$

En este caso la función de activación utilizada es conocida como la función de Heaviside.

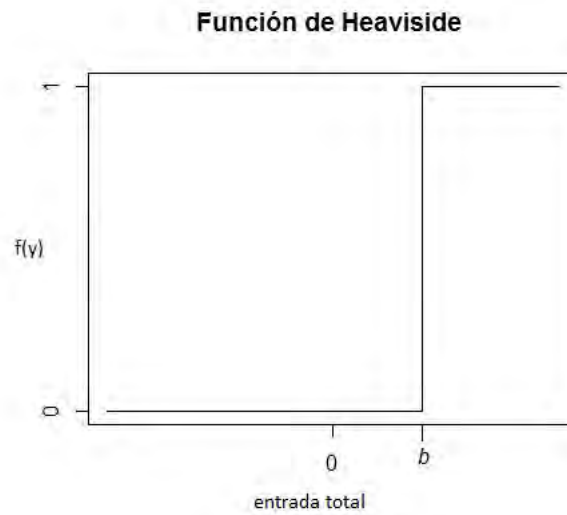


Figura 2.3: Función de Heaviside con umbral b

Otra forma de escribir lo anterior es agregando una entrada adicional x_0 al vector de entradas y w_0 al vector de pesos sinápticos donde x_0 siempre tendrá el valor de $-b$ y w_0 será igual a 1.

$$y = \sum_{i=0}^n x_i w_i \quad (2.4)$$

$$f(y) = \begin{cases} 1 & y > 0 \\ 0 & e.o.c. \end{cases} \quad (2.5)$$

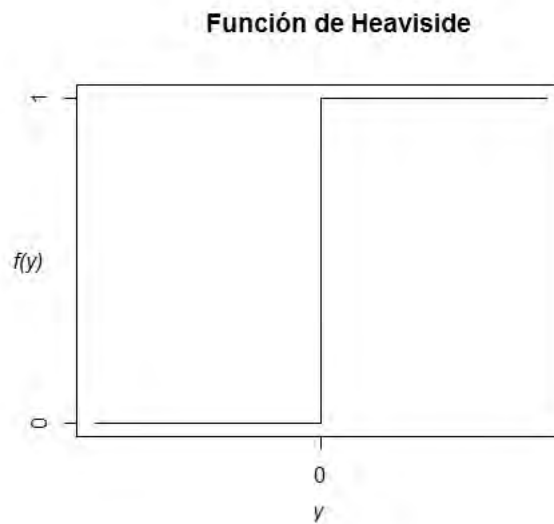


Figura 2.4: Función de Heaviside con umbral 0

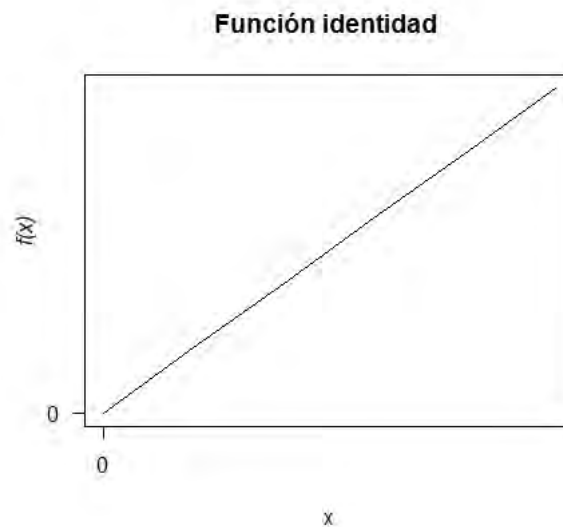
Podemos utilizar la función de Heaviside como función de activación, sin embargo, no es la única, de hecho, esta función no nos será útil con otro tipo de redes de neuronas por razones que se explicarán más adelante, a continuación se mencionarán las funciones de activación más comunes.

Funciones de activación

Identidad

La función identidad también puede ser utilizada como función de activación para las neuronas artificiales, sin embargo, esta función de activación es lineal, y por consiguiente tiene los mismos problemas que la función de Heaviside.

$$f(x) = x \quad (-\infty < x < \infty) \quad (2.6)$$



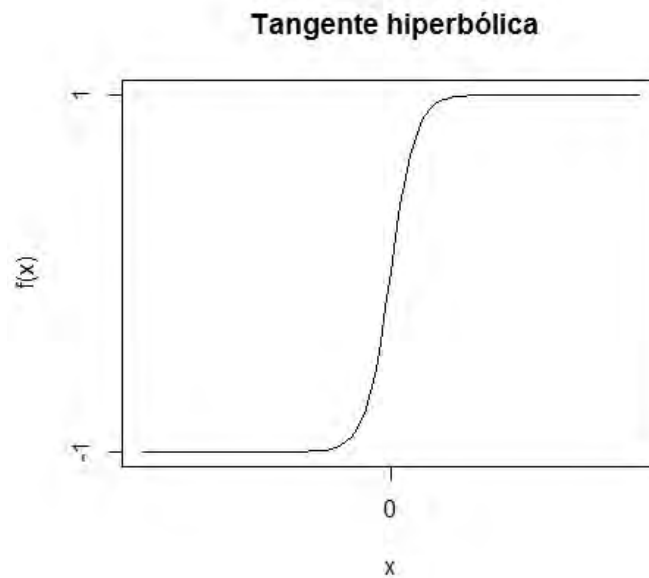
Sigmoide

Son funciones de variable real en forma de S en las que destaca la relación que existe entre el valor de la función y su derivada, lo cual es de utilidad ya que reduce la cantidad de trabajo computacional necesario para el entrenamiento de las redes.

Entre las funciones sigmoide más utilizadas se encuentran: la función logística y tangente hiperbólica.

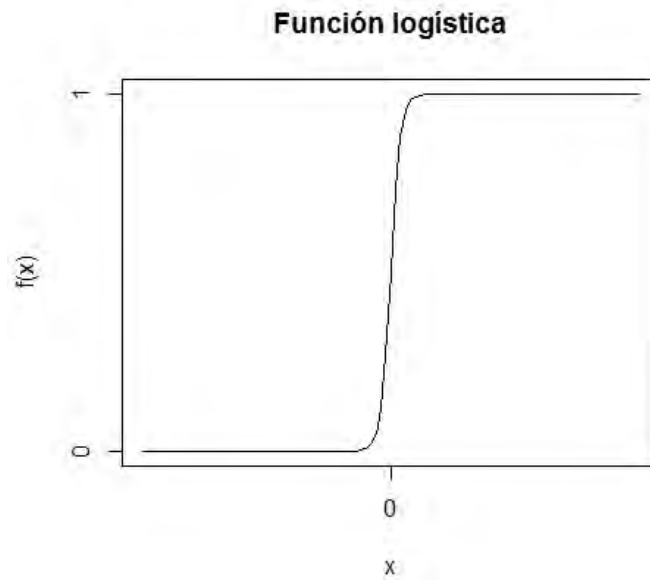
Tangente hiperbólica

$$f(x) = \frac{\sinh(x)}{\cosh(x)} \quad (-\infty < x < \infty) \quad (2.7)$$



Logística

$$f(x) = \frac{1}{1 + e^{kx}} \quad (-\infty < x < \infty); k \in R \quad (2.8)$$



La función logística presenta dos principales inconvenientes: el primero es que en las colas el gradiente es muy pequeño, y por lo tanto en esos caso las red aprende muy lento, este problema se podría presentar al momento de iniciar los pesos, si éstos son demasiado grandes en términos absolutos la neurona podría saturarse y actuar de la manera antes mencionada, el segundo es que esta función no está centrada en 0, lo cual podría ocasionar problemas cuando se aplique el método del gradiente, ya que si los datos que llegan a las neuronas son siempre positivos esto podría influir al momento de realizar el ajuste de los pesos con la posibilidad de que se asignen a éstos siempre valores positivos o negativos ocasionando un zigzag que afecte su correcto ajuste, este problema podría evitarse utilizando la función tangente hiperbólica la cual sí está centrada en 0.

Al modelo neuronal, conectado de manera simple como el que acabamos que describir se le dio el nombre de perceptrón en 1962 por Frank Rosenblatt, el cual en su libro *Principles of Neurodynamics* describe al perceptrón como una red

simplificada en la cual ciertas propiedades del sistema nervioso son exageradas y otras son ignoradas. [39]

A pesar de su simplicidad este modelo es capaz de comportarse como un discriminador lineal, esto quiere decir que puede aprender a clasificar correctamente cada uno de los elementos de dos distintas clases, esto si y sólo si existe por lo menos un hiperplano que sirva de frontera entre ellas (figura 2.5).

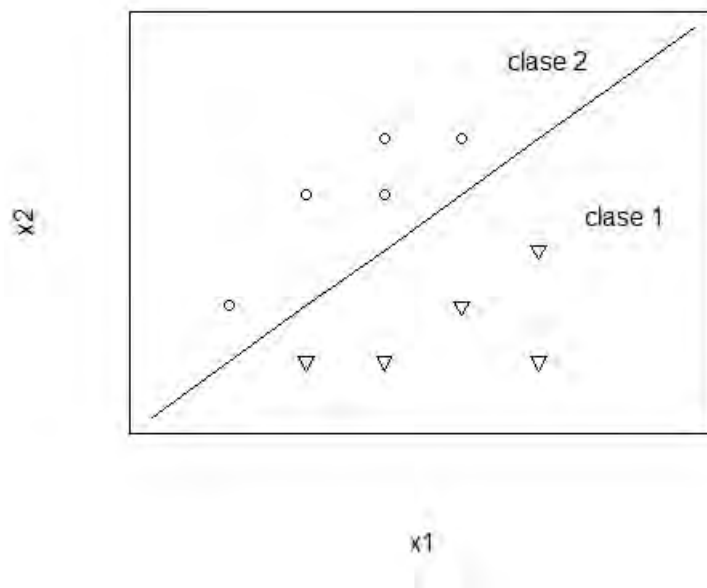


Figura 2.5: Ejemplo de dos clases linealmente separables.

Para que el perceptrón logre aprender es necesario someterlo a un proceso de entrenamiento, éste se basa en la idea de que la neurona aprenda de sus errores. En la primera etapa del entrenamiento se fijan los pesos sinápticos y el umbral de manera aleatoria, posteriormente, la neurona recibe varios elementos y su correspondiente clase, a este conjunto de valores se le denomina conjunto de entrenamiento y servirá como ejemplo que mostrará a la neurona como es que tiene que realizar la clasificación, la neurona procesa cada uno los valores de entrada recibidos para obtener un resultado, si el resultado obtenido por la neurona no concuerda con el resultado deseado se modifican los pesos sinápticos para evitar que el error se vuelva a repetir, a este tipo de entrenamiento se le conoce como entrenamiento supervisado.

Algoritmo de entrenamiento

1) Iniciar pesos sinápticos y umbral.

Sea $w_i(p)$ el peso sináptico asociado a la entrada i al tiempo p con $(0 \leq i \leq n)$, b el umbral, $w_0 = -b$ y $x_0 = 1$.

Los valores $w_i(0)$ se establecen de manera aleatoria tomando preferentemente valores pequeños.

La forma más común de fijar estos valores es utilizando un variable aleatoria con función de distribución uniforme continua en el intervalo $[-1,1]$, sin embargo, dependiendo de las características de la red también se pueden utilizar otros métodos como por ejemplo el de Nguyen-Widrow.

2) Se presenta el vector de entradas (x_1, x_2, \dots, x_n) y la salida deseada $d(p)$.

3) Calcular la salida actual.

$$y(p) = f_h \left[\sum_{i=0}^n w_i(p)x_i(p) \right] \quad (2.9)$$

f_h la función de Heaviside.

4) Actualizar los pesos.

$$\text{Si } y(p) = d(p); \quad w_i(p+1) = w_i(p)$$

$$\text{Si } y(p) = 0 \text{ y } d(p) = 1; \quad w_i(p+1) = w_i(p) + x_i(p)$$

$$\text{Si } y(p) = 1 \text{ y } d(p) = 0; \quad w_i(p+1) = w_i(p) - x_i(p)$$

[39]

Limitaciones

El hecho de que el perceptrón solo logre realizar la clasificación cuando las clases son linealmente separables restringe mucho su utilidad, ya que en muchos de los casos, la frontera entre las clases resulta ser más compleja que una línea recta o un hiperplano.

El ejemplo clásico de este problema es la función lógica XOR (tabla 2.1).

x	y	f(x,y)
0	0	0
1	1	0
0	1	1
1	0	1

Tabla 2.1: Función XOR

Como podemos ver en la figura 2.6 estamos tratando con un problema linealmente inseparable y por lo tanto es imposible que el perceptrón sea entrenado para resolverlo.

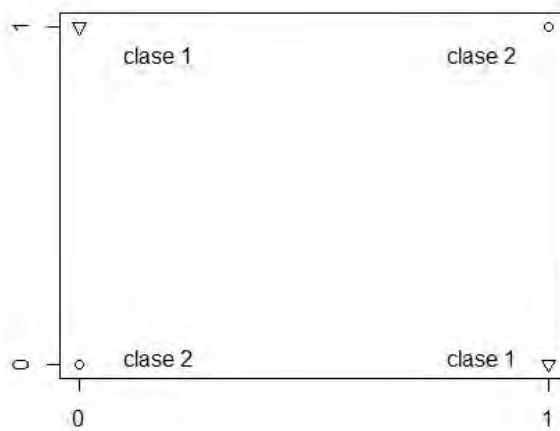


Figura 2.6: Función lógica XOR

Esta función tan simple pone en evidencia la gran cantidad de problemas que el perceptrón no será capaz de abordar de manera exitosa y es por esto que se han desarrollado otras redes de neuronas artificiales que además de resolver los las limitaciones planteadas para el perceptrón serán capaces de abordar otro tipo de cuestiones además de la clasificación.

2.4. Tipos de redes de neuronas artificiales

Las redes de neuronas artificiales pueden ser clasificadas principalmente basándonos en su arquitectura, el método utilizado para ser entrenadas y su flujo de información.

2.4.1. Arquitectura

La disposición de las neuronas en capas y cómo es que éstas se conectan entre sí se conoce como arquitectura de la red, cada capa está compuesta de cierto número de neuronas las cuales procesan en paralelo la información recibida. Tomando

en cuenta lo anterior las redes de neuronas artificiales suelen clasificarse por su arquitectura como monocapa y multicapa.

Red monocapa

Este tipo de red neuronal como su nombre lo indica solo cuenta con una capa para procesar información. Su estructura se compone de una capa de entrada en donde se recibe la información sin realizar ningún tipo de procesamiento y una capa de salida en donde se procesan los datos y se muestra el resultado obtenido por la red, el ejemplo básico para ilustrar este tipo de redes es el perceptrón, sin embargo, existen otros tipos de redes más complejas como la red de Hopfield que entran dentro de esta categoría (figura 2.7).

La red de Hopfield es una red altamente interconectada en donde cada neurona se conecta con el resto, en este tipo de red los pesos sinápticos son simétricos es decir w_{ij} que representa el peso sináptico que va de la neurona i a la j es igual a w_{ji} el cual representa la conexión de j en dirección a la neurona i . [40]. Estas redes cuentan con un valor escalar asociado a cada estado, conocido como energía de la red, y cada nodo tiene un umbral y una función de activación, éstos calculan la salida binaria haciendo la suma ponderada de las entradas con los pesos sinápticos, restándoles el umbral y pasando la entrada total a través la función de activación, sin embargo, lo que distingue a las redes de Hopfield es la manera en que éstas encuentran la solución, primero, esta red es alimentada con el vector de entradas el cual es binario, posteriormente, los nodos van cambiando a través de una sucesión de estados hasta que la red converge a los mínimos locales de la función energía (esto no siempre ocurre) finalizando el proceso. [39] Estas redes son utilizadas como memorias asociativas, es decir, permiten recuperar información a partir del conocimiento parcial de su contenido.

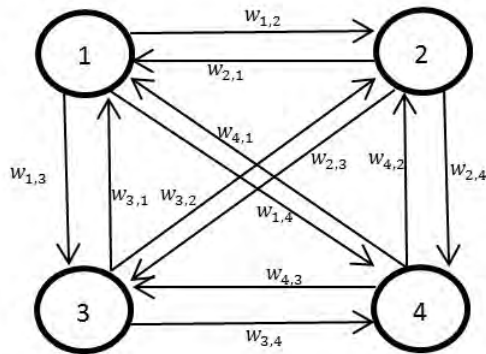


Figura 2.7: Red de hopfield

Red multicapa

Las redes neuronales artificiales multicapa se caracterizan por tener una o más capas entre la capa de entrada y la capa de salida conocidas como capas ocultas, este tipo de red es capaz de resolver problemas más complejos que las redes mono capa, pero su entrenamiento suele ser más difícil [40], un ejemplo de este tipo de redes es el perceptrón multicapa.

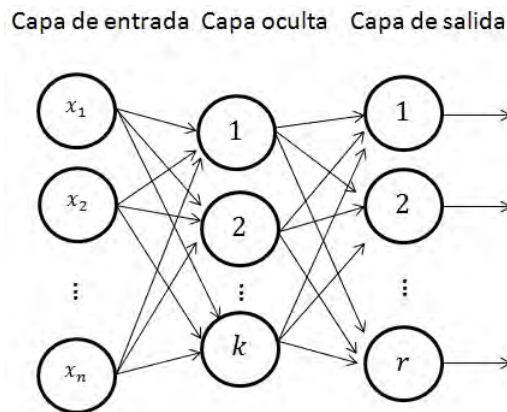


Figura 2.8: Perceptrón multicapa

2.4.2. Tipo de entrenamiento

Para que una red neuronal logre aprender, esta tiene que establecer el grado de conexión entre sus neuronas, lo que es igual a fijar sus pesos sinápticos, para poder

hacerlo la red tiene que someterse a un proceso denominado entrenamiento el cual puede ser clasificado en dos tipos: entrenamiento supervisado y entrenamiento no supervisado.

Entrenamiento supervisado

En este tipo de entrenamiento se le presenta a la red un conjunto de valores de entrada (*inputs*) que a su vez están asociados a un resultado deseado (*target*), la red recibe la información, la procesa y arroja un resultado (*output*), después se procede a comparar el resultado obtenido con el deseado mediante una función, la cual servirá como medida del error cometido por la red y representa la diferencia entre el resultado actual de la red y la salida que queremos que se produzca [39], usualmente se utiliza el error cuadrático medio, aunque existe una infinidad más de funciones que pueden ser utilizadas. Los pesos de la red son modificados mediante un algoritmo de aprendizaje para lograr minimizar el error de la red.

Una red que está siendo entrenada para asociar un conjunto de valores de entrada con un correspondiente conjunto de resultados se le conoce como una memoria asociativa. Si los valores de salida deseados son los mismo que los valores de entrada se le denomina memoria auto-asociativa y si los valores de salida deseados son diferentes a los valores de entrada la red será una memoria hetero-asociativa. [40]

Entrenamiento no supervisado

A diferencia del entrenamiento supervisado en este caso la red agrupará valores de entrada similares sin la necesidad de saber *a priori* el resultado que se desea obtener, primero se le presentará a la red un conjunto de valores de entrada (conjunto de entrenamiento) y ésta modificará sus pesos sinápticos para que aquellos elementos que presenten similitudes sean asignados a un mismo grupo o cluster. Un ejemplo de este tipo de redes son los mapas auto-organizados de Kohonen (figura 2.9).

Los mapas auto-organizados son un tipo de red neuronal entrenada mediante entrenamiento no supervisado, la cual se compone de nodos (neuronas), y para cada neurona hay un vector de pesos asociado. Estas redes describen un mapeo de un espacio de una dimensión mayor en uno de menor dimensión.

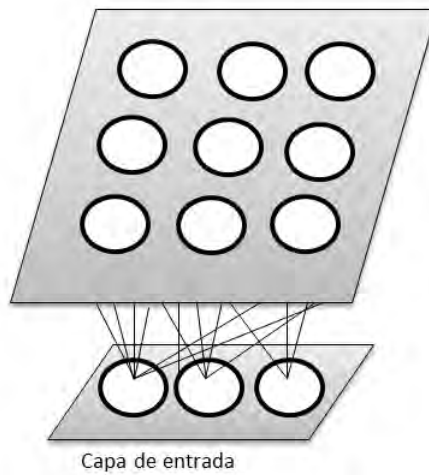


Figura 2.9: Mapas auto-organizados: en este tipo de red cada neurona representa una clase a la cual serán asociados todos los elementos que la red considere similares o del mismo tipo. [39]

2.4.3. Tipo de conexión

Esta clasificación toma como criterio la forma en que fluye la información dentro de las redes. Si ésta fluye desde las unidades de entrada hasta las unidades de salida siempre en dirección hacia adelante diremos que estamos tratando una red del tipo *feedforward* o de alimentación hacia adelante, por otro lado si la red cuenta con circuitos cerrados en donde la información puede ser transmitida de una neurona hacia otra en una capa anterior o hacia ella misma se dirá que estamos tratando con una red del tipo recurrente, esto quiere decir que existe retroalimentación en las neuronas.

2.5. Perceptrón Multicapa

El perceptrón multicapa es una red de neuronas artificiales compuesta por unidades de procesamiento de información simples agrupadas en capas que pueden clasificarse dentro de los siguientes tres tipos: capa de entrada, capas ocultas y capa de salida (figura 2.10).

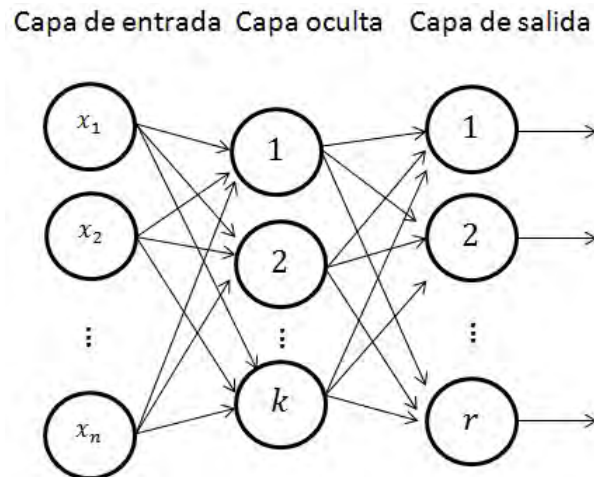


Figura 2.10: Perceptrón multicapa

La capa de entrada es la encargada de recibir la información que se va a procesar y alimentar a la siguiente capa de neuronas. Las capas ocultas están compuestas por perceptrones solo que en este caso no usarán la función de Heaviside como función de activación, esto quedará justificado más adelante. La capa de salida se compone de un conjunto de neuronas las cuales mostrarán el resultado obtenido por la red.

En este caso la red puede estar total o localmente conectada. Una red está totalmente conectada si cada neurona se conecta con todas las neuronas de la capa siguiente (figura 2.10), sin embargo, cuando las neuronas solo se conectan con un subconjunto de neuronas en la capa siguiente se dirá que es una red localmente conectada (figura 2.11).

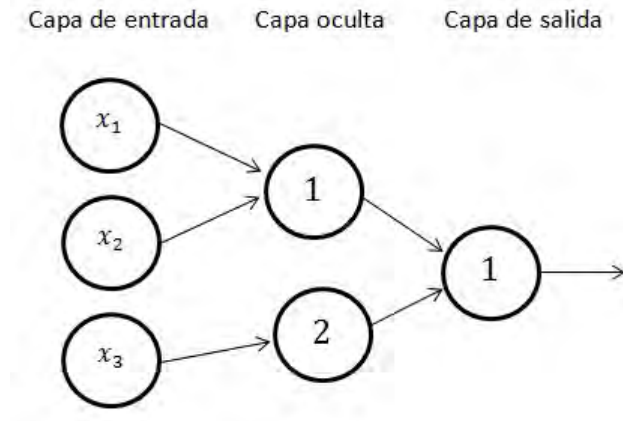


Figura 2.11: Red de neuronas artificiales localmente conectada.

El perceptrón multicapa pertenece al conjunto de redes que ocupan el aprendizaje supervisado, en este caso la capa de entrada recibe las entradas y las manda a la capa oculta siguiente, ahí cada neurona funciona de la misma manera que el perceptrón, solo que en este caso la función de activación es sustituida generalmente por alguna de las funciones antes mencionadas, esto se debe a que la función de Heaviside solo permite conocer la información de entrada a la capa oculta adyacente a la capa de entrada, las demás capas sólo conocen la información con las que están siendo alimentadas, esto es, las salidas de la capa anterior las cuales al ser valores binarios pierden toda la información original, esto imposibilita el correcto ajuste de los pesos sinápticos debido a que no es posible saber en qué proporción hay que ajustarlos para lograr un mejor resultado, a este problema se le conoce como “Problema de asignación de crédito”. [39]

Para lograr el entrenamiento de esta red se ha desarrollado una regla de aprendizaje conocida como propagación hacia atrás (*backpropagation*) o regla delta generalizada (*generalized delta rule*), en ésta se calcula el error de las neuronas de la capa de salida, lo cual es relativamente fácil debido a que en este caso conocemos el valor obtenido por la red y el valor deseado, una vez hecho lo anterior se procederá a transmitir el error hacia atrás y de esta manera ajustar los pesos de cada neurona, el método para ajustar los pesos será abordado a continuación,

pero intuitivamente podemos notar que los pesos que se modificarán en mayor medida son aquellos que estén conectados a las neuronas que producen un resultado muy diferente al deseado mientras que los que produzcan un error más pequeño se modificarán en menor proporción. [39]

Algoritmo de entrenamiento

Algoritmo de propagación hacia atrás utilizando la función sigmoide:

$$f(net) = \frac{1}{e^{-knet}}$$

1) Iniciar pesos sinápticos y umbral.

Se inicializan los pesos sinápticos y los umbrales de manera aleatoria.

Al igual que en el caso del perceptrón, la forma más común de fijar estos valores es utilizando un variable aleatoria con función de distribución uniforme continua en el intervalo $[-1,1]$, sin embargo, dependiendo de las características de la red también se pueden utilizar otros métodos como por ejemplo el de Nguyen-Widrow. 2) Se presenta la entrada $X_p = x_0, x_2, \dots, x_{n-1}$ y la salida deseada $T_p = t_0, t_1, \dots, t_{m-1}$, donde n es el número de neuronas en la capa de entrada y m es el número de neuronas en la capa de salida.

Sea $w_0 = -b$ y $x_0 = 1$. Para asociación de patrones, X_p y T_p representan los patrones que deben de ser asociados, para clasificación, T_p es un conjunto de ceros excepto para un elemento, el cual tendrá el valor de 1 y corresponderá a la clase a la que X_p debe ser asociado.

3) Calcular la salida actual.

Para cada capa se calcula de la siguiente manera:

$$y_{pj} = f \left[\sum_{i=0}^{n-1} w_i x_i \right] \quad (2.10)$$

Posteriormente se pasa a la siguiente capa hasta llegar a la capa de salida. Los valores de la neurona j asociados al patrón p se denotarán como o_{pj} .

4) Actualizar los pesos.

Se inicia con los pesos asociados a la capa de salida y para después ajustar los de las capas anteriores, es decir, se ajustan de la capa de salida hacia atrás.

$$w_{i,j}(s+1) = w_{i,j}(s) + \eta \delta_{pj} o_{pj}$$

Donde $w_{i,j}(s)$ representa el peso sináptico asociado a la neurona i y a la neurona j al tiempo s , η es el coeficiente de aprendizaje y δ_{pj} es el error cometido para el patrón de entrada p .

Para las neuronas de salida el error se define:

$$\delta_{pj} = k o_{pj} (1 - o_{pj}) (t_{pj} - o_{pj})$$

Para las neuronas de las capas ocultas:

$$\delta_{pj} = k o_{pj} (1 - o_{pj}) \sum_k \delta_{pk} w_{j,k}$$

Donde la suma es sobre las k neuronas en la capa siguiente a la de la neurona j .
[39]

Aplicaciones

El perceptrón multicapa ha sido ampliamente utilizado en distintos campos y ha abordado una muy variada cantidad de problemas. En general, cualquier problema que pueda ser enmarcado dentro del área de clasificación o predicción es candidato a ser abordado con esta herramienta [46]. Debido a la gran cantidad de estudios donde éste se ha utilizado sólo se mencionarán algunos ejemplos para ilustrar su versatilidad.

Una aplicación muy famosa es conocida como *NETtalk* [47] desarrollada en 1987 por Sejnowski y Rosenberg, en este caso se entrenó un perceptrón multicapa para leer un texto en inglés lo cual se lograba con cierto porcentaje de error, en el portal de internet *youtube* se puede encontrar un video [48] el cual muestra el comportamiento de dicha red.

En el área de las finanzas el perceptrón multicapa se ha utilizado para predecir el desempeño de rentabilidad de las acciones en el mercado de capitales. [46]

También se ha utilizado este tipo de red para predecir la estancia de los pacientes psiquiátricos en el hospital con base en información recabada durante su admisión al mismo. [49]

Capítulo 3

Análisis de cúmulos

La idea de agrupar objetos con características similares ha estado presente desde las primeras culturas, por ejemplo, éstas tenían que ser capaces de diferenciar entre grupos de plantas comestibles y plantas venenosas, animales inofensivos y aquellos que representaban algún peligro, etc. [50]

Además de ser una actividad básica para el ser humano, la clasificación es también fundamental para la mayoría de las ramas de la ciencia. La biología y la zoología aportaron mucho a este campo con el desarrollo de herramientas numéricas cuyo principal objetivo era lograr realizar clasificaciones objetivas y estables, es decir, que el análisis del mismo conjunto de organismos mediante los mismos métodos produjese el mismo resultado y que la clasificación no cambiara con la adición de nuevos elementos o nuevas características. [50]

Al conjunto de metodologías para la clasificación automática de muestras en un número de grupos usando una medida de asociación se le conoce como análisis de cúmulos. Esta técnica es apropiada para la exploración preliminar de los datos con los que se trabaja y permite hacer una evaluación con respecto a su estructura. [51]

3.1. Conceptos generales

El análisis de cúmulos es el estudio formal de métodos y algoritmos para el agrupamiento de objetos con base en sus similitudes.[51]

En particular el problema de clasificación es complejo debido a que los datos pueden ser agrupados de distintas maneras, en donde no sólo varía el número de grupos formados sino también la combinación de rasgos que caracteriza a los elementos de cada grupo.

Los datos utilizados en el análisis de cúmulos son representados en una matriz de $n \times p$ en donde los renglones de la matriz representan a los elementos y las columnas las variables que los caracterizan, por ejemplo, si estuviésemos analizando un conjunto de personas, cada renglón representaría a un sujeto en específico y cada columna una característica: color de ojos, personalidad, estatura, etc.

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,p} \end{bmatrix}$$

Dado que el objetivo es formar grupos cuyos elementos compartan características, es decir, tengan algún tipo de similitud, es necesario establecer una medida que nos indique de manera clara cuando es que dos elementos son similares en cierto sentido, para esto se utiliza una función de distancia d o de similitud s la cual será abordada con mayor detalle más adelante.

Básicamente el análisis de cúmulos consiste en obtener una partición P del conjunto total de elementos X ; $P = \{G_1, G_2, \dots, G_N\}$ en donde los G_k ; ($1 \leq k \leq N$) son subconjuntos de X tales que :

$$G_1 \cup G_2 \cup \dots \cup G_N = X$$

$$G_i \cap G_j = \emptyset; \quad \forall j \neq i$$

Los elementos G_k de P son llamados cúmulos o *clusters*.

Existe un gran número de algoritmos que se pueden utilizar para formar los cúmulos, estos pueden ser clasificados dentro de dos categorías: jerárquicos y de partición.

En los algoritmos jerárquicos comúnmente se inicia con n grupos, uno para cada observación, y terminamos con un solo grupo que contiene todas las observaciones, en cada paso, una observación o un grupo de observaciones se absorbe en otro grupo. También se puede comenzar con un sólo cluster que contiene todas las n observaciones y terminar con n grupos de un sólo elemento, es decir, realizar el proceso de manera inversa. [52]

Por otro lado los algoritmos de partición consisten en dividir las observaciones en k grupos. Esto puede hacerse comenzando con un particionamiento inicial y luego reasignando las observaciones de acuerdo con algún criterio. [52]

Para poder agrupar a los elementos de la muestra es necesario aplicar el algoritmo de agrupación a los renglones de X , los cuales como ya mencionamos representan a un elemento en específico, sin embargo, también se puede aplicar este método para las columnas (variables), como una alternativa para reducir la dimensionalidad. [50]

Es importante mencionar que no existe una técnica de agrupamiento que sea universalmente aplicable para descubrir la variedad de estructuras presentes en conjuntos de datos multidimensionales [51], es por esto que la correcta comprensión del problema y el buen manejo de varias técnicas de agrupamiento serán primordiales para obtener buenos resultados.

3.2. Medidas de proximidad

Al hacer un análisis de cúmulos es de vital importancia conocer que tan “parecidos” son los elementos con los que estamos trabajando, esto ya que lo que se busca es crear grupos relativamente homogéneos.

Algunas técnicas de análisis de cúmulos comienzan formando una matriz de $n \times n$ en donde cada entrada representa una medida con respecto a la cercanía, distancia,

similitud o en términos más generales la proximidad entre los elementos.

$$D = \begin{bmatrix} d_{1,1} & d_{1,2} & \cdots & d_{1,n} \\ d_{2,1} & d_{2,2} & \cdots & d_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ d_{n,1} & d_{n,2} & \cdots & d_{n,n} \end{bmatrix} \quad (3.1)$$

Como medida de proximidad podemos utilizar una función de distancia o métrica ya que ésta aumenta a medida que los elementos asociados se separan, sin embargo, también es posible usar para esta cuestión una función de similitud, este tipo de funciones presentan valores grandes cuando se aplican a un par de elementos que son parecidos y valores pequeños cuando no lo son. En general, dos elementos están más próximos cuando su distancia asociada es pequeña o su similitud es grande.

A continuación se abordarán algunas medidas de proximidad que son comúnmente utilizadas.

3.2.1. Distancia

Dada una matriz de $n \times p$, con n vectores renglón x_1, x_2, \dots, x_n , las distancias entre los vectores x_s y x_t se definen:

- Distancia euclidiana

$$d_{s,t}^2 = (x_s - x_t)(x_s - x_t)'$$

- Distancia de Mahalanobis

$$d_{s,t}^2 = (x_s - x_t)C^{-1}(x_s - x_t)'$$

C es la matriz de covarianza.

- Distancia de Minkowski

$$d_{s,t} = \sqrt[r]{\sum_{j=1}^p |x_{s,j} - x_{t,j}|^r}$$

- Distancia de Hamming

$$d_{s,t} = \frac{\#(x_{s,j} \neq x_{t,j})}{p}$$

3.2.2. Similitud

Dada una matriz de $n \times p$, con n vectores renglón x_1, x_2, \dots, x_n .

Sea:

a= el número de elementos en x_s y x_t tal que $x_{s,j} = x_{t,j} = 1; (1 \leq j \leq p)$

b= el número de elementos en x_s y x_t tal que $x_{s,j} = 1$ y $x_{t,j} = 0; (1 \leq j \leq p)$

c= el número de elementos en x_s y x_t tal que $x_{s,j} = 0$ y $x_{t,j} = 1; (1 \leq j \leq p)$

d= el número de elementos en x_s y x_t tal que $x_{s,j} = x_{t,j} = 0; (1 \leq j \leq p)$

Definimos las siguientes medidas de similitud.

- Coeficiente de coincidencia simple (simple matching coefficient)

$$S(x_t, x_s) = \frac{a + d}{a + b + c + d}$$

- Coeficiente de Jaccard

$$S(x_t, x_s) = \frac{a}{a + b + c}$$

- Coeficiente de Rao

$$S(x_t, x_s) = \frac{a}{a + b + c + d}$$

Cabe destacar que las medidas antes mencionadas sólo se aplican a vectores de entradas binarias.

Fuente [51]

3.3. Métodos jerárquicos

Los métodos jerárquicos se caracterizan por no definir el número de cúmulos desde un principio. Estos métodos pueden ser clasificados en dos categorías: jerárquicos aglomerativos y jerárquicos divisivos. [51]

3.3.1. Métodos jerárquicos aglomerativos

Los métodos aglomerativos son probablemente los más utilizados dentro de los métodos jerárquicos [50]. En este caso al inicio del proceso se considera a cada elemento como un cúmulo y posteriormente se fusionan los elementos más próximos de acuerdo con algún criterio para formar un nuevo grupo, al final del proceso se llega a un sólo cúmulo, el cual contiene a todos los elementos. Este tipo de agrupamiento va de una partición más fina a una más gruesa.

Dado que en cada iteración del proceso se van agrupando los cúmulos más cercanos o próximos, es necesario abordar cómo se mide la cercanía entre éstos. Las diferentes maneras de medir la distancia entre dos grupos dan lugar a los distintos métodos aglomerativos que se abordarán a continuación.

Vecino más cercano (Single Linkage)

En este método la distancia entre dos cúmulos A, B es definida como el mínimo de las distancias entre los elementos de A y los elementos de B .

$$D(A, B) = \min \{d(x_s, x_t); \text{ para } x_s \text{ en } A \text{ y } x_t \text{ en } B\}$$

Donde d es alguna métrica o distancia entre vectores.

En cada paso de este proceso se calcula la distancia para cada par de cúmulos, posteriormente se fusiona el que tenga la distancia D mínima, este proceso se repite hasta llegar a tener un solo cúmulo.

Vecino más lejano (Complete Linkage)

Para este método la distancia entre dos cúmulos A,B se define como el máximo de las distancias entre los elementos de A y los elementos de B.

$$D(A, B) = \max \{d(x_s, x_t); \text{ para } x_s \text{ en } A \text{ y } x_t \text{ en } B\}$$

Al igual que en el método del vecino más cercano, en cada paso se calcula la distancia entre los cúmulos y se fusiona al par que presente la distancia D mínima.

Distancia promedio (Average Linkage)

En este método la distancia se define como la distancia promedio entre los elementos de A y los elementos de B.

$$D(A, B) = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} d(x_s, x_t)$$

Donde n_A y n_B son el número de elementos en A y B respectivamente.

Distancia entre centros (Centroid)

En este caso, la distancia entre dos cúmulos A,B es definida como la distancia euclidiana entre los vectores de medias (centroides) de los dos cúmulos.

$$D(A, B) = d(\bar{X}_A, \bar{X}_B)$$

Donde \bar{X}_A y \bar{X}_B se definen:

$$\bar{X}_A = \sum_{i=1}^{n_A} \frac{x_i}{n_A}$$

$$\bar{X}_B = \sum_{i=1}^{n_B} \frac{x_i}{n_B}$$

Los dos cúmulos con la distancia mínima asociada serán fusionados y se calculará el centroide del nuevo cúmulo de la siguiente manera.

$$\bar{X}_{AB} = \frac{n_A \bar{X}_A + n_B \bar{X}_B}{n_A + n_B}$$

Distancia mediana (Median)

Al utilizar el método de distancia entre centros, puede ocurrir que al fusionar dos cúmulos A, B el centroide asociado este mucho más cerca de \bar{X}_A o de \bar{X}_B , esto ocurre cuando uno de éstos tiene un mayor número de elementos en comparación con el otro, para evitar esto podemos utilizar la mediana (punto medio) de la línea que une A y B para calcular las distancias entre cúmulos la cual se define como:

$$m(A, B) = \frac{\bar{X}_A + \bar{X}_B}{2}$$

Los dos cúmulos con la distancia más pequeña serán fusionados en cada paso.

Método de Ward

Este método utiliza las distancias al cuadrado dentro de los cúmulos y la distancia al cuadrado entre cúmulos. En este caso se fusionan los dos cúmulos que minimicen el incremento en el valor total de la suma de los cuadrados de las diferencias de cada individuo al centroide del cúmulo definido como:

$$I_{AB} = SSE_{AB} - (SSE_A + SSE_B)$$

Donde:

$$SSE_A = \sum_{i=1}^{n_A} (x_i - \bar{X}_A)'(x_i - \bar{X}_A)$$

$$SSE_B = \sum_{i=1}^{n_B} (x_i - \bar{X}_B)'(x_i - \bar{X}_B)$$

$$SSE_{AB} = \sum_{i=1}^{n_{AB}} (x_i - \bar{X}_{AB})'(x_i - \bar{X}_{AB})$$

Con $n_{AB} = n_A + n_B$, la suma del número de elementos en A y en B.

Otra forma de definir el incremento es la siguiente:

$$I_{AB} = \frac{n_A n_B}{n_A + n_B} (\bar{X}_A - \bar{X}_B)' (\bar{X}_A - \bar{X}_B)$$

Minimizar este incremento es equivalente a minimizar la distancia entre cúmulos, esto se puede ver de una manera más clara cuando cada cúmulo tiene un solo elemento, en este caso A está formado solo por el elemento x_i y B por el elemento x_j , lo cual implica que SSE_A y SSE_B son 0, por lo tanto I_{AB} se reduce a:

$$I_{ij} = SSE_{AB} = \frac{1}{2} (x_i - x_j)' (x_i - x_j) = \frac{1}{2} d^2(x_i, x_j)$$

Dependiendo de la estructura de los datos, los métodos antes mencionados obtendrán cúmulos diferentes, es por esto que es necesario tener en cuenta cómo es que estos realizan las agrupaciones y sus principales limitantes. La siguiente tabla contiene algunas observaciones referentes a los distintos métodos jerárquicos aglomerativos.

Método	
Vecino más cercano	Tiende a producir cúmulos desequilibrados y desordenados (<i>chaining</i>), especialmente en conjuntos de datos grandes.
Vecino más lejano	Tiende a encontrar cúmulos compactos con diámetros iguales.
Distancia promedio	Tiende a unir cúmulos con varianzas pequeñas y es relativamente robusto.
Distancia entre centros	En este caso, al realizar la unión de dos cúmulos, cuando uno presenta un número mayor de elementos, éste domina al cúmulo fusionado y el nuevo centroide se carga hacia el cúmulo más grande.
Distancia mediana	Puede utilizarse para evitar el problema que surge con el método de distancia entre centros.

Método de Ward	Tiende a encontrar cúmulos esféricos del mismo tamaño y es sensible a los valores atípicos.
----------------	---

[50]

3.3.2. Métodos jerárquicos divisivos

Los métodos divisivos funcionan de manera opuesta a los métodos aglomerativos, en este caso se parte de un cúmulo que contiene a todos los elementos para después dividirlo. En cada paso, cada subconjunto se divide en conjuntos más pequeños para que al final se obtengan tantos cúmulos como número de elementos.

Este enfoque es útil cuando es de interés realizar un análisis de cúmulos con un gran número de elementos, es decir, cuando queremos una división más gruesa de los datos, ya que en este caso los cúmulos más grandes se encuentran en los primeros pasos del proceso. Los métodos divisivos tienen el mismo problema que los aglomerativos, una vez que se hace la partición y cada elemento es asignado a un grupo, ya no hay posibilidad de reasignarlo, es por esto que pequeñas fallas al inicio del proceso ocasionan una división no óptima de los datos. [52]

3.3.3. Dendrograma

Los resultados de los métodos jerárquicos pueden ser visualizados en un dendrograma, el cual es una representación gráfica del proceso de agrupación, en este caso los nodos representan a los cúmulos y la altura de los tallos representa la distancia entre éstos.

[50]

Al cortar el dendrograma a una altura determinada se define una partición del conjunto total de modo que los elementos de los cúmulos formados distan entre sí a lo más en la altura del corte, por ejemplo, si se realiza un corte a altura k , todos los elementos de un cúmulo distarán entre sí a lo más en una distancia k ,

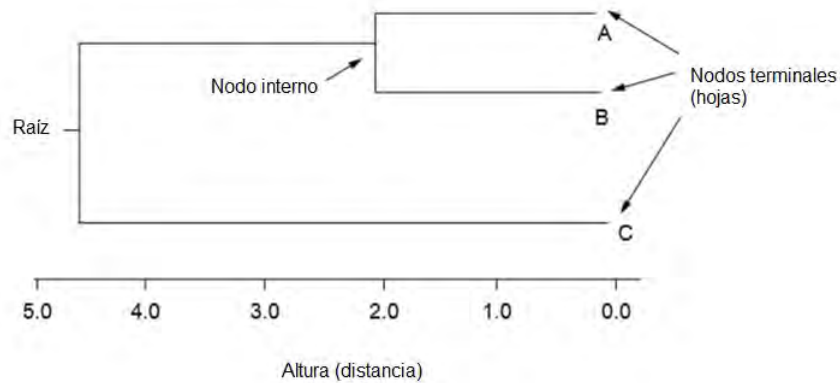


Figura 3.1: Dendrograma

en este caso el número de cúmulos depende completamente de la altura de corte.

Matriz cofenética y Coeficiente de correlación cofenética

Para evaluar que tan bien representa el árbol la estructura original de los datos usualmente es utilizado el coeficiente de correlación cofenética. Éste se define como la correlación entre las distancias cofenéticas obtenidas del árbol y las distancias originales (o disimilitudes).

Las distancias cofenéticas son representadas en una matriz triangular con ceros en la diagonal en donde sus entradas son las alturas en donde dos elementos se volvieron parte del mismo cúmulo. [50]

Este coeficiente será muy cercano a 1 cuando la solución obtenida sea buena, así que puede ser utilizado para comparar distintas soluciones obtenidas utilizando distintos algoritmos.

3.4. Métodos de partición

A diferencia de los métodos jerárquicos en los métodos de partición se define *a priori* el número de conglomerados que se desea. Cada uno de estos conjuntos contiene por lo menos a un elemento y no tienen elementos en común. [53]

Los cúmulos se forman mediante la optimización de una función, es decir, se trata de obtener la mejor partición del conjunto total de datos en el sentido de que los elementos de cada cúmulo estén relacionados entre sí y elementos en distintos cúmulos tengan poca relación.

El criterio más utilizado en estos métodos es la función de error al cuadrado, en este caso se formarán k cúmulos $\{C_1, C_2, \dots, C_k\}$, en donde cada conjunto C_i tiene n_i elementos tal que $\sum_{i=1}^k n_i = N$, con N el número total de elementos en la muestra, posteriormente se define un centroide para cada cúmulo.

$$M_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_{ik}$$

Donde x_{ik} es el i -ésimo elemento del cúmulo k .

El cuadrado del error de cada cúmulo se define de la siguiente manera:

$$e_k^2 = \sum_{i=1}^{n_k} (x_{ik} - M_k)^2$$

Y el cuadrado del error total es la suma de cuadrados de los errores de cada cúmulo.

$$E_k^2 = \sum_{i=1}^k e_k^2$$

El objetivo en este caso es encontrar una partición de la muestra con k cúmulos tal que que minimice el cuadrado del error total para una k fija.

3.4.1. Algoritmo de las k -medias

El algoritmo de las k medias es el algoritmo más simple y comúnmente utilizado que utiliza el error al cuadrado. Éste inicia con una partición inicial elegida de manera aleatoria y en cada iteración reasigna los elementos de la muestra a los cúmulos más cercanos hasta que se cumple el criterio de convergencia, este criterio usualmente se cumple cuando ya no hay reasignaciones que impliquen una disminución en el error al cuadrado total. [51]

Este algoritmo es popular debido a su fácil implementación y a su complejidad en espacio y tiempo relativamente pequeña. Su principal problema es que es sensible a la selección de la partición inicial y esto puede ocasionar que converja a un mínimo local, pero en general, brinda buenos resultados cuando los cúmulos son compactos y muestran una buena separación en el espacio de características.

[51]

Algoritmo

1. Se selecciona una partición inicial de k cúmulos cuyos elementos son elegidos de manera aleatoria.
2. Se genera una nueva partición asignando cada elemento al cúmulo con el centroide más cercano.
3. Una vez formados los nuevos cúmulos se actualizan los centroides.
4. Se repiten los pasos 2 y 3 hasta que no haya más reasignaciones.

[51]

Capítulo 4

Análisis de datos y ajuste de la red neuronal artificial

4.1. Base de datos

El estudio Pacardo[54] fue un estudio de investigación epidemiológica multinacional cuyo objetivo principal era registrar y describir el uso de drogas y los patrones de comportamiento asociados entre jóvenes de edad escolar (12 a 20 años) en América Latina.

Durante el año 1999-2000, se administró un cuestionario sobre el uso de estupefacientes y los comportamientos relacionados a una muestra transversal representativa a nivel nacional que incluyó un total de 12,797 estudiantes de los siete países siguientes: Costa Rica, República Dominicana, El Salvador, Honduras, Nicaragua, Panamá y Guatemala. Se evaluó el uso y la exposición de los sujetos a las siguientes sustancias: alcohol, tabaco, inhalantes, marihuana, cocaína, anfetaminas, metanfetaminas, tranquilizantes, éxtasis y heroína mediante las respuestas recabadas con el cuestionario.

Para este trabajo se utilizó la base de datos *PACARDO: Data on Drug Use and Behavior in School-Aged Children* proveniente del estudio antes mencionado.

Esta base de datos cuenta con 12,797 renglones (el número de sujetos encuesta-

dos) y 374 columnas, las primeras 12 corresponden a la información general del individuo (país, tipo de escuela, grado escolar, etc.) y el resto a las preguntas de la investigación y sus respectivas recodificaciones.

4.1.1. Codificación y recodificación de los datos

Las preguntas hechas en el cuestionario en su mayoría han sido diseñadas a modo tal que la respuesta sea del tipo booleano, a continuación se dará una descripción detallada de las preguntas y su codificación.

La información general referente al individuo (12 primeras columnas de la base) fue excluida de las variables debido a que las redes obtenían un peor rendimiento al utilizarlas para realizar la predicción, sin embargo, al ser variables como la nacionalidad del individuo, el tipo de colegio (privado o público) al que asisten, entre otros datos, éstas también pudieron ser utilizadas para realizar una partición de la base y hacer un análisis enfocado a grupos más homogéneos dentro de la población, por ejemplo: un análisis separando a los individuos por nacionalidad para después comparar los resultados obtenidos, cuestión que se deja abierta para futuras investigaciones.

De la pregunta 1 a la 152 por definición las respuestas son del tipo booleano, por ende no se realizó ninguna modificación a esas variables.

Variables asociadas con el consumo

Para este trabajo se tomó la codificación de la variable de tal forma que la respuesta sea del tipo booleano.

Sustancias legales

Preguntas originales	Recodificación final
162 y 163	163e
164 y 165	165e
168 y 169	169e

Ejemplo:

Preguntas originales

162: ¿Qué edad tenías cuando por primera vez tuviste la oportunidad de probar el tabaco?

163: ¿Qué edad tenías cuando consumiste tabaco por primera vez?

Recodificación final

163e: ¿Haz probado el tabaco?

Respuesta

- 1 sí y
- 0 no.

Las preguntas anteriores reflejan si el usuario ha consumido por lo menos una vez en su vida alguna de las siguientes sustancias:

- Tabaco,
- Alcohol y
- Chicha.

Sustancias ilegales

Preguntas originales	Posibles recodificaciones	Recodificación final
172 y 173	173e	173e
176 y 177	177e	177e
179 y 180	180e	180e
186 y 187	187e	187e
188	188e	188e
189	189e	189e
191	191e	191e
206	206c, 206s, 206yr, 206db y 206wb	206yr
216	216c, 216s, 216yr, 216db y 216wb	216yr
218	218c, 218s, 218yr, 218db y 218wb	218yr
222	222c, 222s, 222yr, 222db y 222wb	222yr

Ejemplos:

Preguntas originales

172: ¿Qué edad tenías cuando por primera vez tuviste la oportunidad de probar crack?

173: ¿Qué edad tenías cuando consumiste crack por primera vez?

Recodificación

173e: ¿Has probado el crack?

- 1 sí y
- 0 no.

Preguntas originales

206: ¿Cuántas veces has hecho las siguientes actividades? Fumar marihuana.

Respuesta

- 1-Ni una sola vez,
- 2-de una a tres veces,
- 3-una vez cada 3 meses,
- 4-una vez al mes,
- 5-algunas veces cada mes,
- 6-de una a tres veces cada semana,
- 7-al menos una vez al día y
- 8-más de una vez al día.

Posibles recodificaciones

206c: En el último año, ¿Cuántas veces has fumado marihuana?

Respuesta

- 0-Ni una sola vez,

- 2-de una a tres veces,
- 4-una vez cada 3 meses,
- 12- algunas veces cada mes,
- 104-de una a tres veces cada semana y
- 365-una vez al día.

206s: En el último año, ¿Cuántas veces has fumado marihuana?

Respuesta (se combinan las respuestas 7 y 8 de la pregunta original en la opción 7 de esta pregunta).

- 1-Ni una sola vez,
- 2-de una a tres veces,
- 3-una vez cada 3 meses,
- 4-una vez al mes,
- 5-algunas veces cada mes,
- 6-de una a tres veces cada semana y
- 7-al menos una vez al día.

206yr:En el último año,¿Has fumado marihuana?.

Respuesta

- 1 sí la respuesta respuesta original está en $\{2, 3, 4, 5, 6, 7, 8\}$ y
- 0 en otro caso.

206db:En el último año,¿Has fumado marihuana?.

Respuesta

- 1 sí la respuesta respuesta original está en $\{7, 8\}$ y

- 0 en otro caso.

206wb: En el último año,¿Has fumado marihuana?.

Respuesta

- 1 sí la respuesta respuesta original está en {6, 7, 8} y
- 0 en otro caso.

Las preguntas anteriores reflejan si el usuario ha consumido por lo menos una vez en su vida alguno de los siguientes estupefacientes:

- Crack,
- Éxtasis (MDMA),
- Inhalantes,
- Opiáceos (heroína, morfina),
- Anfetaminas,
- Metanfetamina,
- Tranquilizantes,
- Marihuana y
- Cocaína.

Estas codificaciones se realizaron para detectar a las personas que habían consumido algún estupefaciente por lo menos una vez en su vida. Aunque esto no implica que el sujeto vaya a desarrollar un problema, el hecho de que ya esté en contacto con sustancias psicoactivas aumenta la probabilidad de que el sujeto vuelva a consumir.

Tomando en cuenta las preguntas anteriores se definió la variable *consumo* de la siguiente manera:

$$consumo = \begin{cases} 1 & \text{si } q_i = 1 \text{ para algún } i \in \{173e, 177e, 180e, 187e, 188e, 189e, \\ & 191e, 206yr, 216yr, 218yr, 222yr\} \\ 0 & \text{e.o.c.} \end{cases} \quad (4.1)$$

Donde q_i representa la recodificación utilizada para la pregunta(s) k con $k \in \{172, 173, 176, 177, 179, 180, 186, 187, 188, 189, 191, 206, 216, 218, 222\}$.

Esta variable divide a la muestra en dos clases (consumidor y no consumidor) y será vital ya que el propósito de este trabajo es lograr clasificar a un sujeto dentro de estas dos categorías utilizando la información del cuestionario.

Sólo 10,382 de los 12,797 sujetos pueden ser clasificados como personas que consumieron o no alguno de los estupefacientes marcados en el estudio, es por eso que tomaremos a esa población como nuestra población inicial, teniendo un total de 1,924 (18.5%) “consumidores”.

Variables asociadas a la disponibilidad de los estupefacientes

En el caso de las variables anteriores, las recodificaciones fueron hechas en el mismo estudio, es por eso que en esos casos, sólo se escogió la que fuese de utilidad. Debido a que para las variables asociadas a la disponibilidad de estupefacientes (VADE) no existía una recodificación se procedió a recodificarlas de la siguiente manera:

Pregunta original	¿Para ti y tus amigos qué tan fácil es conseguir...?	Respuesta
167	Tabaco	1 Muy fácil
171	Alcohol	2 Fácil
175	Crack	3 Difícil
178	Éxtasis (MDMA)	4 Muy difícil
183	Inhalantes	
193	Marihuana	

Recodificación: $\begin{cases} 1 & \text{si la respuesta es } \textit{muy fácil} \text{ o } \textit{fácil}. \\ 0 & \text{si la respuesta es } \textit{muy difícil} \text{ o } \textit{difícil}. \end{cases}$
--

Con esta codificación binaria lo que se logra es separar a la población en dos grupos dependiendo su respuesta: el primero tiene un fácil acceso a este tipo de sustancias y esto aumenta las posibilidades de consumirla, y el segundo que es un grupo al que le resulta complicado conseguir este tipo de sustancias.

Variables asociadas a la percepción que el sujeto tiene con respecto a los riesgos que el consumo de estupefacientes implica

Al igual que en el caso de las VADE, tampoco existía una recodificación para estas variables. Éstas fueron recodificadas de la siguiente manera:

Pregunta original	¿Qué tanto piensa que la gente se arriesga a dañarse, si ellos consumen... regularmente?	Respuesta
166	Tabaco	1 Ningún riesgo
170	Alcohol	2 Riesgo leve
174	Crack	3 Riesgo moderado
181 y 182	Inhalantes	4 Gran riesgo
192	Marihuana	5 No sé

Recodificación:	$\begin{cases} 1 & \text{si la respuesta es } \textit{ningún riesgo}, \textit{ riesgo leve} \textit{ o no sé.} \\ 0 & \text{si la respuesta es } \textit{riesgo moderado} \textit{ o gran riesgo.} \end{cases}$
-----------------	---

En este caso lo que se pretende con esta codificación es separar por su respuesta a los sujetos que estén poco informados de la situación o perciban el riesgo de consumir este tipo de sustancias como algo menor, de los que catalogan el consumo de estupefacientes como una actividad que implica un alto riesgo, ya que esto es un indicio de su postura con respecto a consumir o no.

Es importante mencionar que existe una gran cantidad de datos faltantes en la base de datos. Para darle solución a ese problema se utilizaron sólo los registros que estuviesen completos, sin embargo, para lograr que la reducción de la muestra no fuera tan drástica, primero se redujo el número de variables a utilizar en el análisis y posteriormente se eliminaron los registros con datos faltantes en esas variables en específico.

4.2. Reducción de dimensionalidad

El problema de reducción de dimensionalidad representa el tema central en la simplificación de datos, la elección de la representación, reducción o transformación de las características de estos últimos es una de las cuestiones más importantes al momento su manejo. [51]

Este procesamiento preliminar de la información es importante debido a que:

- Permite la reducción del costo en tiempo: el hecho de que se esté utilizando una menor cantidad de datos implica que la cantidad de operaciones realizadas por los algoritmos también disminuya.
- Ayuda a la obtención de mejores resultados: por lo general, esperamos que al utilizar sólo las características más relevantes, un modelo o algoritmo obtenga resultados de una manera más rápida y precisa, las características irrelevantes pueden inducir a un proceso con errores, mientras que los datos redundantes pueden complicar el proceso y provocar resultados inesperados.
- Permite una mejor representación e interpretación de los datos: la simplicidad de la representación obtenida usualmente con la reducción de dimensionalidad, a menudo implica modelos que pueden ser mejor interpretados.
- Minimiza la cantidad de datos necesarios para el análisis: Al reducir la dimensionalidad podemos establecer una recolección de datos más reducida lo que permite ahorrar tiempo y dinero al eliminar características redundantes o irrelevantes.

Lo ideal sería reducir el tiempo, mejorar la precisión y simplificar la representación al mismo tiempo utilizando la reducción de la dimensionalidad, sin embargo, cuando ganamos en algunos aspectos perdemos en otros, por lo que se debe buscar un equilibrio entre ellos.

Para realizar la reducción de dimensionalidad en este trabajo se utilizó el análisis de cúmulos, no obstante, existen varias técnicas para realizar esta tarea como por ejemplo: el análisis de componentes principales, análisis de factores, mapas auto-organizados, etc. A continuación procederé a dar una descripción general de las técnicas antes mencionadas, para no pasar por alto estos métodos, los cuales también pudieron ser utilizados en esta parte del trabajo.

Análisis de componentes principales

El método estadístico más popular para la reducción de dimensionalidad es el análisis de componentes principales (PCA por sus siglas en inglés). En este método se tienen n observaciones de m variables, es decir, tenemos un conjunto de variables (x_1, x_2, \dots, x_m) , y lo que se busca es representar esa información con un menor número de variables perdiendo la menor cantidad de información posible, para esto se construyen unas nuevas variables (y_1, y_2, \dots, y_m) como una combinación lineal de las variables originales pero con la característica de almacenar la mayor parte de la información (varianza) en las primeras dimensiones y no estar correlacionadas, debido a que la varianza depende de la escala de las variables se estandarizan las variables originales antes de buscar a las nuevas.

$$y_i = a_{i1}x_1 + a_{i2}x_2 + \dots + a_{im}x_m$$

Para obtener los coeficientes que hacen que las variables y_i cumplan con las características antes mencionadas, se construye la matriz de covarianzas de las variables (estandarizadas) con los datos de nuestras observaciones, y se puede demostrar que los coeficientes de las variables y_i resultan ser los vectores propios de esa matriz, en donde los coeficientes de la primera componente principal corresponden al vector propio asociado al valor propio más grande, los coeficientes de la segunda componente principal corresponden al vector asociado al segundo valor propio más grande y así sucesivamente.

Una vez obtenidas las componentes principales, se tiene que decir con cuantos se va a trabajar, para esto existen múltiples métodos para definir el número de componentes, algunos de estos son:

- Seleccionar componentes hasta cubrir una proporción determinada de varianza.
- Desechar aquellos componentes asociados a valores propios inferiores a una cota.

Análisis de factores

El análisis de factores es un modelo estadístico de reducción de datos que pretende expresar un conjunto de p variables observables $X = (x_1, x_2, \dots, x_p)$ como la combinación lineal de k variables denominadas factores $f = (f_1, f_2, \dots, f_k)$ más expresiones de error $u = (u_1, u_2, \dots, u_p)$. [55]

$$\begin{aligned}x_1 &= \alpha_{11}f_1 + \dots + \alpha_{1k}f_k + u_1 \\x_2 &= \alpha_{21}f_1 + \dots + \alpha_{2k}f_k + u_2 \\&\vdots \\x_p &= \alpha_{p1}f_1 + \dots + \alpha_{pk}f_k + u_p\end{aligned}$$

o visto de manera matricial:

$$X = Af' + u'$$

donde:

$$A = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1k} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{p1} & \alpha_{p2} & \cdots & \alpha_{pk} \end{bmatrix} \quad (4.2)$$

En este caso, los términos α_{ij} son los pesos de los factores que muestran el grado de dependencia de las variables originales y los factores, es decir, son las correlaciones entre las variables y los factores. Los términos errores u_i son variables asociadas a las variables originales que se supone no está correlacionadas entre sí, ni con los factores. Otra suposición hecha por este modelo es que tanto las variables de los vectores u y f están estandarizadas.

Debido a las suposiciones hechas se puede demostrar que la varianza de las variables x_i es:

$$\sigma_i^2 = \sum_{j=1}^k \alpha_{ij}^2 + \psi_i$$

Con ψ_i la varianza de la variable u_i .

Lo cual quiere decir que la varianza de las variables observadas se puede separar en dos partes, la primera suele denotarse como h^2 , denominada comunalidad.

$$h_i^2 = \sum_{j=1}^k \alpha_{ij}^2$$

Ésta representa la varianza de las variables originales explicada por los factores, y la segunda parte ψ_i , también llamada unicidad, representa la varianza explicada por el factor único u_i . [55]

Dada la construcción del modelo, también se puede demostrar que la matriz de covarianzas Σ para las variables originales se puede ver como el producto de la matriz de pesos de los factores por su transpuesta más una matriz ψ con las varianzas de las variables u_i en la diagonal. [55]

$$\Sigma = AA' + \psi$$

En la práctica, lo que se hace es estimar los parámetros del modelo a partir de un conjunto de observaciones, es decir, se busca las matrices \hat{A} y $\hat{\psi}$, tal que:

$$S \approx \hat{A}\hat{A}' + \hat{\psi}$$

Donde S es la matriz de covarianzas muestral. [55]

Mapas auto-organizados de Kohonen

Las redes de neuronas artificiales también ofrecen una herramienta que puede ser útil para reducir la dimensionalidad: los mapas auto-organizados de Kohonen. Como ya mencioné en la sección 2.4.2 éstos son un tipo de red neuronal de entrenamiento no supervisado que se compone de nodos (neuronas), y para cada neurona hay un vector de pesos asociado. Estas redes describen un mapeo de un espacio de una dimensión mayor en uno de menor dimensión, básicamente lo que hacen es asociar un vector con la neurona cuyo vector de pesos se encuentre más cerca (utilizando alguna métrica o medida de similitud), afectando el vector de

pesos de las neuronas adyacentes, este proceso se realiza con un conjunto de vectores de entrada conocido como conjunto de entrenamiento, para posteriormente lograr clasificar otros vectores que no han sido pasados por la red.

Al agrupar un conjunto de vectores que presenten un mayor grado de similitud, esta herramienta puede ser utilizada para agrupar el conjunto de variables en grupos relativamente homogéneos que se pretende que no estén relacionados entre sí.

Una vez mencionados de manera muy breve algunos métodos para realizar la reducción de dimensionalidad, pasaré a la explicación del método utilizado en este trabajo para manejar la gran cantidad de variables en la muestra.

4.2.1. Variables no relacionadas con el consumo

Como mencionamos anteriormente, la decisión de consumir o no algún estupefaciente está basada en un conjunto de factores y cómo es que estos se relacionan, es por eso que en primer lugar se eliminaron aquellas variables que no pudiesen ser consideradas como un factor de riesgo, como por ejemplo la variable 194, la cual nos dice si el sujeto ve televisión o la variable 221 que muestra si el sujeto se cepilla los dientes, las variables eliminadas en esta sección fueron las siguientes.

Tabla 4.1: Variables no relacionadas con el consumo.

153	¿De qué grupos étnicos es tu familia?
154	¿Cómo te sientes con relación a tu grupo étnico o cultural?
155	¿Te gustaría aprender más acerca de la historia y las costumbres de tu grupo étnico o cultural?
156	¿Qué piensas de tu grupo étnico o cultural?
157	¿Cuál es tu religión?
158	¿Cuántos años fue a la escuela tu papá?
159	¿Cuántos años fue a la escuela tu mamá?

Tabla 4.1: Variables no relacionadas con el consumo.

184*	Acerca de la droga Cadrina ¿A qué edad tuvo su primera oportunidad de consumir la droga Cadrina?
185*	¿A qué edad probó la droga Cadrina por primera vez?
194	En el último año, ¿Cuántas veces has hecho las siguientes actividades?: Ver televisión.
195	En el último año, ¿Cuántas veces has hecho las siguientes actividades?: Jugar fútbol, béisbol, baloncesto u otro juego de equipo.
196	En el último año, ¿Cuántas veces has hecho las siguientes actividades?: Escuchar la radio o música.
197	En el último año, ¿Cuántas veces has hecho las siguientes actividades?: Tocar un instrumento musical.
198	En el último año, ¿Cuántas veces has hecho las siguientes actividades?: Trabajar para ganar dinero.
200	En el último año, ¿Cuántas veces has hecho las siguientes actividades?: Ir de citas o cortejar.
205	En el último año, ¿Cuántas veces has hecho las siguientes actividades?: Cocinar.
207	En el último año, ¿Cuántas veces has hecho las siguientes actividades?: Practicar deportes competitivos que implican mucho ejercicio como correr largas distancias o correr a campo traviesa.
209	En el último año, ¿Cuántas veces has hecho las siguientes actividades?: Cuidar niños
211	En el último año, ¿Cuántas veces has hecho las siguientes actividades?: Ir a conciertos de música rock.
213	En el último año, ¿Cuántas veces has hecho las siguientes actividades?: Leer el periódico.
215	En el último año, ¿Cuántas veces has hecho las siguientes actividades?: Salir a bailar.

Tabla 4.1: Variables no relacionadas con el consumo.

217	En el último año, ¿Cuántas veces has hecho las siguientes actividades?: Ir al dentista.
221	En el último año, ¿Cuántas veces has hecho las siguientes actividades?: Cepillarse los dientes.
224	En el último año, ¿Cuántas veces has hecho las siguientes actividades?: Trotar, caminar o hacer ejercicios ligeros.

* Estas preguntas fueron eliminadas debido a que no se contaba con respuestas para ellas en ninguno de los registros.

4.2.2. Análisis de cúmulos

Análisis de cúmulos por columnas

Una vez eliminadas las variables no relacionadas al consumo se procedió a realizar un análisis de cúmulos por columnas (variables), con lo cual se pretende reducir aún más el número de variables.

En este caso se utilizó un método jerárquico aglomerativo de distancia promedio, utilizando como medida de distancia la métrica Hamming, ya que ésta nos da una buena idea de que tan parecidas son las variables, no solo en las respuestas de los sujetos sino también en la no respuesta, esta métrica tendrá un valor pequeño asociado a dos variables cuando tengan no solo las mismas respuestas sino también cuando compartan los mismo datos faltantes, esto quiere decir que si algún sujeto no responde a una de estas preguntas es muy probable que tampoco lo haga para la otra.

Para formar los cúmulos se realizó un corte en el dendrograma 4.1 a una altura de 0.2, lo cual implica (dada la métrica utilizada) que las variables que están en un mismo cúmulo tienen en alrededor del 80 % de sus entradas el mismo valor indicado un alto grado de similitud, es por eso que se decidió dejar como variable

explicativa a un representante de cada cúmulo, eligiendo a aquella que presentara el menor número de datos faltantes y eliminando las otras.

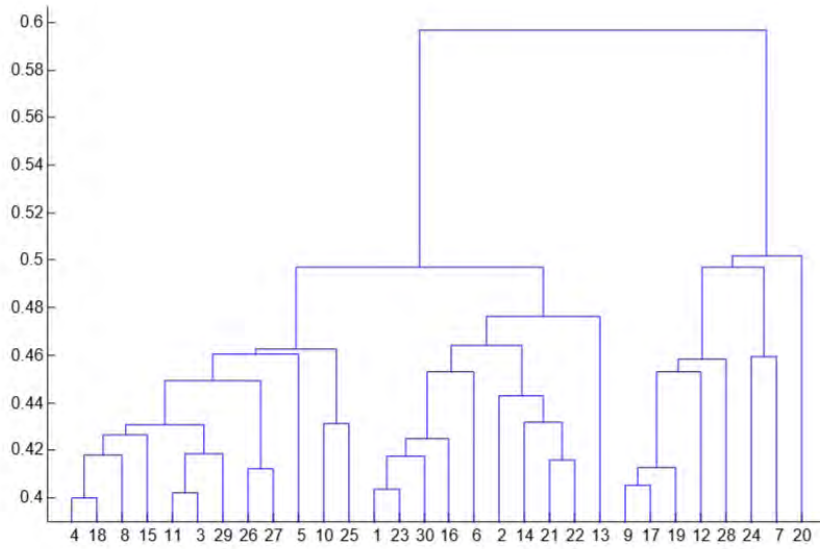


Figura 4.1: Dendrograma

Es importante hacer notar que debido al gran número de variables que estamos analizando, el representarlas todas en el dendrograma 4.1 dificultaría su apreciación, es por esto que en este caso se realizó la unión de los cúmulos más cercanos mostrando sólo los últimos 30, es decir, los 30 cúmulos más lejanos.

Los cúmulos que cumplen la condición de tener más de un elemento y que estos a su vez tengan una distancia asociada menor igual a 0.2 fueron 8, a continuación se procederá a mostrar los resultados obtenidos.

Cúmulo 1	
Variables asociadas 22, 47 y 65	
22	¿Siempre pides permiso a tus padres cuando sales de la casa a divertirte?
47	¿Generalmente tus padres o encargados han estado enterados de dónde estás y qué haces?
65	Algunos jóvenes siempre le dicen a sus padres a donde van cuando salen de la casa y otros jóvenes a veces no les cuentan a sus padres a donde van. Durante los últimos 6 meses, cuando has salido de la casa, ¿Siempre le has dicho a tus padres a dónde vas?
Variable con menor número de datos faltantes: 22	

Cúmulo 2	
Variables asociadas 3, 19, 31, 35, 54, 105, 119, 125 y 152	
3	¿La gente que me rodea considera que soy una persona amigable?
19	¿He disfrutado el estar con otras personas?
31	¿Durante los últimos 6 meses, generalmente has sido una persona sociable?
35	¿Has sentido que eres importante para tus padres o encargados?
54	Algunos jóvenes piensan que es buena idea usar drogas y otros piensan que es una mala idea usar drogas. ¿Algunos amigos piensan que usar drogas es una mala idea?
105	A algunas personas les gusta estar con otras personas, mientras que otros prefieren estar solos. Durante los últimos 6 meses, ¿has querido estar con otras personas durante tu tiempo libre?
119	¿Has tenido muchos amigos que piensan que usar drogas es una mala idea?
125	¿He disfrutado el mirar televisión?
152	¿Para tu familia, respetar la ley es importante?
Variable con menor número de datos faltantes: 152	
Cúmulo 3	
Variables asociadas 48 y 58	
48	¿Has herido o hecho daño a los animales?
58	Durante los últimos 12 meses, ¿has molestado animales?
Variable con menor número de datos faltantes: 48	
Cúmulo 4	
Variables asociadas 28 y 127	
28	He tenido excelentes relaciones con los profesores (en general).
127	He tenido excelentes relaciones con mis profesores.
Variable con menor número de datos faltantes: 127	

Cúmulo 5	
Variables asociadas 167 y 171	
167	Para Ud. y sus amigos, ¿Qué tan fácil o difícil es conseguir tabaco?
171	Para Ud. y sus amigos, ¿Qué tan fácil o difícil es conseguir alcohol?
Variable con menor número de datos faltantes: 167	
Cúmulo 6	
Variables asociadas 103,108 y 109	
103	¿Has tenido muchos amigos que han tratado de tener una conducta excelente?
108	Durante los últimos 6 meses, ¿has tenido algún amigo que le gusta obedecer las reglas?
109	A veces tenemos amigos que tratan de ser muy buenos y de tener una conducta excelente. Durante los últimos 6 meses, ¿has tenido algún amigo que trata de tener una conducta excelente?
Variable con menor número de datos faltantes: 103	
Cúmulo 7	
Variables asociadas 21 y 115	
21	¿Te has aprovechado de otras personas?
115	Rara vez he olvidado lo que aprendí en el instituto o colegio.
Variable con menor número de datos faltantes: 115	

Cúmulo 8	
Variables asociadas 25, 37, 55, 59, 77, 92, 102, 116, 117, 122, 123, 124, 148, 149, 150, 199, 201, 203 y 212	
25	¿Has faltado a la escuela con frecuencia?
37	¿Has pertenecido a alguna pandilla?
55	¿Has pensado seriamente en abandonar los estudios?
59	Durante los últimos 12 meses, ¿has tenido problemas con la policía?
77	¿Has sido suspendido del instituto o colegio?
92	Algunos jóvenes prefieren ir a trabajar en vez de ir a la escuela. Durante los últimos 6 meses, ¿has preferido ir a trabajar que ir a la escuela?
102	He pensado en abandonar el instituto o colegio completamente.
116	¿Has tenido amigos a quienes les gusta inhalar pegamento o gasolina?
117	Algunos jóvenes están más contentos cuando piensan en trabajar en vez de ir al colegio o el instituto. ¿Has estado más contento pensando en ir a trabajar?
122	Algunos jóvenes han comenzado a usar pasta base, cocaína o crack. ¿Has tenido algún amigo que haya usado pasta base, cocaína, o crack?
123	¿Has tenido muchos amigos que han usado pasta base, cocaína, o crack?
124	A veces la gente joven dice 'ir al instituto ó colegio es una pérdida de tiempo'. Para ti, los últimos 6 meses, ¿han sido una pérdida de tiempo ir al instituto ó colegio?
148	Ha consumido algún miembro de tu familia (Madre, Padre, Hermano(a)) marihuana?
149	¿Has sentido que corrías peligro en tu casa?

150	¿Ha consumido algún miembro de tu familia (Madre, Padre, Hermano(a)) cocaína?
199	En el último año, ¿Cuántas veces has hecho las siguientes actividades?: Ir al bar ó cantina.
201	En el último año, ¿Cuántas veces has hecho las siguientes actividades?: Fumar tabaco.
203	En el último año, ¿Cuántas veces has hecho las siguientes actividades?: Apostar por dinero.
212	En el último año ¿Cuántas veces has hecho las siguientes actividades?: Consumir alcohol, cerveza, guaro, etc.
Variable con menor número de datos faltantes: 102	

Como podemos ver el algoritmo utilizado para formar estos cúmulos en la mayoría de los casos agrupa variables que intuitivamente podríamos pensar que están relacionadas como por ejemplo: en el primer cúmulo todas las variables hacen referencia a la comunicación entre padres e hijos, en el cúmulo tres las variables se relacionan con maltrato animal, en el cúmulo seis todas las variables están relacionadas a las conductas de los amigos, sin embargo, para evaluar la calidad de este análisis, se calculó el coeficiente cofenético (0.7928), el cual es relativamente alto, y cabe mencionar, mayor al obtenido con otros métodos como por ejemplo: método jerárquico aglomerativo distancia más cercana, distancia promedio, ambos utilizando la misma métrica.

Una vez eliminadas estas variables y tomando en cuenta las que ya habíamos quitado (variables no relacionadas con el consumo) en total se obtienen 144 variables las cuales servirán para entrenar a la red de neuronas artificiales, por lo tanto, se les dará el nombre de variables explicativas..

Variables explicativas
001, 002, 004, 005, 006, 007, 008, 009, 010, 011, 012, 013, 014, 015, 016, 017, 018, 020, 022, 023, 024, 026, 027, 029, 030, 032, 033, 034, 036, 038, 039, 040, 041, 042, 043, 044, 045, 046, 048, 049, 050, 051, 052, 053, 056, 057, 060, 061, 062, 063, 064, 066, 067, 068, 069, 070, 071, 072, 073, 074, 075, 076, 078, 079, 080, 081, 082, 083, 084, 085, 086, 087, 088, 089, 090, 091, 093, 094, 095, 096, 097, 098, 099, 100, 101, 102, 103, 104, 106, 107, 110, 111, 112, 113, 114, 115, 118, 120, 121, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 151, 152, 163, 165, 166, 167, 169, 170, 174, 175, 178, 181, 182, 183, 192, 193, 202, 204, 208, 214, 219, 220 y 223.

En esta parte del estudio se eliminaron también los renglones con datos faltantes en esas 144 variables en específico con lo cual la base de datos se reduce a 5,111 registros con los datos completos.

Análisis de cúmulos por renglones

Para facilitar el entrenamiento de la RNA se hizo un análisis de cúmulos por renglones (sujetos), en el cual debido a que todas las respuestas son del tipo booleano se utilizó el método de distancia promedio (métrica Hamming) con lo cual al realizar el corte a altura 0.2 en el dendrograma la base se redujo de 5,111 a 4,445 registros.

Cabe destacar que debido al corte a esa altura en específico los sujetos de un mismo conglomerado comparten las respuestas en alrededor del 80 % de las preguntas, mostrando una gran similitud por lo cual se consideró un sujeto por conglomerado. De los 4,445 sujetos a analizar, 862 pertenecen a la clase “consumidor”, los cuales representan el 19.39 % del total, esta proporción resulta parecida a la que teníamos en la muestra de datos inicial en donde el 18.5 % del total había consumido algún estupefaciente, lo cual refleja que la proporción entre consumidores y no consumidores no cambió de una manera significativa.

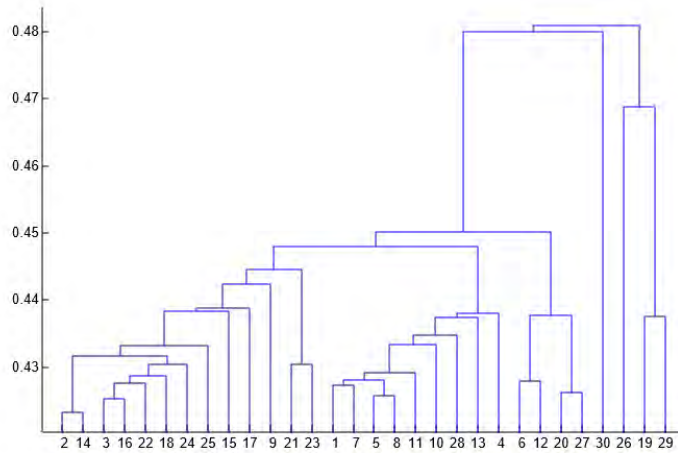


Figura 4.2: Dendrograma: al igual que en el dendrograma 4.1 sólo se muestran los 30 cúmulos más lejanos (debido a la gran cantidad de datos).

4.3. Redes de neuronas artificiales utilizadas

4.3.1. Descripción de las redes

Para lograr obtener mejores resultados al momento de realizar la predicción se entrenó un total de 5 redes de neuronas artificiales las cuales constan de 144 entradas (variables explicativas), 1 capa oculta con 50 neuronas con función de activación sigmoidea (tangente hiperbólica) y una capa de salida con una neurona que tiene la misma función de activación.

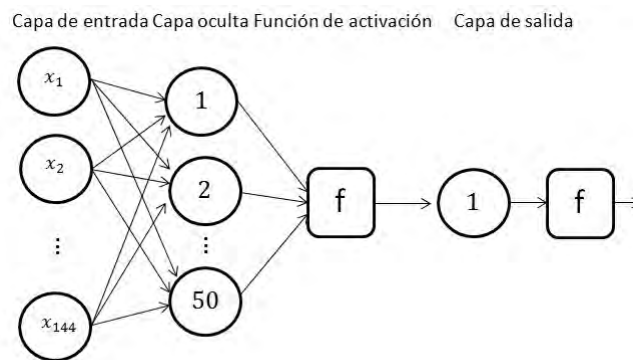


Figura 4.3: Estructura de las redes de neuronas artificiales.

4.3.2. Entrenamiento

Fase 1

Se separó el conjunto de datos (4,445 sujetos) en dos subconjuntos, el subconjunto E con el 75 % de las observaciones para el entrenamiento y el resto (subconjunto P) se utilizó para la predicción, cabe destacar que se tomó el 75 % de los consumidores y el 75 % de los no consumidores para que se mantuviesen las proporciones.

Del subconjunto de entrenamiento 647 sujetos habían consumido algún estupefaciente de la lista de interés por lo menos una vez en su vida, es decir, presentaban el valor 1 en la variable *consumo* y 2,687 un valor de 0.

Fase 2

En esta parte se procedió al entrenamiento de las redes, las cuales difieren entre sí sólo en el conjunto de datos que se les dio para ser entrenadas, esto se verá con mayor claridad con la descripción del entrenamiento para cada una de las redes.

Red 1

Para esta red se seleccionó del subconjunto E un conjunto de valores de entrenamiento con la misma cantidad de consumidores y no consumidores, es decir, 647 sujetos con la variable consumo igual a 1 y 647 sujetos con la variable consumo igual a 0.

Al igual que las redes que veremos después, ésta se entrenó mediante el algoritmo de propagación hacia atrás utilizando como función error: la suma de cuadrados de los errores(SCE). Esta red tiene un SCE de 0.1198 al final del entrenamiento. Al momento de realizar la predicción la red logra identificar cada una de las clases de manera correcta el 53.56 % de las veces, y aunque en general es un mal resultado, ésta nos será de utilidad ya que logra identificar a los consumidores relativamente bien (81.86 % de efectividad).

Red 1 (Predicción)	Resultado deseado	
Resultado de la red	1	0
1	176	477
0	39	419
Error	18.14 %	53.24 %
Clasificación correcta	53.56 %	

Red 2

Para esta red se utilizó un conjunto de entrenamiento de 647 consumidores y 971 no consumidores o sea una proporción de 1-1.75, obteniendo en la suma de cuadrados de los errores un valor de 0.1347 al final del entrenamiento.

Red 2 (Predicción)	Resultado deseado	
Resultado de la red	1	0
1	139	268
0	76	628
Error	35.35 %	29.91 %
Clasificación correcta	69.04 %	

Como vemos en este caso se obtienen resultados más equilibrados con respecto a la identificación de las dos clases, 64.65 % de resultados correctos al clasificar a los consumidores y 70.09 % en el caso de los no consumidores.

Red 3

En este caso el conjunto de entrenamiento está conformado por 647 consumidores y 1,294 no consumidores (proporción 1-2), obteniendo una SCE de 0.1317 al final del entrenamiento.

Red 3 (Predicción)	Resultado deseado	
Resultado de la red	1	0
1	127	220
0	88	676
Error	40.93 %	24.55 %
Clasificación correcta	72.27 %	

Esta red identifica de manera correcta al 59.07 % de los consumidores y al 75.45 % de los no consumidores.

Red 4

La proporción del conjunto de entrenamiento para esta red fue 1-2.5, 647 consumidores 1,618 no consumidores, obteniendo un valor en la suma de cuadrados de los errores de 0.1442.

Red 4 (Predicción)	Resultado deseado	
Resultado de la red	1	0
1	114	123
0	101	773
Error	46.98 %	13.73 %
Clasificación correcta	79.8 %	

Esta red identifica de manera correcta al 53.02 % de los consumidores y al 86.27 % de los no consumidores.

Red 5

El conjunto de entrenamiento se conforma por 647 consumidores y 1941 no consumidores (proporción 1-3), obteniendo un SCE de 0.1442.

Red 5 (Predicción)	Resultado deseado	
Resultado de la red	1	0
1	83	95
0	132	801
Error	61.4 %	10.6 %
Clasificación correcta	79.6 %	

Esta red identifica de manera correcta al 38.6 % de los consumidores y al 89.4 % de los no consumidores.

A continuación se procederá a graficar el desempeño de las 5 redes en el espacio ROC, pero antes de hacerlo se darán las definiciones necesarias para su construcción.

Al tener una tabla de confusión como en nuestro caso, las clasificaciones de la red se pueden agrupar en 4 grupos:

Red	Resultado deseado	
Resultado	1	0
1	Verdaderos positivos (VP)	Falsos positivos (FP)
0	Falsos negativos (FN)	Verdaderos negativos (VN)
Total	P=VP+FN	N=FP+VN

Dado lo anterior, también podemos definir:

La razón de falsos positivos

$$FPR = \frac{FP}{FP + VN}$$

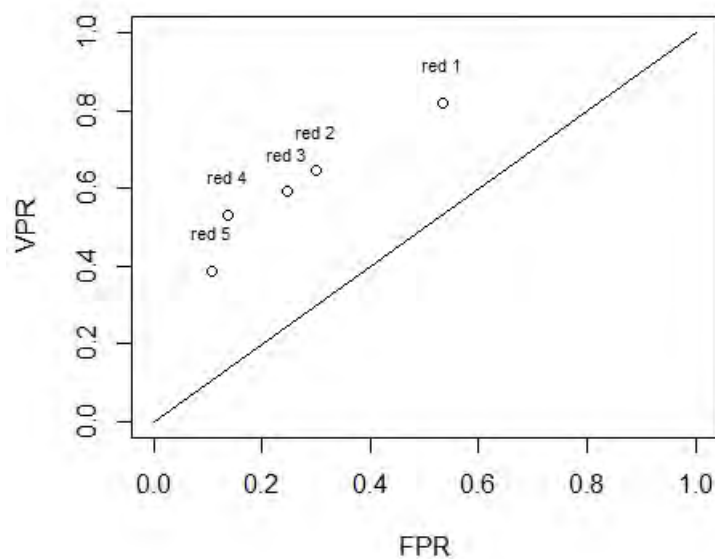
La razón de verdaderos positivos

$$VPR = \frac{VP}{VP + FN}$$

El espacio ROC se define de la siguiente manera: el eje x representa la razón de falsos positivos (FPR), y el eje y la razón de verdaderos positivos (VPR).

En esta gráfica, el punto correspondiente a una red cuyas predicciones son las mejores posibles se situaría en la esquina superior izquierda, o coordenada (0,1) del espacio ROC, lo cual indica que no existe ningún falso negativo y que tampoco existe ningún falso positivo. A este punto (0,1) también se le llama una clasificación perfecta. Por el contrario, una clasificación totalmente aleatoria daría un punto a lo largo de la línea diagonal, que se llama también línea de no-discriminación, desde el extremo inferior izquierdo hasta la esquina superior derecha, esto quiere decir que entre más cerca se encuentre nuestra red de la coordenada (0,1) sus resultados serán mejores.

La siguiente gráfica muestra a nuestras redes en el espacio ROC:



Como vemos, en este caso todas las redes se encuentran por encima de la diagonal, lo cual indica que sus predicciones son relativamente buenas, en términos generales.

Es importante destacar que los resultados mostrados en las tablas anteriores corresponden a los datos del subconjunto P , es decir, es la predicción de las redes para sujetos que no estuvieron involucrados en el entrenamiento.

4.3.3. Resultados

Como vemos, al modificar la proporción de datos del conjunto de entrenamiento se observa un patrón en los resultados, cuando se le dan más elementos de una clase a la red, ésta tiende a identificarlos de una mejor manera, sin embargo, para la otra clase la red pierde en cierta medida su efectividad, mientras que unas identifican mejor a los consumidores otras lo hacen mejor para los no consumidores, es por eso que se decidió utilizar la combinación de varias redes.

Para predecir si los sujetos eran o no “consumidores” se definió la variable *consumo red* de la siguiente manera:

$$consumo\ red = \begin{cases} 1 & \text{si } \sum_{i=1}^5 red_i > 2 \\ 0 & \text{En otro caso.} \end{cases}$$

Donde red_i representa el resultado obtenido por la red i para ese sujeto, esto quiere decir que clasificamos al sujeto como “consumidor” cuando por lo menos tres de las cinco redes lo hacen.

Realizando lo anterior, los resultados obtenidos para el conjunto de predicción P fueron los siguientes:

Consumo red	Resultado deseado	
	1	0
Resultado de la red	1	0
1	134	189
0	81	707
Error	37.67 %	21.09 %
Clasificación correcta	75.7 %	

En este caso, al combinar las redes se logró un mayor equilibrio al momento de la clasificación, dicho de otra forma, la variable *consumo red* obtiene mejores resultados para las dos clases y no sólo para una como las redes vistas de manera individual.

El 62.33 % de los consumidores y el 78.91 % de los no consumidores fueron clasificados de manera correcta, lo que al final nos arroja resultados satisfactorios el 75.7 % de las veces.

En general, utilizando todo el conjunto de datos, es decir, los 4,445 sujetos de la muestra obtenemos una clasificación correcta en el 80.16 % de los sujetos, esto ya que en este conjunto están incluidos los sujetos utilizados para el entrenamiento, a continuación se mostraran los resultados para toda la muestra.

Consumo red	Resultado deseado	
Resultado de la red	1	0
1	530	550
0	332	3033
Error	38.52 %	15.35 %
Clasificación correcta	80.16 %	

Capítulo 5

Conclusiones

Este trabajo representa una aplicación de las redes de neuronas artificiales y el análisis de cúmulos en lo que considero un problema importante con gran impacto en todos los sectores de la sociedad, esto es: el consumo de estupefacientes. Debido a que la venta de estas sustancias es parte importante del financiamiento del crimen organizado, esto ocasiona olas de violencia, terror, corrupción, etc. que aunados a los problemas de salud, económicos y muchos otros relacionados dañan gravemente a la sociedad.

En este trabajo se abordó la problemática específica referente a la identificación de los consumidores de estupefacientes de manera automatizada, esto con el fin de mostrar la capacidad que tienen ciertas herramientas, en este caso las redes de neuronas artificiales, para realizar dicha tarea. Es importante destacar la importancia de la creación de métodos para identificar el consumo de este tipo de sustancias en etapas tempranas y de manera automática, ya que en las primeras etapas de consumo el tratamiento resulta ser más efectivo y el hecho de que sea de manera automática ayuda a la reducción de recursos y tiempo invertidos, por lo que puede analizarse a un sector más amplio de la población.

Para lograr identificar a los consumidores se trabajó con un conjunto inicial de 224 variables que brindaban información de varios factores en la vida del individuo, dichos factores son los encargados de brindar a la red los elementos necesarios

para realizar la clasificación, sin embargo, se realizó una reducción de dimensión para que al final, con una menor cantidad de preguntas se obtuviera el análisis del sujeto, esto se hizo descartando las variables que no estuviesen relacionadas con el fenómeno, lo cual resultó en la eliminación de 24 variables, posteriormente, se realizó un análisis de cúmulos para eliminar las variables que estuviesen correlacionadas, esta herramienta arrojó grupos que al ser analizados mostraban como las preguntas asociadas a un cúmulo tenían que ver con cierta característica, lo cual intuitivamente mostraba como se agruparon las preguntas, sin embargo, para averiguar la calidad de este análisis, se calculó el coeficiente cofenético (0.7928), el cual mostro un valor relativamente alto y resultado mayor al obtenido utilizando otros métodos de agrupación como por ejemplo: método jerárquico aglomerativo distancia más cercana, método jerárquico aglomerativo distancia promedio, etc. Al finalizar con el proceso de reducción de dimensionalidad, el conjunto de variables se redujo a 144 variables explicativas, sin embargo, hay que destacar que la reducción de variables es un problema sumamente complejo e importante, en donde existe una gran cantidad de herramientas diseñadas para realizar esta tarea, como por ejemplo: el análisis de componentes principales, análisis de factores, etc., lo cual implica que tan solo el estudio detallado con respecto al análisis de las variables mediante distintos métodos resultaría el tema central de otro trabajo. Una vez establecidas las variables con las que se iba a alimentar a la red, se procedió a eliminar a los sujetos que tuviesen valores faltantes en esas variables en específico, lo cual fue posible debido a la gran cantidad de registros con los que contaba la base de datos, no obstante, es de interés para futuras investigaciones, abordar el problema de datos faltantes mediante otros métodos.

El procesamiento de la base de datos fue importante para el entrenamiento de las redes de neuronas artificiales, además de la reducción de variables, también se eliminaron los registros que tuviesen las mismas respuestas aproximadamente el 80 % de las veces, esto ya que la red utilizada en este caso: perceptrón multicapa, tiene un entrenamiento supervisado, es decir, mediante ejemplos, lo cual implica que si los sujetos cuyas respuestas son muy similares pertenecen a clases diferentes,

la red recibirá mensajes contradictorios y esto dificultará la correcta asignación de los pesos sinápticos.

Para realizar la clasificación se utilizó un total de 5 redes de neuronas artificiales que se diferenciaban entre sí solamente en el conjunto de entrenamiento E , esto mostró un patrón en donde al modificar el conjunto E la red clasificaba con un menor porcentaje de error aquella clase que presentara mayor presencia en el conjunto de entrenamiento y no precisamente mostraba buenos resultados para ambas, fue por eso que se decidió combinar los resultados obtenidos por las redes para mejorar la clasificación en las dos clases, obteniendo resultados satisfactorios al momento de la predicción el 75.7% de la veces, es importante mencionar que estos resultados sugieren desde mi punto de vista que esta herramienta puede ser utilizada como un auxiliar al momento de la clasificación, en donde especialistas en el campo tienen que complementarlas para lograr resultados mucho mejores. Con este trabajo intenté mostrar como es que un problema puede enfrentarse apoyándose en muchas áreas de conocimiento, utilizando herramientas que muchas veces fueron desarrolladas en áreas que no están directamente relacionadas con los fenómenos estudiados, sin embargo, resultan útiles para comprenderlos.

Apéndice A

Cuestionario

PAIS

PACARDO 1999

ESTE CUESTIONARIO ES ANÓNIMO, POR ESO NO DEBÉS ESCRIBIR TU NOMBRE.

INSTRUCCIONES:

Lo que vas a realizar es fácil y no se necesita haber estudiado para poder contestar las preguntas.

No es un examen y por lo tanto no hay respuestas buenas ni malas.

Lo que importa es que las repuestas que des, sean verdaderas.

Si te surgiera alguna duda o si quisieras hacer algún comentario, nosotros con gusto atenderemos tus inquietudes.

No dudes en preguntarnos si lo necesitas.

La información de este cuestionario se manejará en secreto y su finalidad es conocer opiniones y características de grupos de jóvenes estudiantes para el desarrollo y aplicación de programas preventivos

¿Tenés alguna pregunta?

PACARDO - 1999

NUMERO DE ID. DEL ASESOR PRINCIPAL:								FECHA DE HOY:						9	9	HORA PROGRAMADA:			:	
-------------------------------------	--	--	--	--	--	--	--	---------------	--	--	--	--	--	---	---	------------------	--	--	---	--

Zona A.	
1. País: _____	
2. ¿Qué edad tienes? _____	
3. Cual es tu sexo?: Masc. () Fem. ()	
4. Instituto o Colegio: _____	

Zona B.	
1. Tipo de Instituto o Colegio: 1. Privado () 2. Público ()	
2. ¿En que grado estás? _____	
3. Salón de clase: _____	
4. Comunidad de su casa: _____	

INSTRUCCIONES: Por favor, responda todas las preguntas, aunque no se ajusten exactamente a su experiencia. Marque <SI>, si la mayoría de las veces es cierto, o <NO>, si la mayoría de las veces es falso. Marque con una (X) la respuesta correspondiente. Para eso, considere los últimos 6 meses.

Zona 1: RESPONDA CON RELACION A LOS ÚLTIMOS 6 MESES.		
	Si	No
1. Durante los últimos 6 meses, ¿has estado bien de salud?	<input type="radio"/>	<input type="radio"/>
2. En mi tiempo libre frecuentemente estoy con otras personas.	<input type="radio"/>	<input type="radio"/>
3. La gente que me rodea considera que soy una persona amigable.	<input type="radio"/>	<input type="radio"/>
4. ¿Te has puesto triste con frecuencia?	<input type="radio"/>	<input type="radio"/>
5. ¿Han conocido tus padres o encargados lo que pensabas o sentías sobre las cosas importantes para ti?	<input type="radio"/>	<input type="radio"/>
6. Durante los últimos 6 meses, he estado sin nauseas.	<input type="radio"/>	<input type="radio"/>
	<input checked="" type="checkbox"/>	

Zona 3: RESPONDA CON RELACION A LOS ÚLTIMOS 6 MESES.		
	Si	No
16. ¿Has realizado la mayoría de tus actividades recreativas con otros en vez de sólo?	<input type="radio"/>	<input type="radio"/>
17. ¿Tus padres o encargados, han sido conscientes de lo que te gustaba o no te gustaba?	<input type="radio"/>	<input type="radio"/>
18. ¿Tus amigos han robado, o han causado daño a propósito a las cosas de otras personas?	<input type="radio"/>	<input type="radio"/>
19. He disfrutado el estar con otras personas.	<input type="radio"/>	<input type="radio"/>
20. Durante los últimos 6 meses, ¿has discutido mucho?	<input type="radio"/>	<input type="radio"/>
21. ¿Te has aprovechado de otras personas?	<input type="radio"/>	<input type="radio"/>
	<input checked="" type="checkbox"/>	

Zona 2. RESPONDA CON RELACION A LOS ÚLTIMOS 6 MESES.		
	Si	No
7. ¿Has estado sin dolores de estómago?	<input type="radio"/>	<input type="radio"/>
8. ¿Te has enojado con frecuencia?	<input type="radio"/>	<input type="radio"/>
9. No he dañado intencionalmente las cosas de otras personas.	<input type="radio"/>	<input type="radio"/>
10. ¿Alguno de tus amigos se ha metido en problemas con la policía?	<input type="radio"/>	<input type="radio"/>
11. ¿Es cierto que la mayoría de las veces no haces las tareas del colegio?	<input type="radio"/>	<input type="radio"/>
12. ¿Tus notas escolares han sido mejores que las de la mayoría de las de tus compañeros de clase?	<input type="radio"/>	<input type="radio"/>
13. ¿Has pasado aburrido la mayor parte del tiempo?	<input type="radio"/>	<input type="radio"/>
14. Durante los últimos 6 meses, ¿has robado algo?	<input type="radio"/>	<input type="radio"/>
15. ¿Has tenido <u>muchos</u> amigos que piensan que usar drogas es una mala idea?	<input type="radio"/>	<input type="radio"/>

Zona 4. RESPONDA CON RELACION A LOS ÚLTIMOS 6 MESES.		
	Si	No
22. ¿Siempre pides permiso a tus padres cuando sales de la casa a divertirse?	<input type="radio"/>	<input type="radio"/>
23. ¿Te has sentido calmado y en paz contigo mismo y con el resto del mundo?	<input type="radio"/>	<input type="radio"/>
24. ¿Has tenido pocos amigos en comparación con la mayoría de tus compañeros de la misma edad?	<input type="radio"/>	<input type="radio"/>
25. ¿Has faltado a la escuela con frecuencia?	<input type="radio"/>	<input type="radio"/>
26. ¿Te has sentido nervioso?	<input type="radio"/>	<input type="radio"/>
27. No he hecho nada riesgoso o peligroso.	<input type="radio"/>	<input type="radio"/>
28. He tenido excelentes relaciones con los profesores.	<input type="radio"/>	<input type="radio"/>
29. Muchos de mis amigos han fumado cigarrillos.	<input type="radio"/>	<input type="radio"/>
30. Durante los últimos 6 meses, ¿has estado mas enfermo que la mayoría de tus compañeros de la misma edad?	<input type="radio"/>	<input type="radio"/>

Zona 5	
RESPONDA CON RELACION A LOS ÚLTIMOS 6 MESES.	
	SI No
31. ¿Durante los últimos 6 meses, generalmente has sido una persona sociable?	<input type="radio"/> SI <input type="radio"/> No
32. ¿Has tenido ganas de llorar con frecuencia?	<input type="radio"/> SI <input type="radio"/> No
33. ¿Ha sido fácil para ti defender tus opiniones o puntos de vista?	<input type="radio"/> SI <input type="radio"/> No
34. ¿Has amenazado con lastimar o hacer daño a otras personas?	<input type="radio"/> SI <input type="radio"/> No
35. ¿Has sentido que eres importante para tus padres o encargados?	<input type="radio"/> SI <input type="radio"/> No
36. ¿He hecho bien mis tareas del instituto o colegio?	<input type="radio"/> SI <input type="radio"/> No
37. ¿Has pertenecido a alguna pandilla?	<input type="radio"/> SI <input type="radio"/> No
38. ¿Alguna vez has dejado de cumplir con los deberes escolares que necesitabas hacer?	<input type="radio"/> SI <input type="radio"/> No
39. A veces, la gente actúa sin pensar. Durante los últimos 6 meses, nunca he actuado sin pensar primero lo que serían las consecuencias de mis acciones.	<input type="radio"/> SI <input type="radio"/> No
40. ¿Has escuchado ruidos o voces alrededor de ti, que nadie más oía?	<input type="radio"/> SI <input type="radio"/> No
41. ¿Generalmente te has sentido mal por tu desempeño en las actividades con tus compañeros?	<input type="radio"/> SI <input type="radio"/> No
42. ¿Has tenido problemas con mantener la concentración durante las clases?	<input type="radio"/> SI <input type="radio"/> No
43. ¿Te has molestado con facilidad?	<input type="radio"/> SI <input type="radio"/> No
44. ¿Has tenido muchos amigos que rompen las reglas?	<input type="radio"/> SI <input type="radio"/> No
45. Durante los últimos 6 meses, ¿has sido generalmente una persona buena y honesta?	<input type="radio"/> SI <input type="radio"/> No

Zona 6	
RESPONDA CON RELACION A LOS ÚLTIMOS 6 MESES.	
	SI No
46. ¿Has sido capaz de mantener la concentración cuando estudias?	<input type="radio"/> SI <input type="radio"/> No
47. ¿Generalmente tus padres o encargados han estado enterados de dónde estás y qué haces?	<input type="radio"/> SI <input type="radio"/> No
48. ¿Has herido o hecho daño a los animales?	<input type="radio"/> SI <input type="radio"/> No
49. ¿Has tenido alguna vez la tentación de robar algo?	<input type="radio"/> SI <input type="radio"/> No
50. ¿Muy pocos de tus amigos copian en los exámenes.	<input type="radio"/> SI <input type="radio"/> No
51. ¿Te has preocupado mucho?	<input type="radio"/> SI <input type="radio"/> No
52. ¿Te ha sido difícil pedir ayuda a otros?	<input type="radio"/> SI <input type="radio"/> No
53. ¿Han faltado al colegio mucho de tus amigos?	<input type="radio"/> SI <input type="radio"/> No
54. Algunos jóvenes piensan que es una buena idea usar drogas y otros piensan que es una mala idea usar drogas. ¿Algunos amigos piensan que usar drogas es una mala idea?	<input type="radio"/> SI <input type="radio"/> No
55. ¿Has pensado seriamente en abandonar los estudios?	<input type="radio"/> SI <input type="radio"/> No
56. ¿Te has exaltado con frecuencia al hablar?	<input type="radio"/> SI <input type="radio"/> No
57. Durante los últimos 6 meses, ¿has sentido temor o miedo?	<input type="radio"/> SI <input type="radio"/> No
RESPONDA CON RELACION A LOS ÚLTIMOS 12 MESES.	
58. Durante los últimos 12 meses, ¿has molestado animales?	<input type="radio"/> SI <input type="radio"/> No
59. Durante los últimos 12 meses, ¿has tenido problemas con la policía?	<input type="radio"/> SI <input type="radio"/> No
60. Durante los últimos 12 meses, ir a la iglesia los domingos ha sido una actividad importante para ti.	<input type="radio"/> SI <input type="radio"/> No

Zona 7		Zona 8			
RESPONDA CON RELACION A LOS ÚLTIMOS 6 MESES.		RESPONDA CON RELACION A LOS ÚLTIMOS 6 MESES.			
	SI	NO	SI	NO	
61. Durante los últimos 6 meses, ¿Te ha sido fácil hacer nuevos amigos?	<input type="radio"/>	<input type="radio"/>	76. ¿Has faltado a la escuela más de dos días seguidos al mes?	<input type="radio"/>	<input type="radio"/>
62. ¿Has tenido problemas para concentrarte?	<input type="radio"/>	<input type="radio"/>	77. ¿Has sido suspendido del instituto o colegio?	<input type="radio"/>	<input type="radio"/>
63. ¿Has molesto o usado lenguaje soez (vulgar) con frecuencia?	<input type="radio"/>	<input type="radio"/>	78. ¿Te ha sido fácil hacer amigos en grupos nuevos?	<input type="radio"/>	<input type="radio"/>
64. ¿Te has sentido de mal humor?	<input type="radio"/>	<input type="radio"/>	79. ¿Has tenido dificultad para quitarte algún pensamiento fijo?	<input type="radio"/>	<input type="radio"/>
65. Algunos jóvenes siempre le dicen a sus padres a donde van cuando salen de la casa y otros jóvenes a veces no les cuentan a sus padres a donde van. Durante los últimos 6 meses, cuando has salido de la casa, ¿siempre le has dicho a tus padres a donde vas?	<input type="radio"/>	<input type="radio"/>	80. A veces los jóvenes regresan a casa después del colegio y no encuentran a nadie. Durante los últimos 6 meses, cuando has regresado a casa después del instituto, colegio o trabajo, ¿han estado en casa tu papá, tu mamá o algún otro adulto?	<input type="radio"/>	<input type="radio"/>
66. ¿Has sido una persona solitaria?	<input type="radio"/>	<input type="radio"/>	81. ¿Has molestado a otros con frecuencia?	<input type="radio"/>	<input type="radio"/>
67. ¿Usualmente has hablado con tus padres sobre lo que vas a hacer el día siguiente?	<input type="radio"/>	<input type="radio"/>	82. ¿Te has molestado cuando personas te han hecho esperar?	<input type="radio"/>	<input type="radio"/>
68. A veces, los padres les dicen a sus hijos que no se junten con personas que se meten en problemas. Durante los últimos 6 meses, ¿te han dicho tus padres que no te juntes con personas que puedan meterte en problemas?	<input type="radio"/>	<input type="radio"/>	83. Algunos jóvenes piensan que está bien hacer trampa para ganar en un juego y otros piensan que no es correcto hacer trampa. Durante los últimos 6 meses, ¿has tenido algún amigo cercano que piense que no es correcto hacer trampa?	<input type="radio"/>	<input type="radio"/>
69. ¿Existen reglas claras en tu casa acerca de lo que puedes o no puedes hacer?	<input type="radio"/>	<input type="radio"/>	84. ¿Has sido terco?	<input type="radio"/>	<input type="radio"/>
70. Durante los últimos 6 meses, raramente he sentido miedo.	<input type="radio"/>	<input type="radio"/>	85. ¿Te has sentido rechazado en actividades grupales fuera o dentro del instituto o colegio?	<input type="radio"/>	<input type="radio"/>
71. ¿Han sido tus amigos rechazados por tus padres o enorgados?	<input type="radio"/>	<input type="radio"/>	86. ¿Has desobedecido a los adultos en algunas ocasiones?	<input type="radio"/>	<input type="radio"/>
72. ¿Has sentido que la gente te mira fijamente?	<input type="radio"/>	<input type="radio"/>	87. ¿Han empeorado tus notas este año?	<input type="radio"/>	<input type="radio"/>
73. ¿Muy pocos de tus amigos empiezan peleas.	<input type="radio"/>	<input type="radio"/>	88. Algunos jóvenes prefieren faltar en vez de asistir al instituto o colegio. Durante los últimos 6 meses, ¿has preferido ir a la escuela en vez de faltar?	<input type="radio"/>	<input type="radio"/>
74. ¿Has tenido muchos amigos que roban y hacen trampa?	<input type="radio"/>	<input type="radio"/>	89. Algunos de mis amigos han fumado marijuana.	<input type="radio"/>	<input type="radio"/>
75. ¿Has sido rechazado por tus compañeros u otros jóvenes de tu edad?	<input type="radio"/>	<input type="radio"/>	90. Durante los últimos 6 meses, ¿has dormido bien?	<input type="radio"/>	<input type="radio"/>

Zona 9		
RESPONDA CON RELACION A LOS ÚLTIMOS 6 MESES.		
	Si	No
91. Algunos jóvenes prefieren que se les deje solos. ¿Frecuentemente has querido que te dejen sólo?	<input type="radio"/>	<input type="radio"/>
92. Algunos jóvenes prefieren ir a trabajar en vez de ir a la escuela. Durante los últimos 6 meses, ¿has preferido ir a trabajar que ir a la escuela?	<input type="radio"/>	<input type="radio"/>
93. Durante los últimos 6 meses, <u>ninguna de mis amigos</u> ha fumado tabaco.	<input type="radio"/>	<input type="radio"/>
94. <u>Muchos de mis amigos</u> han fumado tabaco.	<input type="radio"/>	<input type="radio"/>
95. <u>Pocos amigos míos</u> han fumado tabaco.	<input type="radio"/>	<input type="radio"/>
96. A algunos jóvenes les gusta desobedecer las reglas y otros prefieren obedecerlas. Durante los últimos 6 meses, ¿has tenido <u>algún amigo</u> que le guste desobedecer las reglas?	<input type="radio"/>	<input type="radio"/>
97. ¿ <u>Mucho</u> de tus amigos desobedecen las reglas?	<input type="radio"/>	<input type="radio"/>
98. Durante los últimos 6 meses, ¿en general, te has mantenido tranquilo cuando no has podido hacer lo que querías?	<input type="radio"/>	<input type="radio"/>
99. ¿Has tenido <u>algún amigo</u> que robe o haga trampa?	<input type="radio"/>	<input type="radio"/>
100. Algunos jóvenes se sienten contentos cuando piensan en ir a la escuela. Durante los últimos 6 meses, ¿frecuentemente te has sentido contento al pensar en ir al colegio o instituto?	<input type="radio"/>	<input type="radio"/>
101. Durante los últimos 6 meses, ¿has tenido un amigo o una amiga a quien le podrías contar acerca de los problemas y cosas que ocurren en tu vida?	<input type="radio"/>	<input type="radio"/>
102. He pensado en abandonar el instituto o colegio completamente.	<input type="radio"/>	<input type="radio"/>

Zona 10		
RESPONDA CON RELACION A LOS ÚLTIMOS 6 MESES.		
	Si	No
103. ¿Has tenido muchos amigos que han tratado de tener una conducta excelente?	<input type="radio"/>	<input type="radio"/>
104. A veces, las personas tiran o quiebran cosas cuando se enojan. ¿Has tirado o quebrado cosas cuando te enojas?	<input type="radio"/>	<input type="radio"/>
105. A algunas personas les gusta estar con otras personas, mientras que otros prefieren estar solos. Durante los últimos 6 meses, ¿has querido estar con otras personas durante tu tiempo libre?	<input type="radio"/>	<input type="radio"/>
106. Algunas personas generalmente están de mal humor, mientras que otras siempre están de buen humor. Durante los últimos 6 meses, ¿has estado siempre de buen humor?	<input type="radio"/>	<input type="radio"/>
107. ¿Has sido deshonesto aprovechándote de otras personas?	<input type="radio"/>	<input type="radio"/>
108. Durante los últimos 6 meses, ¿has tenido algún amigo que le gusta obedecer las reglas?	<input type="radio"/>	<input type="radio"/>
109. A veces tenemos amigos que tratan de ser muy buenos y de tener una conducta excelente. Durante los últimos 6 meses, ¿has tenido algún amigo que trata de tener una conducta excelente?	<input type="radio"/>	<input type="radio"/>
110. Algunos jóvenes prefieren ir al colegio o instituto que ir a trabajar. ¿Tu preferirías ir al colegio o instituto en lugar de estar trabajando?	<input type="radio"/>	<input type="radio"/>
111. ¿Te has enojado con frecuencia cuando no podías hacer lo que querías?	<input type="radio"/>	<input type="radio"/>
112. A algunas personas les gusta ir al colegio o instituto y otros se sienten descontentos cuando piensan en ir al colegio o instituto. ¿A menudo has estado descontento cuando piensas en ir al colegio o instituto?	<input type="radio"/>	<input type="radio"/>
113. Durante los últimos 6 meses, ¿te has sentido muy alegre?	<input type="radio"/>	<input type="radio"/>
114. ¿Has discutido poco?	<input type="radio"/>	<input type="radio"/>

Zona 11		Si	No
RESPONDA CON RELACION A LOS ÚLTIMOS 6 MESES.			
115.	Rara vez he olvidado lo que aprendí en el instituto o colegio.	<input type="radio"/>	<input type="radio"/>
116.	¿Has tenido amigos a quienes les gusta inhalar pegamento o gasolina?	<input type="radio"/>	<input type="radio"/>
117.	Algunos jóvenes están más contentos cuando piensan en trabajar en vez de ir al colegio o el instituto. ¿Has estado más contento pensando en ir a trabajar?	<input type="radio"/>	<input type="radio"/>
118.	A veces los jóvenes dicen "todo está muy bien conmigo". Durante los últimos 6 meses, ¿has estado todo bien contigo?	<input type="radio"/>	<input type="radio"/>
119.	¿Has tenido muchos amigos que piensan que usar drogas es una mala idea?	<input type="radio"/>	<input type="radio"/>
120.	De vez en cuando, ¿has estado de mal humor?	<input type="radio"/>	<input type="radio"/>
121.	Durante los últimos 6 meses, no he hecho nada peligroso.	<input type="radio"/>	<input type="radio"/>
122.	Algunos jóvenes han comenzado a usar pasta base, cocaína o crack. ¿Has tenido algún amigo que haya usado pasta base, cocaína, o crack?	<input type="radio"/>	<input type="radio"/>
123.	¿Has tenido <u>muchos amigos</u> que han usado pasta base, cocaína, o crack?	<input type="radio"/>	<input type="radio"/>
124.	A veces la gente joven dice "ir al instituto ó colegio es una pérdida de tiempo". Para ti, los últimos 6 meses, ¿han sido una pérdida de tiempo ir al instituto ó colegio?	<input type="radio"/>	<input type="radio"/>
125.	He disfrutado el mirar televisión.	<input type="radio"/>	<input type="radio"/>
126.	¿Has tenido muchos amigos que son malos con otras personas?	<input type="radio"/>	<input type="radio"/>
127.	He tenido excelentes relaciones con mis profesores.	<input type="radio"/>	<input type="radio"/>
128.	Algunas personas prefieren obedecer las reglas mientras que otras prefieren romperlas. Durante los últimos 6 meses, ¿has tenido <u>muchos amigos</u> a quienes les gusta obedecer las reglas?	<input type="radio"/>	<input type="radio"/>

Zona 12		Si	No
RESPONDA CON RELACION A LOS ÚLTIMOS 6 MESES.			
129.	Hay suficientes lugares seguros para caminar o jugar en mi barrio o vecindario.	<input type="radio"/>	<input type="radio"/>
130.	Algunas veces al mes algún joven en mi barrio es agredido.	<input type="radio"/>	<input type="radio"/>
131.	En mi barrio o vecindario, muchos patios o calles tienen botellas quebradas y basura.	<input type="radio"/>	<input type="radio"/>
132.	Yo he visto personas usando o vendiendo drogas en mi barrio o vecindario.	<input type="radio"/>	<input type="radio"/>
133.	Temprano en la mañana o al atardecer, frecuentemente veo personas borrachas en las calles de mi barrio o vecindario.	<input type="radio"/>	<input type="radio"/>
134.	La mayoría de los adultos en mi barrio o vecindario respetan la ley.	<input type="radio"/>	<input type="radio"/>
135.	Me siento seguro cuando camino solo en mi barrio o vecindario.	<input type="radio"/>	<input type="radio"/>
136.	Las personas que viven en mi barrio o vecindario frecuentemente dañan o roban la propiedad de otros.	<input type="radio"/>	<input type="radio"/>
137.	Para muchas personas en mi barrio o vecindario, ir a la Iglesia los domingos es una actividad importante.	<input type="radio"/>	<input type="radio"/>
138.	Durante los últimos 6 meses, con respecto a ti, ¿has participado en actividades de iglesia?	<input type="radio"/>	<input type="radio"/>
139.	Durante los últimos 6 meses, ¿has visto campañas de prevención de drogas por la T.V.?	<input type="radio"/>	<input type="radio"/>
140.	Existe algún programa de prevención de drogas en tu colegio?	<input type="radio"/>	<input type="radio"/>
141.	¿Has participado en actividades de prevención en tu colegio?	<input type="radio"/>	<input type="radio"/>
142.	Durante los últimos 6 meses, ¿has participado en actividades de prevención en tu barrio o vecindario ?	<input type="radio"/>	<input type="radio"/>

Zona 13. LA FAMILIA: RESPONDA CON RELACION A LOS ÚLTIMOS 6 MESES.		Sí	No
143.	¿Durante los últimos 6 meses, ha consumido algún miembro de tu familia (Madre, Padre, Hermano(a)) <u>tabaco</u> ?	<input type="radio"/>	<input type="radio"/>
144.	Ha consumido algún miembro de tu familia (Madre, Padre, Hermano(a)) <u>bebidas alcohólicas</u> ?	<input type="radio"/>	<input type="radio"/>
145.	¿Con frecuencia has tenido discusiones con tus padres que han terminado a gritos?	<input type="radio"/>	<input type="radio"/>
146.	¿Tus padres han discutido mucho entre ellos?	<input type="radio"/>	<input type="radio"/>
147.	¿Ha ingerido alcohol algún miembro de tu familia hasta el punto de causar problemas en la casa, en el trabajo o con los amigos?	<input type="radio"/>	<input type="radio"/>
148.	¿Ha consumido algún miembro de tu familia (Madre, Padre, Hermano(a)) <u>marijuana</u> ?	<input type="radio"/>	<input type="radio"/>
149.	¿Has sentido que corrías peligro en tu casa?	<input type="radio"/>	<input type="radio"/>
150.	¿Ha consumido algún miembro de tu familia (Madre, Padre, Hermano(a)) <u>cocaína</u> ?	<input type="radio"/>	<input type="radio"/>
151.	Para tu familia, ir a la Iglesia los domingos es una actividad importante.	<input type="radio"/>	<input type="radio"/>
152.	Para tu familia, respetar la ley es importante.	<input type="radio"/>	<input type="radio"/>

Zona 14. POR FAVOR CONTESTA LAS SIGUIENTES PREGUNTAS ACERCA DE TU CULTURA Y TU FAMILIA			
153.	¿De qué grupos étnicos es tu familia?		

Zona 14, continuado			
154.	¿Cómo <u>te sientes</u> con relación a tu grupo étnico o cultural?		
	No pienso en eso.....	1	
	No me gusta.....	2	
	Pienso que está bien (OK).....	3	
	Me gusta.....	4	
	¡Me gusta mucho!.....	5	
155.	¿Te gustaría <u>aprender más</u> acerca de la historia y de las costumbres de tu grupo étnico o cultural?		
	No, no quiero aprender más.....	1	
	No estoy seguro si quiero aprender más.....	2	
	Sí, me gustaría aprender un poquito más.....	3	
	Sí, me gustaría aprender más.....	4	
156.	¿Qué piensas acerca de pertenecer a un grupo étnico o cultural?		
	No pienso en eso.....	1	
	No me siento orgulloso - Lo odio.....	2	
	Me siento un poco orgulloso - Esta bien.....	3	
	Estoy orgulloso - Me gusta.....	4	
	Estoy muy orgulloso - ¡Me gusta muchísimo!.....	5	
157.	¿Cuál es tu religión? [Católico, Cristiano o Protestante (favor indicar cuál), Otro (favor indicar cuál), Ninguna]		

158.	¿Cuántos años fue a la escuela o colegio tu papá (o la persona que es como tu papá)?		

159.	¿Cuántos años fue a la escuela o colegio tu mamá (o la persona que es como tu mamá)?		

160.	¿Qué tipo de trabajo tiene tu papá (o la persona que es como tu papá)?		

161.	¿Qué tipo de trabajo tiene tu mamá (o la persona que es como tu mamá)?		

Zona 15	
RESPONDA CON RELACION A TODA SU VIDA.	
162. Las primeras preguntas son sobre la 'chicha', que es una bebida que contiene alcohol. Si usted no sabe sobre la chicha o si usted nunca ha tenido una oportunidad de consumir chicha, escriba el número '99' como su respuesta a estas preguntas. ¿A qué edad tuvo su primera oportunidad de consumir chicha?	EDAD: <input type="text"/> <input type="text"/>
163. ¿A qué edad probó chicha por primera vez?	EDAD: <input type="text"/> <input type="text"/>
164. Acerca del tabaco. ¿A qué edad tuvo su primera oportunidad de fumar tabaco?	EDAD: <input type="text"/> <input type="text"/>
165. ¿A qué edad probó tabaco por primera vez?	EDAD: <input type="text"/> <input type="text"/>
166. ¿Qué tanto piensa que la gente se arriesga a dañarse (físicamente o de otras maneras), si ellos fuman veinte o más cigarrillos de tabaco por día? MARQUE SOLO UNA OPCION.	
1. Ningún riesgo 3. Riesgo moderado 5. No sé	
2. Riesgo leve 4. Gran riesgo	<input type="text"/>
167. Para Ud. y sus amigos, ¿Qué tan fácil o difícil es conseguir tabaco? MARQUE SOLO UNA OPCION.	
1. Muy fácil 2. Fácil 3. Difícil 4. Muy difícil	<input type="text"/>
168. Acerca de las bebidas alcohólicas. ¿A qué edad tuvo su primera oportunidad de consumir bebidas alcohólicas?	EDAD: <input type="text"/> <input type="text"/>
169. ¿A qué edad probó alcohol por primera vez?	EDAD: <input type="text"/> <input type="text"/>
170. ¿Qué tanto piensa que la gente se arriesga a dañarse (físicamente o de otras maneras), si ellos consumen una o dos bebidas alcohólicas? MARQUE SOLO UNA OPCION.	
1. Ningún riesgo 3. Riesgo moderado 5. No sé	
2. Riesgo leve 4. Gran riesgo	<input type="text"/>
171. Para Ud. y sus amigos, ¿Qué tan fácil o difícil es conseguir alcohol? MARQUE SOLO UNA OPCION.	
1. Muy fácil 2. Fácil 3. Difícil 4. Muy difícil	<input type="text"/>
172. Acerca del crack o pasta base. ¿A qué edad tuvo su primera oportunidad de consumir crack o pasta base?	EDAD: <input type="text"/> <input type="text"/>
173. ¿A qué edad probó crack o pasta base por primera vez?	EDAD: <input type="text"/> <input type="text"/>
174. ¿Qué tanto piensa que la gente se arriesga a dañarse (físicamente o de otras maneras), si ellos consumen crack o pasta base una vez o dos veces? MARQUE SOLO UNA OPCION.	
1. Ningún riesgo 3. Riesgo moderado 5. No sé	
2. Riesgo leve 4. Gran riesgo	<input type="text"/>

Zona 16	
RESPONDA CON RELACION A TODA SU VIDA.	
175. Para Ud. y sus amigos, ¿Qué tan fácil o difícil es conseguir crack o pasta base? MARQUE SOLO UNA OPCION.	
1. Muy fácil 2. Fácil 3. Difícil 4. Muy difícil	<input type="text"/>
176. Acerca de las drogas Ecstasy, XTC, o MDMA. ¿A qué edad tuvo su primera oportunidad de consumir las drogas Ecstasy, XTC, o MDMA?	EDAD: <input type="text"/> <input type="text"/>
177. ¿A qué edad probó las drogas Ecstasy, XTC, o MDMA?	EDAD: <input type="text"/> <input type="text"/>
178. Para Ud. y sus amigos, ¿Qué tan fácil o difícil es conseguir las drogas Ecstasy, XTC, o MDMA? MARQUE SOLO UNA OPCION.	
1. Muy fácil 2. Fácil 3. Difícil 4. Muy difícil	<input type="text"/>
179. Acerca del pegamento del zapatero, gasolina, éter, u otras sustancias inhalantes. ¿A qué edad tuvo su primera oportunidad de consumir inhalantes de este tipo?	EDAD: <input type="text"/> <input type="text"/>
180. ¿A qué edad probó inhalantes por primera vez?	EDAD: <input type="text"/> <input type="text"/>
181. ¿Qué tanto piensa que la gente se arriesga a dañarse (físicamente o de otras maneras), si ellos consumen inhalantes de este tipo una vez o dos veces? MARQUE SOLO UNA OPCION.	
1. Ningún riesgo 3. Riesgo moderado 5. No sé	
2. Riesgo leve 4. Gran riesgo	<input type="text"/>
182. ¿Qué tanto piensa que la gente se arriesga a dañarse (físicamente o de otras maneras), si ellos consumen inhalantes de este tipo regularmente? MARQUE SOLO UNA OPCION.	
1. Ningún riesgo 3. Riesgo moderado 5. No sé	
2. Riesgo leve 4. Gran riesgo	<input type="text"/>
183. Para Ud. y sus amigos, ¿Qué tan fácil o difícil es conseguir inhalantes? MARQUE SOLO UNA OPCION.	
1. Muy fácil 2. Fácil 3. Difícil 4. Muy difícil	<input type="text"/>
184. Acerca de la droga Cadrina ¿A qué edad tuvo su primera oportunidad de consumir la droga Cadrina?	EDAD: <input type="text"/> <input type="text"/>
185. ¿A qué edad probó la droga Cadrina por primera vez?	EDAD: <input type="text"/> <input type="text"/>
186. Acerca de las drogas opiáceas como heroína o morfina. ¿A qué edad tuvo su primera oportunidad de consumir 'por razones no-médicas' las drogas opiáceas como heroína o morfina?	EDAD: <input type="text"/> <input type="text"/>
187. ¿A qué edad probó por primera vez y 'por razones no-médicas' drogas opiáceas como heroína o morfina?	EDAD: <input type="text"/> <input type="text"/>

Zona 17	
RESPONDA CON RELACION A TODA SU VIDA	
188. Acerca de las anfetaminas, metanfetamina, de la droga 'ice', u otras drogas estimulantes. ¿ A qué edad probó por primera vez y 'por razones no-médicas' estimulantes?	EDAD: <input type="text"/> <input type="text"/>
189. Acerca el diazepam, alprazolam, chicala, u otras pastillas tranquilizantes o sedantes. ¿ A qué edad probó por primera vez las pastillas tranquilizantes o sedantes 'por razones no-médicas'?	EDAD: <input type="text"/> <input type="text"/>
190. Acerca del marihuana. ¿ A qué edad tuvo su primera oportunidad de fumar marihuana?	EDAD: <input type="text"/> <input type="text"/>
191. ¿A qué edad probó marihuana por primera vez?	EDAD: <input type="text"/> <input type="text"/>
192. ¿Qué tanto piensa que la gente se arriesga a dañarse (físicamente o de otras maneras), si ellos fuman marihuana regularmente? MARQUE SOLO UNA OPCION.	
1. Ningún riesgo 3. Riesgo moderado 5. No sé 2. Riesgo leve 4. Gran riesgo	<input type="text"/>
193. Para Ud. y sus amigos, ¿Qué tan fácil o difícil es conseguir marihuana? MARQUE SOLO UNA OPCION.	
1. Muy fácil 2. Fácil 3. Difícil 4. Muy difícil	<input type="text"/>

Zona 18. Las siguientes preguntas son sobre tu vida durante el último año.

En el último año, CUANTAS VECES HAS HECHO LAS SIGUIENTES ACTIVIDADES?

Para empezar, ¿Cuan seguido ha visto la televisión? Dirías nunca en mi vida, ni una sola vez en el año, una a tres veces este año, mas o menos una vez cada tres meses, una vez al mes, de dos a tres veces cada mes, de una a tres veces cada semana, como una vez al día, o mas de una vez al día?
MARQUE SOLO UNA OPCION POR PREGUNTA.

	NI UNA SOLA VEZ ESTE AÑO	UNA A TRES VECES ESTE AÑO	MAS O MENOS UNA VEZ CADA TRES MESES	UNA VEZ AL MES	DE DOS A TRES VECES CADA MES	DE UNA A TRES VECES CADA SEMANA	COMO UNA VEZ AL DIA	MAS DE UNA VEZ AL DIA
194. Ver televisión	A	B	C	D	E	F	G	H
195. Jugar fútbol, béisbol, baloncesto u otro juego de equipo	A	B	C	D	E	F	G	H
196. Escuchar la radio o música	A	B	C	D	E	F	G	H
197. Tocar un instrumento musical	A	B	C	D	E	F	G	H
198. Trabajar para ganar dinero	A	B	C	D	E	F	G	H
199. Ir al bar ó cantina	A	B	C	D	E	F	G	H
200. Ir de citas o cortejar	A	B	C	D	E	F	G	H
201. Fumar tabaco	A	B	C	D	E	F	G	H
202. Pasar el tiempo con mi familia	A	B	C	D	E	F	G	H
203. Apostar por dinero	A	B	C	D	E	F	G	H
204. Rezar, leer la Biblia o participar en actividades religiosas	A	B	C	D	E	F	G	H
205. Cocinar	A	B	C	D	E	F	G	H
206. Fumar marijuana	A	B	C	D	E	F	G	H
207. Practicar deportes competitivos que implican mucho ejercicio como correr larga distancias o correr a campo traviesa	A	B	C	D	E	F	G	H

Zona 19. En el último año, CUANTAS VECES HAS HECHO LAS SIGUIENTES ACTIVIDADES? (MARQUE SOLO UNA OPCIÓN POR PREGUNTA.)

0		NI UNA SOLA VEZ ESTE AÑO	UNA A TRES VECES ESTE AÑO	MÁS O MENOS UNA VEZ CADA TRES MESES	UNA VEZ AL MES	DE DOS A TRES VECES CADA MES	DE UNA A TRES VECES CADA SEMANA	COMO UNA VEZ AL DÍA	MÁS DE UNA VEZ AL DÍA
208.	Ir a actividades de la iglesia como clases o actividades sociales o ferias	A	B	C	D	E	F	G	H
209.	Cuidar niños	A	B	C	D	E	F	G	H
210.	Ir a encuentros o cruzadas religiosos	A	B	C	D	E	F	G	H
211.	Ir a conciertos de música rock	A	B	C	D	E	F	G	H
212.	Consumir alcohol, cerveza, guano, etc.	A	B	C	D	E	F	G	H
213.	Leer el periódico	A	B	C	D	E	F	G	H
214.	Estudiar para los exámenes	A	B	C	D	E	F	G	H
215.	Salir a bailar	A	B	C	D	E	F	G	H
216.	Consumir crack o pasta base	A	B	C	D	E	F	G	H
217.	Ir al dentista	A	B	C	D	E	F	G	H
218.	Consumir cocaína	A	B	C	D	E	F	G	H
219.	Intentar a bajar de peso	A	B	C	D	E	F	G	H
220.	Dormir por lo menos ocho horas	A	B	C	D	E	F	G	H
221.	Cepillarse los dientes	A	B	C	D	E	F	G	H
222.	Consumir heroína	A	B	C	D	E	F	G	H
223.	Ir a encuentros o cruzadas religiosos	A	B	C	D	E	F	G	H
224.	Trotar, caminar o hacer ejercicios ligeros	A	B	C	D	E	F	G	*H

MUCHAS GRACIAS POR TUS RESPUESTAS!

INFORMACION SOBRE COMO CONTACTARLE EN EL FUTURO

INFORMACIÓN QUE NECESITAMOS SI QUISIERA PARTICIPAR EN LA RIFA PARA EL REGALO DE CUMPLEAÑOS Y PARA AYUDARNOS CONTACTARLES EN EL FUTURO PARA VER COMO SIGUEN.

Por favor escriban con letras de molde para que sea mas fácil de leer

1.	Su nombre completo: _____											
2.	Numero de teléfono, si tienen _____ o de otra quien sabe como contactarles indica su relación con esa persona :											
3.	Dirección de correo de su casa:											
4.	Dirección de su casa o como llegar a su casa si es diferente que su dirección correo:											
5.	Favor de darnos los nombres de algunas personas como vecinos, amigos o parientes quienes supieran como contarles aunque se mudara de su casa.	<table border="0"> <tr> <td data-bbox="836 1014 983 1117">Vecinos:</td> <td data-bbox="983 1014 1434 1117">_____</td> </tr> <tr> <td data-bbox="836 1117 983 1220">Amigos:</td> <td data-bbox="983 1117 1434 1220">_____</td> </tr> <tr> <td data-bbox="836 1220 983 1323">Parientes:</td> <td data-bbox="983 1220 1434 1323">_____</td> </tr> <tr> <td data-bbox="836 1323 983 1426">Lideres de la iglesia</td> <td data-bbox="983 1323 1434 1426">_____</td> </tr> <tr> <td data-bbox="836 1426 983 1503">Otros</td> <td data-bbox="983 1426 1434 1503">_____</td> </tr> </table>	Vecinos:	_____	Amigos:	_____	Parientes:	_____	Lideres de la iglesia	_____	Otros	_____
Vecinos:	_____											
Amigos:	_____											
Parientes:	_____											
Lideres de la iglesia	_____											
Otros	_____											
6.	Nombres de sus padres y/o encargados	_____ _____										
7.	Su Fecha de Nacimiento	Día/Mes/Año: _____										

Bibliografía

- [1] Oficina de las Naciones Unidas contra la Droga y el Delito. Informe mundial sobre las drogas, 2013.
- [2] Alfonso Palmer Juan J. Montaña y Amador Calafat. Predicción del consumo de éxtasis a partir de redes neuronales artificiales. *Original*, (12), 2000.
- [3] Massimo Buscema. Squashing theory: A prediction approach for drug behavior. *Drugs & Society*, (8), 1995.
- [4] José Theodoro Corrêa de Carvalho. Historia de las drogas y de la guerra de su difusión, 2007.
- [5] Institute of Medicine. *Dispelling the Myths About Addiction: Strategies to Increase Understanding and Strengthen Research*. The National Academies Press, Washington, DC, 1997.
- [6] World Health Organization. Lexicon of alcohol and drug terms, 1994.
- [7] Organización de las Naciones Unidas. Convención única de 1961 sobre estupefacientes, enmendada por el protocolo de 1972 de modificación de la convención única de 1961 sobre estupefacientes, 1972.
- [8] Organización de las Naciones Unidas. Convenio sobre sustancias sicotrópicas de 1971, 1971.
- [9] Aram Barra y Rubén Diazconti. Las diferencias entre el uso, el abuso y la dependencia a las drogas.

- [10] Oficina de las Naciones Unidas contra la Droga y el Delito. La prevención en manos de las y los jóvenes.
- [11] José Félix Martínez López Antonio Martín Páez, Margarita Valero Sánchez y Miguel Salguero Piedras. Síndrome de abstinencia. Technical report, HOSPITAL REGIONAL CARLOS HAYA.
- [12] Elisardo Becoña Iglesias. *Bases científicas de la prevención de las drogodependencias*. Delegación del Gobierno para el Plan Nacional sobre Drogas, Madrid, 2002.
- [13] Instituto Nacional sobre el Abuso de Drogas. La ciencia de la adicción: Las drogas, el cerebro y el comportamiento. *NIH*, (15-5605), 2014.
- [14] Ramón Estruch. Efectos del alcohol en la fisiología humana. *ADICCIONES*, (14), 2002.
- [15] Alvaro Urbano Marquez, Ramon Estruch, Francisco Navarro Lopez, Jose Maria Grau, Lluís Mont, and Emanuel Rubin. The effects of alcoholism on skeletal and cardiac muscle. *NEJM*, (320), 1989.
- [16] María Teresa Aguilera, Alejandro de la Sierra, Antonio Coca, Ramón Estruch, Joaquín Fernández Solá, and Alvaro Urbano Márquez. Effect of alcohol abstinence on blood pressure:assessment by 24-hour ambulatory blood pressure monitoring. *Hypertension*, (33), 1999.
- [17] Ramón Estruch. Efectos cardiovasculares del alcohol. *Med Clin*, (105), 1995.
- [18] U.S. Department of Justice. National drug threat assessment 2010, 2010.
- [19] Paul J. Goldstein. The drugs/violence nexus: A tripartite conceptual framework. *Journal of Drug Issues*, (39), 1985.
- [20] Menachem Amir. Patterns in forcible rape. *University of Chicago Press*, 1971.

- [21] Richard Rada. Alcoholism and forcible rape. *The American Journal of Psychiatry*, (132), 1975.
- [22] Marvin E. Wolfgang. *Patterns in Criminal Homicide*. University of Philadelphia Press, Philadelphia, 1958.
- [23] Bruce D. Johnson. *Taking care of business: the economics of crime by heroin abusers*. Lexington Books, Lexington, 1985.
- [24] James A. Inciardi and Carl D. Chambers. *Drugs and the criminal justice system*. Sage Publications, Beverly Hills, 1974.
- [25] Diana Carolina Durán Núñez. La coca y el alto impacto ambiental. *El Espectador*, 2008.
- [26] Raktim Pal, Mallavarapu Megharaj, Paul K. Kirkbride, and Ravi Naidu. Illicit drugs and the environment. *Science of the Total Environment*, 2012.
- [27] Meyer Glantz and Roy Pickens (Eds.). *Vulnerability to Drug Abuse*. American Psychological Association, Washington, DC, 1992.
- [28] Gilbert J. Botvin and Elizabeth M. Botvin. Adolescent tobacco, alcohol, and drug abuse: Prevention strategies, empirical findings, and assessment issues. *Developmental and Behavioral Pediatrics*, (13), 1992.
- [29] Francisco González Calleja Mar García Señorán y Salvador G. González González. Consumo de drogas en la adolescencia. *Psicothema*, (8), 1996.
- [30] Karol L. Kumpfer and Charles Turner. The social ecology model of adolescent substance abuse: Implications for prevention. *The International Journal of Addictions*, (25), 1991.
- [31] García Señorán y María del Mar. *Educación para la salud el consumo de drogas en los adolescentes de Galicia: propuesta educativa para su prevención*.

- PhD thesis, Universidad de Santiago de Compostela, Santiago de Compostela, 1994.
- [32] Lisa J. Crockett and Anne C. Petersen. *Promoting the Health of Adolescents*, chapter 2, pages 13–37. Oxford University Press, New York, 1993.
- [33] Rhonda E. Dento and Charlene M. Kampfe. The relationship between family variables and adolescent substance abuse: A literature review. *Adolescence*, (29), 1994.
- [34] Comisión Nacional contra las Adicciones. Nom-028-ssa2-2009, 2009.
- [35] Aram Barra. La política de drogas en México: prevención, reducción de daños y reasignación presupuestal. *Perspectivas*, (1), 2013.
- [36] Juan J. Montaña Elena Gervilla, Berta Cajal and Alfonso Palmer. Data mining classification techniques: an application to tobacco consumption in teenagers. *Anales de psicología*, (30), 2013.
- [37] Andreas M. Olligschlaeger. Artificial neural networks and crime mapping.
- [38] Cretson L. Dalmadge and Monica Cain. Utilizing neural networks to assess community-level vulnerability to the methamphetamine epidemic. *Issues in Information Systems*, (2), 2008.
- [39] Terence Jackson Russell Beale. *Neural Computing: An Introduction*. Adam Hilger, Bristol, 1990.
- [40] Laurene V. Fausett. *Fundamentals of Neural Networks: Architectures, Algorithms And Applications*. 1994.
- [41] Walter Pitts Warren S. mcculloch. A logical calculus of the ideas immanent in nervous activity. *BULLETIN OF MATHEMATICAL BIOPHYSICS*, (5), 1943.
- [42] Donald O. Hebb. *The organization of behavior*. Wiley, New York, 1948.

- [43] Frank Rosenblatt. *Principles of neurodynamics*. Spartan, New York, 1962.
- [44] Seymour A. Papert Marvin L. Minsky. *Perceptrons: an introduction to computational geometry*. M.I.T. Press, Cambridge, 1969.
- [45] Bertha Higashida. *Ciencias de la salud*. Mc Graw Hill, México, DF., 2008.
- [46] Kevin Gurney. *An introduction to neural networks*. UCL Press, London, 1997.
- [47] Terrence J. Sejnowski and Charles R. Rosenberg. Parallel networks that learn to pronounce english text, 1987.
- [48] NETtalk Test. <https://www.youtube.com/watch?v=gakjlr3gece>, Consultado el 12 de febrero de 2017.
- [49] Geoffrey L. Davis George Davis, Walter E. Lowell. A neural network that predicts psychiatric length of stay. *MD Computing*, (10), 1993.
- [50] Morven Leese Brian S. Everitt, Sabine Landau and Daniel Stahl. *Cluster Analysis*. John Wiley & Sons, United Kingdom, 2011.
- [51] Mehmed Kantardzic. *Data Mining: Concepts, Models, Methods, and Algorithms*. Jhon Wiley & Sons, Hoboken, New Jersey, 2011.
- [52] Alvin C. Rencher. *Methods of Multivariate Analysis*. Jhon Wiley & Sons, United States of America, 2002.
- [53] Leonard Kaufman and Peter J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, Hoboken, New Jersey, 1990.
- [54] Anthony James. Pacardo: Data on drug use and behavior in school-aged children and teenagers in panama, central america, and the dominican republic, 1999-2000.

- [55] Carles M. Cuadras. *Nuevos métodos de análisis multivariante*. CMC Editions, Barcelona, Spain, 2014.