

Instituto Tecnológico y de Estudios Superiores de Occidente

Reconocimiento de validez oficial de estudios de nivel superior según acuerdo secretarial 15018, publicado en el Diario Oficial de la Federación del 29 de noviembre de 1976.

Departamento de Matemáticas y Física
Maestría en Ciencia de Datos



Cribado del trastorno límite de la personalidad por medio de aprendizaje automático

TESIS que para obtener el **GRADO** de
MAESTRA EN CIENCIA DE DATOS

Presenta: **JESSICA ARACELI REYES TREJO**

Asesora **MTRA. ANA LUZ DOMÍNGUEZ CARRANZA**

Asesora **DRA. ROCÍO CARRASCO NAVARRO**

Tlaquepaque, Jalisco. 11 de septiembre de 2021.

AGRADECIMIENTOS

Quiero agradecer a mis asesoras quienes me apoyaron de manera significativa para la realización de esta tesis y su motivación para llevarla al mundo real.

También quiero agradecer a mis padres de quienes siempre he recibido su apoyo incondicional para lograr cada uno de mis sueños.

DEDICATORIA

A mi familia, que siempre ha estado para mí en todos los momentos de mi vida.

Para los que ya no están físicamente, pero que siempre me acompañarán en el corazón.

A mis amigos y a Carolina, por su apoyo, paciencia y compañía.

¡Viva Michoacán!

RESUMEN

La presente tesis aborda el cribado del trastorno límite de la personalidad o TLP desde el aprendizaje automático. Comienza describiendo el trastorno y el problema que representa para los profesionales de la salud mental diagnosticarlo, lo cual tiene repercusiones directas en los pacientes ya que el proceso de diagnóstico puede resultar complicado al presentarse en comorbilidad con otros trastornos motivo de consulta; por esto el objetivo principal es aplicar un modelo de clasificación por medio de aprendizaje automático que pueda ser utilizado por profesionales de la salud mental para identificar a personas con alta probabilidad de tener trastorno límite de la personalidad.

Para poder lograrlo se generó una base de datos con más de 1300 registros; asimismo, se identificaron las características demográficas que tienen las personas con mayor probabilidad de tener TLP y los criterios diagnósticos propuestos por el DSM-V adaptados a partir de otras escalas de cribado para trastornos de personalidad.

Por otro lado, se comparó el porcentaje de la muestra con mayor probabilidad de tener TLP con las tasas oficiales de la OMS y el gobierno de México que indican que entre el 1.5% y el 2% de la población tienen este trastorno, mientras que lo observado en la muestra tomada, que es un 17.7%, está muy por encima de las tasas oficiales mencionadas.

Por último, se implementaron 3 modelos de aprendizaje automático para la clasificación, los cuales son regresión logística, árbol de decisión y red neuronal y se compararon para ver los mejores resultados. El modelo que tuvo mejor sensibilidad a los datos fue el modelo de red neuronal, el cual tuvo una sensibilidad del 97.6%.

Palabras clave: TLP, Borderline, trastorno límite de la personalidad, aprendizaje automático

TABLA DE CONTENIDO

MAESTRÍA EN CIENCIA DE DATOS	1
1. INTRODUCCIÓN	9
1.1. ESTADO DEL ARTE.....	14
1.2. PLANTEAMIENTO DEL PROBLEMA.....	18
1.3. JUSTIFICACIÓN.....	19
1.4. OBJETIVOS.....	21
1.4.1. Objetivo General.....	21
1.4.2. Objetivos Específicos.....	21
1.5. ESTRUCTURA DE LA TESIS.....	22
2. DESARROLLO METODOLÓGICO Y ANÁLISIS DESCRIPTIVO DE LA BASE DE DATOS	23
2.1 ASPECTOS METODOLÓGICOS.....	23
2.1.1. Enfoque y diseño de la investigación.....	23
2.1.2. Hipótesis.....	24
2.1.3. Variables.....	24
2.1.4. Conceptualización de las variables.....	25
2.1.5. Operacionalización de las variables.....	26
2.2 POBLACIÓN.....	27
2.3 SELECCIÓN DE LA MUESTRA Y MUESTREO.....	28
2.4 INSTRUMENTO DE MEDICIÓN.....	29
2.5 RECOLECCIÓN DE DATOS.....	30
2.6 ANÁLISIS DE LOS DATOS.....	31
3. FUNDAMENTOS MATEMÁTICOS	32
3.1. REGRESIÓN LOGÍSTICA.....	32
3.1.1. Regresión logística simple.....	32
3.1.2. Estimación de los coeficientes de regresión.....	34
3.1.3. Regresión logística múltiple.....	35
3.2. ÁRBOL DE DECISIÓN.....	36
3.2.1. Árbol de decisión para clasificación.....	36
3.3. RED NEURONAL.....	38
4. RESULTADOS Y DISCUSIÓN	41
5. CONCLUSIONES Y TRABAJO FUTURO	54
5.1 CONCLUSIONES.....	54
5.2 TRABAJO FUTURO.....	56
6. BIBLIOGRAFÍA	57

LISTA DE FIGURAS

Ilustración 1 - Función logística (función sigmoide)	33
Ilustración 2 - Red neuronal con una capa oculta	38
Ilustración 3 - Resumen base de datos de Pandas	41
Ilustración 4 – Frecuencia de sexo de la muestra tomada	42
Ilustración 5 - Histograma de edades	42
Ilustración 6 - Frecuencia de religión de la muestra tomada	43
Ilustración 7 - Histograma de profesiones.....	43
Ilustración 8 - Histograma de estado civil.....	44
Ilustración 9 - Frecuencia de lugar de residencia	44
Ilustración 10 - Histograma de personas que han sido atendidas por un profesional de la salud mental .	45
Ilustración 11 - Variables significativas.....	47
Ilustración 12 - Histograma de respuesta y	48
Ilustración 13 - Frecuencia de personas con y si probabilidades de tener TLP	48
Ilustración 14 - Métricas para la regresión logística.....	50
Ilustración 15 - Matriz de confusión para la regresión logística.....	50
Ilustración 16 - Métricas para el árbol de decisión	51
Ilustración 17 - Matriz de confusión para el árbol de decisión	52
Ilustración 18 - Métricas para la red neuronal	52
Ilustración 19 - Matriz de confusión para la red neuronal	52
Ilustración 20 - Métricas para medir los modelos de clasificación.....	54

LISTA DE TABLAS

Tabla 1 - Variables independientes	24
Tabla 2 - Variables dependientes	24
Tabla 3 - Comparación de modelos	55

LISTA DE ACRÓNIMOS Y ABREVIATURAS

TLP		Trastorno Límite de la Personalidad
"Borderline Disorder"		Trastorno Límite de la Personalidad
DSM		Manual Diagnóstico y Estadístico de los Trastornos Mentales
APA		Asociación Americana de Psicología
CIE		Clasificación Internacional de Enfermedades
Machine learning		Aprendizaje Automático
Blog		Diario registro del ancho mundo de la red
Podcast		Serie episódica de archivos de audio o video
YouTube		Sitio web dedicado a compartir videos
Bot		Programa informático que efectúa automáticamente tareas reiterativas mediante Internet

1. INTRODUCCIÓN

De acuerdo con la Asociación Americana de Psicología (APA) en el Manual Diagnóstico y Estadístico de los Trastornos Mentales (DSM-IV & DSM-5), el trastorno límite de personalidad (TLP) o “borderline”, “es un trastorno de la personalidad caracterizado por un patrón de larga duración de inestabilidad en el estado de ánimo, relaciones interpersonales y autoimagen que es lo suficientemente grave como para causar angustia extrema o interferir con el funcionamiento social y ocupacional. Entre las manifestaciones de este trastorno se encuentran (a) comportamiento autodestructivo, como autolesiones, comer en exceso o consumo de sustancias; (b) relaciones intensas pero inestables; (c) arrebatos de temperamento incontrolables; (d) incertidumbre sobre la propia imagen, género, metas y lealtades; (e) cambios de humor; (f) comportamiento contraproducente, como peleas, gestos suicidas o automutilación; y (g) sentimientos crónicos de vacío y aburrimiento.” (American Psychological Association, 2020)

El término “borderline” fue utilizado por primera vez en 1884 por el psiquiatra C. Hughes, y posteriormente por el psicoanalista Adolph Stern en 1938 para caracterizar afecciones psiquiátricas que superaban la neurosis, pero que no alcanzaban la psicosis (muchos autores califican los síntomas del TLP como pseudopsicóticos). (Fernández-Guerrero, 2017)

Aunque hoy en día existen criterios más claros para el diagnóstico del TLP, no basta con contestar una serie de preguntas; para poder diagnosticar de manera adecuada el TLP es necesario que se lleve a cabo una entrevista detallada por parte del profesional de la salud mental, así como una evaluación psicológica la cual puede incluir completar cuestionarios, hacer una revisión de la historia clínica y un análisis de los síntomas que el paciente está presentando. Todo esto debido a la complejidad que tiene el diagnóstico del TLP pues este

comparte muchas de las características que se pueden apreciar en alguien que tiene trastorno bipolar y pueden llegar a presentarse casos mal diagnosticados.

Además, debido al entorno cultural en el que estamos, donde la salud mental no tiene la importancia que debería, las personas suelen normalizar los rasgos que pueden indicar que alguien tiene TLP y atribuirlos a su mal carácter, o incluso a características del género de la persona; esto último debido a las costumbres y creencias del entorno cultural donde la persona se desenvuelve, ya que hay rasgos que pueden ser vistos como aceptables en un sexo y no en el otro; de esta forma si se trata de una mujer, los rasgos propios del TLP pueden justificarse con lo sensibles o dramáticas que son y de la misma manera, si se tratase de un hombre, estos rasgos pueden justificarse con las características masculinas que estamos acostumbrados a ver, por ejemplo, la agresividad.

Por otro lado, a pesar de que el término “borderline” tiene más de 100 años de ser utilizado este ha sido un tema de mucha controversia entre los profesionales de la salud. Aunque el DSM III incluyó el “borderline” en 1980, la organización mundial de la salud incluyó el término como tal hasta la Clasificación Internacional de Enfermedades (CIE-11), lanzada en el 2018; en la CIE-10, publicada en 1992, estaba incluido como trastorno de inestabilidad emocional de la personalidad.

Estas discrepancias también se vieron reflejadas entre la comunidad de profesionales de la salud mental. En un estudio realizado en Buenos Aires, Argentina en 1989 se entrevistaron a más de 400 profesionales de la salud mental (psicólogos, psiquiatras y psicoanalistas) sobre la aceptación, rechazo y sistemas empleados para formular el diagnóstico. En este estudio se encontró que el 77% de los entrevistados estaba de acuerdo en incluir el TLP como categoría psicopatológica, el 18% lo rechazó y el 5% se abstuvo de dar una posición. (Corsaro, 1992)

Otro ejemplo donde se puede observar las discrepancias con el diagnóstico del TLP es en el estudio, ¿Los pacientes “borderline” son bien diagnosticados? (Corsaro, 1992) donde se menciona que para algunos de los profesionales de la salud mental el diagnóstico se ha usado como caja de sastre, pues se les dificulta la diferenciación de la entidad nosológica. Para este grupo de profesionales la mayoría de los pacientes son “borderlines”; esto también se puede ver en el texto “De Elisabeth von R. a Lisbeth S.: todos somos borderline” de Yago Franco,

donde se habla de cómo la sociedad, en términos generales, se ha vuelto laxa; los valores, creencias, principios y el juicio moral se van justificando con muchos argumentos lo que lleva a que la formación de la conciencia moral y el control de los impulsos se vuelvan a su vez demasiado laxos.

Todos los criterios clínicos previamente mencionados (trastorno recientemente reconocido de forma oficial, diagnóstico difícil de identificar, falta de importancia por parte de profesionales de la salud y población en general) nos dejan con la sensación de que el porcentaje de casos en la población en general reconocido por la Organización Mundial de la Salud (OMS) del 2% es mucho menor a la que realmente existe. Lo cual se ha dejado ver en diferentes partes del mundo en diferentes épocas:

Estados Unidos. “Un trabajo relativamente preliminar acerca de este trastorno nos sugiere que ocurre en el 2-4% de la población general. La gente con este síndrome utiliza desproporcionalmente los servicios de atención de la salud mental; por ejemplo, algunos estudios sugieren que es un diagnóstico primario para el 10-20% de los pacientes psiquiátricos hospitalarios internos” (American Journal Psychiatry, 1985)

Inglaterra. Aquí, “un cuarto de los psiquiatras de Inglaterra utiliza el término. Tales casos representan alrededor del 15 al 25% del diagnóstico de personalidad dentro de una guerra local: de ellos comúnmente se informaba aquí que sufrían de desórdenes de personalidad explosiva o histérica, o de una enfermedad depresiva”. (Tarnopolsky A, 1987)

España. “Si bien este diagnóstico puede ser usado por los profesionales de acuerdo con su grado de información y formación, estadísticamente no hay registros debido a que oficialmente se clasifica según la OMS”. (Balgañón, 1990)

Francia. “Los psicoanalistas muellemente instalados en sus consultorios no ven más neurosis clásicas. No reciben prácticamente otros casos que estados límites”. (Bergeret, 1990)

Venezuela. “Tengo una larga historia de luchas a nivel hospitalario para lograr que a estos pacientes se les permita la hospitalización, y luego, tan pronto se determina que no se trata de un esquizofrénico o un maniacodepresivo se procede a darlo de alta, como si no se tratara de enfermos sino de simuladores o manipuladores” (Pomonta, 1991)

México. “Alrededor de 1.5 por ciento de la población mexicana presenta Trastorno Límite de la Personalidad (TLP), el cual predispone a quien lo padece a tener conductas de riesgo para la salud como abuso de sustancias, amenazas suicidas recurrentes y automutilación.” (Secretaría de Salud MX, 2018)

Esta última cita es parte de un comunicado hecho por la secretaría de salud del Gobierno de México, por el psiquiatra José Ibarreche Beltrán, quien es jefe del área de Consulta Externa del Hospital Psiquiátrico Fray Bernardino Álvarez. Lo que más llama la atención de su declaración es que menciona que solo el 1.5% de la población mexicana padece TLP, lo cual está por debajo del 2% que asegura la OMS.

Dentro de las declaraciones hechas en otras partes del mundo se puede ver un claro problema tanto para los profesionales de la salud mental como para las personas que tienen TLP y sus familiares. En el caso de los profesionales de la salud mental lograr un diagnóstico acertado del TLP puede llegar a ser un proceso complejo; de acuerdo con la psicóloga clínica Mercedes Ruedas Lizana, lo que hace complejo el diagnóstico es que este trastorno tiene una gran variedad en el grado de intensidad de los síntomas y de tipología clínica (Fundación Pública Andaluza para la Integración Social de Personas con Enfermedad Mental, 2021).

Por otro lado, para las personas con TLP y sus familiares, el proceso actual de diagnóstico de TLP es un proceso muy largo que se concluye después de mucho sufrimiento, lo cual afecta la funcionalidad y adaptación del paciente a su entorno.

Se trata de un proceso de evoluciones parciales, incidiendo de manera tardía en el abordaje de los pacientes quienes, en buena parte, cansados de no saber qué es lo que sucede con ellos desisten de buscar ayuda, teniendo que ser intervenidos nuevamente hasta que aparece la siguiente crisis, las cuales pueden terminar en suicidio.

Este problema nos lleva a formularnos la siguiente pregunta: ¿Por medio de un modelo de aprendizaje automático, es posible realizar un cribado de alta probabilidad diagnóstica temprana y eficiente del Trastorno Límite de Personalidad?, de ser así no solo se daría un gran paso para reducir el tiempo actual de diagnóstico de los pacientes con TLP, sino que también se podría diagnosticar a una mayor cantidad de personas por especialista de la salud mental. Por otro lado, también se podría responder a la pregunta: ¿el porcentaje de personas

que pueden padecer TLP confirmarán los datos emitidos por el gobierno de México y la OMS del 1.5% y 2% respectivamente?

1.1. Estado del Arte

Poco a poco el TLP ha cobrado más relevancia en la sociedad de nuestro país, en redes sociales se pueden encontrar grupos de ayuda para personas que están diagnosticadas como “borderline” así como para sus familiares o círculo más cercano, también es posible encontrar cuestionarios gratuitos que te indican si es posible que tengas TLP o no y que ofrecen teléfonos de especialistas de la salud para completar un diagnóstico, así como también hay blogs, podcasts, canales de YouTube y varios libros dirigidos tanto a las personas que padecen el trastorno como a sus familiares, parejas y amigos. Sin embargo, es muy difícil encontrar información cuando buscas en fuentes oficiales del gobierno mexicano sobre datos estadísticos, la importancia que tiene el trastorno, como afecta actualmente a la sociedad o cómo y dónde tratarlo. Lo cual indica que toda la ayuda e información viene de la misma sociedad que lo padece o de instituciones privadas. Esto hace que la opción de analizar los datos mediante un modelo de aprendizaje automático sea casi imposible pues no existe una base de datos que contenga la información necesaria para hacerlo.

Esta falta de información obliga a abrir la frontera a otras partes del mundo para saber qué se está haciendo con el tema, así como el tipo de tecnologías que se están aplicando al diagnóstico y tratamiento del TLP. Por ejemplo, en España el centro tecnológico de Cataluña está aplicando el aprendizaje automático para gestionar las crisis emocionales de los pacientes que tienen TLP por medio de una aplicación móvil, la cual permite a los pacientes auto monitorear sus crisis emocionales y les aporta herramientas de apoyo psicológico para afrontarlas. (Eurecat, 2019)

Incluso en la universidad de Yale se habla sobre un nuevo enfoque en la psiquiatría denominado psiquiatría computacional; el cual permite extraer datos a partir de observaciones de larga duración y vincularlos con las teorías matemáticas del conocimiento por medio del aprendizaje automatizado. Para dar un ejemplo de esto se puede ver el trabajo de la investigadora Sarah Fineberg, quien junto a su equipo está investigando el impacto que tiene la psiquiatría computacional en el TLP, demostrando como esta nueva disciplina influye de manera significativa en cómo los profesionales de la salud mental estudian y diagnostican este trastorno. (Emerging Technology From The Arxiv, 2017)

El trabajo de Sarah y su equipo consiste en medir el rechazo social por medio del juego de ordenador "Cyberball", en el juego se tienen a 3 personajes virtuales que se pasan una bola dentro de la pantalla. El sujeto de estudio controla uno de los jugadores y cree que otras personas dirigen a los otros 2, sin embargo, los otros jugadores son controlados por el ordenador. Una característica clave del juego es que sin que el sujeto lo sepa, los investigadores pueden controlar cuántas veces recibe la pelota. "Al variar el porcentaje del tiempo en el que el sujeto recibe la pelota se pueden generar sentimientos de rechazo social", explican Fineberg y sus colaboradores. (Emerging Technology From The Arxiv, 2017)

En el caso más extremo, el sujeto pasa el balón a otro de los jugadores, que luego se dedican a lanzárselo entre ellos sin incluir al humano real. "Esta experiencia provoca tristeza y enfado en tan sólo 6 rondas del juego", describe la investigación. Este fenómeno permite estudiar la diferencia entre los sentimientos que experimentan las personas con y sin el trastorno límite de la personalidad, pues estos últimos viven las emociones con mucha más intensidad. Y, lo que es más interesante, las personas con trastorno límite de la personalidad también se sienten excluidas cuando reciben la pelota a un nivel justo "las emociones negativas se reducen, pero no desaparecen del todo, aunque los sujetos con trastorno límite de personalidad reciban la pelota más veces que cualquier otro jugador", señalan los investigadores. (Emerging Technology From The Arxiv, 2017)

Otra aplicación del aprendizaje automatizado para el tratamiento del TLP es la expuesta por Judit Szalai de la universidad de Eötvös Loránd en Budapest, su propuesta consiste en que el aprendizaje automatizado utilice el procesamiento de lenguaje natural para ayudar a los pacientes a rescribir sus propias narrativas en secuencias más coherentes y significativas, en las que se vean a sí mismos más como agentes y menos como sujetos a control externo, así como a asignar roles más consistentes a los demás. (Szalai, 2020)

Hablando sobre aplicaciones de modelos de aprendizaje automático al trastorno límite de personalidad más relacionadas con el objetivo de esta investigación se pueden encontrar 3 las cuales mencionaré a continuación.

La primera aplicación es sobre un modelo de aprendizaje automatizado basado en firmas para distinguir el trastorno bipolar del trastorno límite de la personalidad. Este estudio hecho

por investigadores de Oxford toma a 130 participantes entre los cuales se encuentran 48 con trastorno bipolar, 31 con TLP y 51 voluntarios sanos, quienes completaron evaluaciones diarias de su estado de ánimo utilizando una aplicación móvil personalizada durante un año. (Imanol Perez, 2018)

Mediante el modelo de aprendizaje se pueden capturar las interrelaciones en evolución entre los diferentes elementos del estado de ánimo y explotar esta información para clasificar el diagnóstico de los participantes y predecir su estado de ánimo posterior. Los 3 grupos de participantes se pudieron distinguir entre sí sobre la base del estado de ánimo autoinformado utilizando la metodología de firma. Dicha metodología logró clasificar al 75% de los participantes en el grupo de diagnóstico correcto, en comparación con el 54% que utilizó enfoques estándar. Las calificaciones posteriores del estado de ánimo se predijeron correctamente con más del 70% de precisión. La predicción del estado de ánimo fue más precisa en voluntarios sanos (89-98%) en comparación con el trastorno bipolar (82-90%) y el trastorno límite de la personalidad (70-78%). (Imanol Perez, 2018)

La segunda aplicación tiene un enfoque de aprendizaje automatizado capaz de identificar predictores de trastorno límite de la personalidad. El investigador Joseph Beeney ha identificado predictores críticos del TLP al final de la adolescencia, utilizando un enfoque de aprendizaje automático. Beeney y sus colegas aprovecharon los datos de un gran conjunto de datos longitudinales prospectivos de más de 2,400 niñas que fueron evaluadas anualmente por varios criterios clínicos, psicosociales y demográficos. Su modelo predictivo óptimo incluía 19 predictores de 128 posibles. Los predictores principales incluían síntomas de ansiedad y estado de ánimo, falta de autocontrol, castigos severos y mal funcionamiento. Estos predictores fueron relativamente estables en tres períodos clave de desarrollo: niñez tardía, adolescencia temprana y adolescencia media, lo que sugiere que los factores de riesgo de TLP no cambian notablemente con el tiempo. (Edwards, 2020)

Por último, la tercera aplicación trata sobre una forma de probar modelos integrales de patología de la personalidad es con el modelo de la Red Bayesiana Temporal (TBN), en el que las relaciones entre múltiples subcomponentes de un modelo pueden especificarse y

examinarse durante un período de tiempo dinámico, lo que permite además modelar procesos de retroalimentación positiva a la utilidad integral del modelo. (Selby, 2021)

En este estudio, se aplicó el modelo de TBN para examinar el modelo de cascada emocional en una muestra de adolescentes y adultos jóvenes que se autolesionan activamente, incluidos aquellos con TLP. El modelo TBN se aplicó a los datos de evaluación ecológica momentánea proporcionados a través de las evaluaciones de los teléfonos inteligentes de los participantes durante un período de 2 semanas. (Selby, 2021)

El análisis de TBN sugirió que el modelo de cascada emocional tiene una utilidad predictiva considerable, demostrando una precisión sustancial en la predicción del diagnóstico de TLP (con estimaciones de precisión de alrededor del 90%) y la predicción momentánea de rumiación, emoción negativa y conductas desreguladas (con estimaciones de precisión consistentemente superiores al 70% y alcanzando hasta el 100%, dependiendo del nivel de especificidad de predicción momentánea). Estos hallazgos brindan apoyo y validez a la noción de que el TLP puede surgir de una interacción dinámica entre cascadas emocionales y conductas desreguladas. (Selby, 2021)

1.2. Planteamiento del problema

Como ya se mencionó anteriormente, el reconocimiento oficial del TLP como trastorno mental es relativamente nuevo y a pesar de que en el DSM-V se tienen los criterios necesarios para poder hacer un diagnóstico adecuado este tarda mucho tiempo, llegando a semanas o incluso meses en realizarse por parte de los especialistas de la salud mental y muchas veces es erróneo debido a que no existen herramientas adicionales a los manuales y a los criterios que tenga cada uno.

Una posible solución a este problema es el desarrollo de una herramienta digital que, por medio de un modelo de aprendizaje automático, permita a los profesionales de la salud mental hacer un cribado inicial a los pacientes para indicar si es posible o no que tengan TLP basándose en las respuestas de un cuestionario inicial.

El aprendizaje automático es el subconjunto de la inteligencia artificial (IA) que se centra en desarrollar sistemas que aprenden, o mejoran el rendimiento, en función de los datos que consumen. (Oracle, 2021)

El objetivo del aprendizaje automatizado es mejorar significativamente las capacidades y contribuciones humanas y no reemplazar a los humanos como se muestra habitualmente. (Oracle, 2021)

El aprendizaje automático se basa en diferentes modelos que dependen del tipo de datos y propósito del problema que se está abordando. Estos modelos pueden ser de regresión o clasificación dependiendo del tipo de resultado que se quiera obtener. Cuando se usa regresión, el resultado es un valor numérico y se utiliza cuando se quiere, por ejemplo, hacer predicciones sobre ventas o tiempo. Por otro lado, cuando se usan modelos de clasificación, el resultado es una clase entre un número limitado de clases.

Como el objetivo de esta tesis es determinar si una persona tiene o no altas probabilidades de tener TLP se aplicarán 3 modelos de clasificación los cuales son: regresión logística, red neuronal y árbol de decisión. Estos modelos se explicarán más detalladamente en la sección de fundamentos matemáticos.

1.3. Justificación

En México existen alrededor de 350,000 psicólogos matriculados, (Cavan, 2018) si se considera que la población del país es de 126 millones de habitantes (INEGI, 2020) se tiene que por cada 360 personas existe un psicólogo. Si se hace este mismo ejercicio con los psiquiatras, en México existen aproximadamente 5000 (Gerard Heinze, 2016), por lo que se puede decir que por cada 25,203 habitantes se tiene un psiquiatra.

A pesar de que no todos los habitantes del país acuden regularmente a terapia se puede ver claramente que hay escases de profesionales de la salud mental, quienes además hacen sus diagnósticos a los pacientes de la forma tradicional, con formularios que son llenados a mano y revisados uno a uno por ellos mismos. Este método, aunque efectivo, es poco eficiente y productivo pues analizar los datos en un papel es una tarea larga que se presta a errores humanos, incluso si los registros se pasaran a una computadora el proceso requiere mucha inversión de tiempo.

Por otro lado, no basta con ver las respuestas que los pacientes contestaron en el cuestionario, además es necesario tener una consulta personal con cada paciente donde se le hacen preguntas más detalladas sobre su padecimiento que ayudan a determinar si este tiene algún trastorno y cual es.

Con base a lo anterior, el objetivo principal de esta tesis es desarrollar un modelo de clasificación capaz de hacer un cribado inicial a pacientes con posible TLP. Este cribado inicial podría ayudar de manera significativa a los profesionales de la salud mental para poder identificar desde la primera cita del paciente si se trata de alguien con altas posibilidades de tener TLP; en caso de ser así, el especialista de la salud mental puede proceder con los siguientes pasos de entrevista y/o pruebas adicionales para confirmar el diagnóstico o descartarlo, ahorrando mucho tiempo y sufrimiento tanto a los pacientes como a su familia y abriendo la posibilidad de comenzar un tratamiento temprano.

Si el modelo tiene éxito se puede convertir en una herramienta para hospitales psiquiátricos y/o consultorios públicos o privados de especialistas de la salud mental; lo cual no sólo ayudaría a psicólogos y psiquiatras, sino que agilizaría por completo el proceso de

ingreso y atención a los pacientes pues, en la mayoría de los hospitales psiquiátricos, un paciente pasa por 3 diagnósticos antes de llegar con el especialista que lo atenderá durante su proceso.

Pensando un poco más a futuro, esta primera herramienta puede adaptarse a los otros 6 trastornos de personalidad que existen actualmente para así considerar todos los escenarios posibles de diagnóstico.

Además, con este trabajo de tesis se podrá observar que otra relación hay entre las personas que padecen TLP, como rango de edad, religión, sexo, estado donde viven, etcétera, lo que abre la posibilidad de describir demográficamente a la población con mayor posibilidad de tener TLP.

1.4. Objetivos

1.4.1. Objetivo General

Desarrollar un modelo de clasificación por medio de aprendizaje automático que pueda ser utilizado por profesionales de la salud mental para que, por medio de un cribado inicial, puedan identificar a personas con alta probabilidad de tener trastorno límite de la personalidad.

1.4.2. Objetivos Específicos

1. Generar una base de datos con al menos 1000 registros para poder realizar el análisis deseado.
2. Identificar qué características demográficas tienen las personas con mayor probabilidad de tener TLP.
3. Conocer los criterios diagnósticos propuestos por el DSM-V adaptados a partir de otras escalas de cribado para trastornos de personalidad más comunes en la muestra.
4. Comparar la tasa de posibles personas que tienen TLP con las tasas oficiales estipuladas por el gobierno de México y la OMS.
5. Implementar 3 modelos de aprendizaje automático para clasificación los cuales son regresión logística, red neuronal y árbol de decisión.

1.5. Estructura de la tesis

En el capítulo 2 describirá la metodología utilizada para lograr los objetivos expuestos anteriormente, así mismo se hará un análisis descriptivo de la base de datos utilizada. Este capítulo incluye la población que conforma la base de datos, el proceso de selección de la muestra, el instrumento de medición utilizado, así como la recolección y el análisis de los datos.

En el capítulo 3 se revisarán los fundamentos matemáticos en los que se basan cada uno de los modelos propuestos.

En el capítulo 4 se analizarán los resultados del modelo elegido y por último, en el capítulo 5, se plasmarán las conclusiones del proyecto así como el trabajo futuro.

2. DESARROLLO METODOLÓGICO Y ANÁLISIS DESCRIPTIVO DE LA BASE DE DATOS

2.1 Aspectos metodológicos

2.1.1. Enfoque y diseño de la investigación

El tipo de enfoque que se le dio a este trabajo es cuantitativo y corresponde a un estudio de caso descriptivo comparativo. Para poder realizar el análisis se construyó una base de datos con la información necesaria para cumplir con los objetivos de la investigación y dar una propuesta para el problema planteado.

La forma de lograrlo fue mediante la aplicación de un cuestionario diseñado para determinar si una persona tiene altas probabilidades de tener TLP, este cuestionario fue diseñado en conjunto con la maestra en psicología Ana Luz Domínguez quien es experta en el tema.

La presente tesis es un estudio descriptivo ya que con base en los resultados del cuestionario se podrá observar la relación entre las preguntas incluidas y la respuesta de la variable dependiente que indica si una persona tiene o no probabilidades de tener TLP.

Este estudio se considera transversal ya que se examina la relación entre la probabilidad de tener TLP y las respuestas a las preguntas incluidas en el cuestionario respondido por la población determinada en un periodo de un mes.

2.1.2. Hipótesis

Para poder responder a la pregunta de investigación se tienen las siguientes hipótesis.

H₁: Un modelo de clasificación de aprendizaje automático es capaz de hacer un cribado inicial confiable, mayor al 90% de sensibilidad, del TLP.

H₀: Un modelo de clasificación de aprendizaje automático no es capaz de hacer un cribado inicial confiable, mayor al 90% de sensibilidad, del TLP

2.1.3. Variables

El cuestionario tiene solamente el propósito de reunir información así que ninguna retroalimentación fue dada a las personas que lo respondieron; todas las preguntas se pueden traducir a variables independientes mostradas a continuación, para mayor detalle el valor asignado a cada variable puede ser encontrado en el anexo 1.

Tabla 1 - Variables independientes

Variable	Descripción
Sexo	Sexo de la persona respondiendo el cuestionario
Edad	Edad de la persona respondiendo el cuestionario (12 a 61 o más)
Religión	Religión de la persona respondiendo el cuestionario
Nivel educativo	Último grado educativo terminado
Estado civil	Estado civil de la persona respondiendo el cuestionario
Lugar de residencia	Estado de la república donde vive la persona respondiendo el cuestionario
a	¿Has sido tratado por un profesional de la salud mental?
X ₁ – X ₁₇	Preguntas diseñadas para determinar si una persona tiene altas probabilidades de tener trastorno límite de la personalidad

Por otro lado, solo hay 2 variables dependientes las cuales se muestran a continuación:

Tabla 2 - Variables dependientes

Variable	Descripción
b	Número de preguntas positivas en las variables X ₁ a X ₁₇
y	Si la persona tiene o no trastorno límite de personalidad en el cribado inicial

2.1.4. Conceptualización de las variables

Las variables independientes consideradas en el cuestionario pueden dividirse en las siguientes 2 categorías:

- Demográficas: las cuales solicitan información básica a las personas que contestaron el cuestionario como sexo, edad, religión, nivel educativo, estado civil y residencia.
- Criterios clínicos para determinar si una persona tiene TLP: Estas preguntas se hicieron con base en los 9 criterios clínicos Manual Estadístico de Trastornos Mentales (DSM – V). Los 9 criterios son:
 1. Esfuerzos desesperados para evitar el desamparo real o imaginario (no incluir el suicida)
 2. Patrón de relaciones interpersonales inestables e intensas que se caracteriza por una alternancia entre los extremos de idealización y de devaluación.
 3. Alteración de la identidad: inestabilidad intensa y persistente de la autoimagen y del sentido del yo.
 4. Impulsividad en dos o más áreas que son potencialmente autolesivas (por ejemplo: gastos, sexo, drogas, conducción temeraria, atracones alimentarios)
 5. Comportamiento, actitud o amenazas recurrentes de suicidio, o comportamientos de automutilación.
 6. Inestabilidad afectiva debida a una reactividad notable del estado de ánimo (por ejemplo: episodios intensos de disforia, irritabilidad o ansiedad que generalmente duran unas horas y, rara vez, más de unos días).
 7. Sensación crónica de vacío.
 8. Enfado inapropiado e intenso, o dificultad para controlar la ira (por ejemplo: exhibición frecuente de genio, enfado constante, peleas físicas recurrentes).
 9. Ideas paranoides transitorias relacionadas con el estrés o síntomas disociativos graves.

2.1.5. Operacionalización de las variables

Las variables independientes se pueden dividir en demográficas y criterios clínicos.

Las variables demográficas están compuestas por las preguntas: 2, 3, 4, 5, 6, 7 y 8.

Por otro lado, las variables que miden los criterios clínicos utilizados se miden con las preguntas: 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 y 25.

Por último, las variables dependientes que indican la cantidad de respuestas positivas a las variables que miden los criterios clínicos y si el paciente tiene o no altas probabilidades de tener TLP están dadas por las preguntas: 26 y 27.

2.2 Población

Para esta investigación se hizo uso de medios digitales con el objetivo de obtener una muestra aleatoria de personas que vivieran en México y fueran mayores de 12 años, por lo que no importaba su escolaridad, sexo, religión o cualquier otra característica demográfica.

El cuestionario fue creado por medio de Google docs ya que permite el acceso a cualquier persona que tenga el enlace para acceder al cuestionario, así como el almacenaje de los resultados y la descarga de los mismos de forma segura y con acceso las 24 horas del día.

2.3 Selección de la muestra y muestreo

El método de muestreo para recolectar los datos fue de modo aleatorio y no probabilístico, ya que el cuestionario utilizado para la recolección de datos fue puesto al alcance de todo tipo de personas, el único requisito era tener un dispositivo con acceso a internet.

Por lo que al concluir el periodo de recolección de datos se obtuvo una muestra de 1373 personas de todo el país, cumpliendo así con el primer objetivo específico de generar una base con al menos 1000 registros.

2.4 Instrumento de medición

El cuestionario incluye algunas preguntas demográficas como sexo, edad, religión, nivel educativo, estado civil, lugar de residencia y si la persona está o ha sido atendido por una profesional de la salud mental, así como las preguntas necesarias para hacer un diagnóstico inicial del trastorno límite de la personalidad. Este cuestionario puede ser encontrado en los anexos de la presente tesis.

El test incluye 17 preguntas necesarias para determinar si una persona tiene altas probabilidades de tener trastorno límite de la personalidad, estas están basadas en los 9 criterios clínicos definidos por la quinta edición del Manual Estadístico de Trastornos Mentales (DSM – V) y para cada uno de los criterios clínicos se crearon 2 preguntas, excepto para el factor que habla acerca de la dificultad para controlar el enojo y la ira intensa e inapropiada ya que había otra pregunta sobre la ira que podía ser aplicada a 2 de los criterios. Las variables asignadas a estas preguntas van de x_1 a x_{17} .

Estas preguntas fueron basadas en el Cuestionario Exploratorio de Personalidad III (CEPER-III) y del Cuestionario Salamanca para el Screening del Trastorno Límite de la Personalidad, que son instrumentos ya estandarizados; además fueron aprobadas por la especialista en el trastorno límite de la personalidad Ana Luz Domínguez Carranza, quien ha trabajado con este trastorno durante los últimos 12 años. Las preguntas fueron diseñadas de manera que las personas que respondieran el cuestionario pudieran hacerlo con una respuesta de “sí” o “no”.

2.5 Recolección de datos

Existen diferentes formas de registrar un nuevo paciente en la base de datos de un hospital psiquiátrico o en un consultorio privado de psicología o psiquiatría. Diferentes formatos se llenan dependiendo de cada establecimiento y casi ninguno de ellos son digitales o están disponibles en alguna plataforma para que el público en general pueda consultarlos; incluso el Instituto Nacional de Estadística y Geografía no tiene base de datos relacionada al trastorno límite de la personalidad en México.

Para resolver este problema, un cuestionario en línea fue diseñado para conseguirla, este cuestionario se hizo con la intención de recolectar los datos de forma anónima de personas que viven en México.

La encuesta fue lanzada el 9 de Febrero del 2021 y cerrada el 9 de Marzo del mismo año por medio de Google docs y se publicó en Facebook y Twitter para darle una mayor difusión.

2.6 Análisis de los datos

Los datos obtenidos en el cuestionario se analizaron de dos formas, la primera fue por medio de estadística descriptiva para identificar los rasgos más significativos de la población en cuanto a las variables demográficas.

Por otro lado, los datos se analizaron por medio de 3 modelos de aprendizaje automático con la finalidad de descubrir si estos eran capaces de hacer un cribado inicial para identificar a las personas con mayor probabilidad de tener Trastorno Límite de la Personalidad.

Estos 3 modelos fueron comparados entre sí para seleccionar el que mejor se adecuaba al tipo de datos que se tenían y poder seleccionar al que tuviera una mayor sensibilidad al momento de clasificar los resultados.

3. FUNDAMENTOS MATEMÁTICOS

La pregunta que responde el cuestionario utilizado para este análisis es si una persona tiene o no altas probabilidades de tener TLP, al ser una variable cualitativa es necesario utilizar modelos de clasificación. En esta sección se verán los tres modelos de aprendizaje automático para clasificación que fueron utilizados. Estos modelos son regresión logística, árbol de decisión y red neuronal.

3.1. Regresión logística

El primer modelo es la regresión logística, este modelo permite estimar la probabilidad de que una persona pueda tener TLP; en este caso, la variable de respuesta puede tomar únicamente dos valores: 1, presencia (con probabilidad p); y 0, ausencia (con probabilidad $1-p$).

3.1.1. Regresión logística simple

Para modelar la relación de la probabilidad de la variable de respuesta Y dado la variable independiente X : $p(X) = P(Y = 1|X)$ y X se debe modelar la probabilidad de X $p(X)$ usando una función que de como resultado valores entre 0 y 1 para todos los valores de X . Aunque hay varias funciones que cumplen con este criterio para la regresión logística se usa la función logística: (Gareth James, 2021)

$$p(X) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (1)$$

La función logística siempre va a producir una curva en forma de S cuyos valores más bajos serán muy cercanos a 0 pero nunca por debajo de 0 y los valores más altos serán muy cercanos a 1 pero nunca por encima de 1. (Gareth James, 2021)

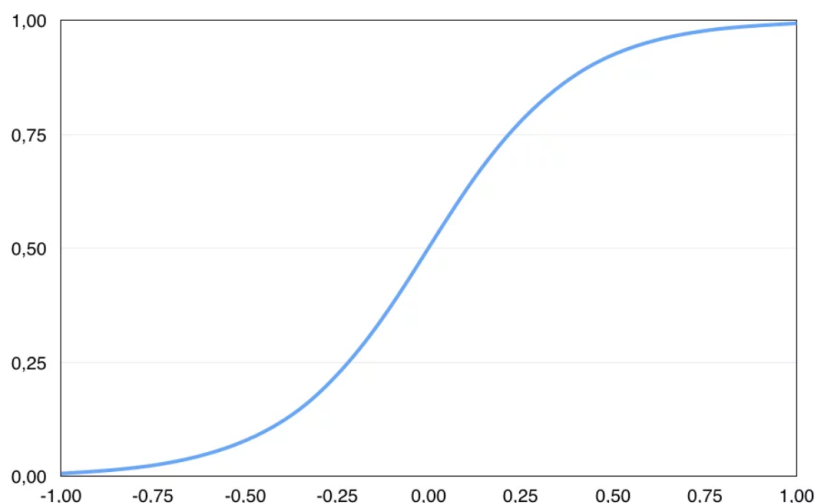


Ilustración 1 - Función logística (función sigmoide)

Si se manipula un poco con la ecuación de arriba se puede obtener que

$$\frac{p(X)}{1-p(X)} = e^{\beta_0 + \beta_1 x} \quad (2)$$

La cantidad $p(X)/(1 - p(X))$ es conocida como odd (posibilidad) y puede tomar cualquier valor entre 1 e infinito (∞); los valores de la posibilidad cercanos a 0 y a ∞ indican probabilidades muy bajas y altas respectivamente. (Gareth James, 2021)

Si se aplica el logaritmo a ambos lados de la ecuación se tiene:

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 x \quad (3)$$

La parte izquierda de la ecuación se llama logit o log odds y este modelo regresión logística tiene un logit que es lineal en X . En la regresión logística incrementar una unidad en X cambia el log odds por β_1 . Equivalentemente, multiplica las odds (posibilidades) por e^{β_1} . Sin embargo, ya que la relación entre $p(X)$ y X no es una línea recta, β_1 no corresponde al cambio en $p(X)$ asociado con el incremento en una unidad en X . La cantidad que $p(X)$

cambia con respecto a una unidad de cambio de X depende del valor actual de X . Pero sin importar el valor de X , si β_1 es positivo entonces el incremento en X estará asociado con el incremento de $p(X)$, y si β_1 es negativo entonces el incremento en X estará asociado con el decremento en $p(X)$. (Gareth James, 2021)

3.1.2. Estimación de los coeficientes de regresión

Los coeficientes β_0 y β_1 son desconocidos y tienen que ser estimados con base en los datos disponibles para entrenamiento. El método más común para ajustar el modelo es el método de máxima verosimilitud ya que tiene mejores propiedades estadísticas. El método consiste en lo siguiente: buscar valores estimados para β_0 y β_1 tales que la probabilidad predicha $\hat{p}(x_i)$ predeterminada para cada individuo usando la función logística, corresponda, tan cercano como sea posible, al estado predeterminado del individuo observado i . (Gareth James, 2021)

En otras palabras, se trata de encontrar β_0 y β_1 de manera que al insertar estas estimaciones en el modelo para $p(X)$, dado en la función logística, se obtenga un número cercano a uno para todos los individuos que incumplieron, y un número cercano a cero para todas las personas que no lo hicieron. Este proceso se puede formalizar utilizando la función de verosimilitud: (Gareth James, 2021)

$$\ell(\beta_0, \beta_1) = \prod_{i: y_i=1} p(x_i) \prod_{i': y_{i'}=0} (1 - p(x_{i'})) \quad (4)$$

Donde ℓ es la función de verosimilitud parametrizada por β_0 y β_1 . Para cada muestra o punto de entrenamiento i se tiene un vector de características x_i y clases observadas y_i . La probabilidad de esta clase es $p(x_i)$, si $y_i = 1$ o $1 - p(x_{i'})$ si $y_i = 0$. Los estimadores β_0 y β_1 son escogidos para maximizar la función de máxima verosimilitud.

3.1.3. Regresión logística múltiple

La regresión logística múltiple se utiliza cuando se tiene un problema con una respuesta binaria que utiliza varios predictores. Con la extensión de la regresión lineal múltiple se puede generalizar:

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p \quad (5)$$

Donde $X = (X_1, \dots, X_p)$ son p predictores. La ecuación de arriba se puede reescribir de la siguiente forma:

$$p(X) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}} \quad (6)$$

Al igual que con la regresión logística simple, se utiliza el método de máxima verosimilitud para estimar $\beta_0, \beta_1, \dots, \beta_p$. (Gareth James, 2021)

3.2. Árbol de decisión

El segundo modelo abordado es el árbol de decisión, estos se pueden aplicar para problemas de regresión y clasificación.

3.2.1. Árbol de decisión para clasificación

El árbol de decisión para clasificación se utiliza cuando la variable de respuesta tiene valores cualitativos. Para un árbol de regresión, la respuesta predicha para una observación está dada por la respuesta promedio de las observaciones de entrenamiento que pertenecen al mismo nodo terminal. (Gareth James, Tree-Based Methods, 2021)

Por otro lado, para un árbol de clasificación, se predice que cada observación pertenece a la clase de observaciones de entrenamiento que ocurren con más frecuencia en la región a la que pertenece. Al interpretar los resultados de un árbol de clasificación, el interés no está solamente en la predicción de clase correspondiente a una región de nodo terminal en particular, sino también en las proporciones de clase entre las observaciones de entrenamiento que caen en esa región. (Gareth James, Tree-Based Methods, 2021)

Al igual que en la configuración de un árbol de regresión, se utiliza la división binaria recursiva para hacer crecer un árbol de clasificación. Sin embargo, en la configuración de clasificación, la suma de cuadrados de los residuos no se puede utilizar como criterio para realizar divisiones binarias. (Gareth James, Tree-Based Methods, 2021)

Una alternativa a la suma de cuadrados de los residuos es la tasa de error de clasificación. Dado que se planea la clasificación para asignar una observación en una región dada a la clase de tasa de error más común de las observaciones de entrenamiento en esa región, la tasa de error de clasificación es la fracción de las observaciones de entrenamiento en esa región que no pertenecen a las clases más comunes: (Gareth James, Tree-Based Methods, 2021)

$$E = 1 - \max_k(\hat{p}_{m,k}) \quad (7)$$

En esta ecuación $\hat{p}_{m,k}$ representa la proporción de las observaciones de entrenamiento en la región m que son de la clase k . Sin embargo, el error de clasificación no es lo suficientemente sensible para el crecimiento de árboles y, en la práctica, son preferibles otras dos medidas. (Gareth James, Tree-Based Methods, 2021)

El índice Gini está definido por:

$$G = \sum_{k=1}^K \hat{p}_{m,k}(1 - \hat{p}_{m,k}) \quad (8)$$

Una medida de la varianza total a través de las K clases. El índice Gini toma un valor pequeño si todas las $\hat{p}_{m,k}$ son cercanas a 0 o 1. Por esta razón el índice Gini es referido como una unidad de medición de la pureza del nodo, un valor pequeño indica que el nodo contiene predominantemente observaciones de una misma clase. (Gareth James, Tree-Based Methods, 2021)

Una alternativa al índice Gini es la entropía la cual está dada por:

$$D = -\sum_{k=1}^K \hat{p}_{m,k} \log \hat{p}_{m,k} \quad (9)$$

Ya que $0 \leq \hat{p}_{m,k} \leq 1$ se tiene que $0 \leq \hat{p}_{m,k} \log \hat{p}_{m,k}$. La entropía va a tomar un valor cercano a 0 si los $\hat{p}_{m,k}$ son todos cercanos a 0 o 1. Por lo tanto, como el índice Gini, la entropía va a tomar un valor pequeño si el nodo m th es puro. (Gareth James, Tree-Based Methods, 2021)

Cuando se construye un árbol de clasificación, el índice Gini o el de entropía son típicamente utilizados para evaluar la calidad de una división en particular, ya que estos son acercamientos más sensitivos a la pureza del nodo que el error de clasificación. Cualquiera de estos 2 acercamientos se puede utilizar cuando se poda el árbol, pero el error de clasificación es preferible si la precisión de la predicción del árbol podado final es el objetivo. (Gareth James, Tree-Based Methods, 2021)

3.3. Red neuronal

Por último, el tercer modelo que se abordará es el modelo de red neuronal, el cuál es un modelo simple del funcionamiento del sistema nervioso. Este modelo emula el modo en que el cerebro humano procesa la información, funciona ejecutando de forma simultánea un número elevado de unidades de procesamiento interconectadas que parecen versiones abstractas de neuronas.

Una red neuronal toma como entrada un vector de p variables $X = (X_1, X_2, \dots, X_p)$ y crea una función no lineal $f(x)$ para predecir la respuesta Y . Lo que diferencia una red neuronal de otros métodos es la estructura del modelo. En la siguiente figura se muestra una red neuronal de avance simple para modelar una respuesta cuantitativa $f(x)$ o Y utilizando $p = 4$ predictores. (Gareth James, Deep Learning, 2021)

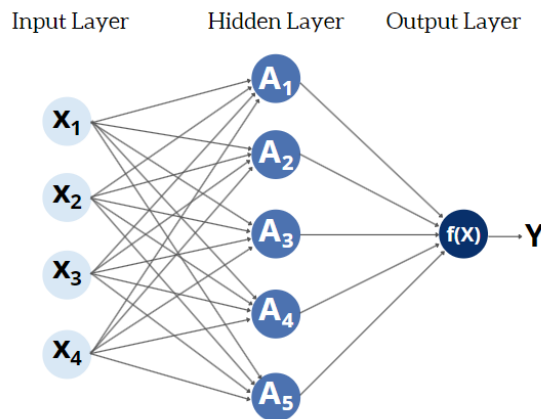


Ilustración 2 - Red neuronal con una capa oculta

La capa oculta calcula activaciones $A_k = h_k(X)$ que son transformaciones no lineales de combinaciones lineales de las entradas X_1, X_2, \dots, X_p . Por tanto, estos A_k no se observan directamente. Las funciones $h_k(\cdot)$ no se fijan de antemano, sino que se aprenden durante el entrenamiento de la red. La capa de salida es un modelo lineal que usa estas activaciones A_k como entradas, dando como resultado una función $f(x)$. (Gareth James, Deep Learning, 2021)

En la terminología de redes neuronales las 4 características X_1, \dots, X_4 forman las unidades de la capa de entrada. Las flechas indican que cada una de las entradas de la capa

de entrada alimentan cada una de las unidades escondidas K , en este caso se tienen 5 unidades K . Por lo que la red neuronal tiene la forma: (Gareth James, Deep Learning, 2021)

$$\begin{aligned} f(X) &= \beta_0 + \sum_{k=1}^K \beta_k h_k(X) \\ &= \beta_0 + \sum_{k=1}^K \beta_k g(w_{k0} + \sum_{j=1}^p w_{kj} X_j) \end{aligned} \quad (10)$$

Este modelo está formado en 2 pasos. Primero las K activaciones $A_k, k = 1, \dots, K$, en la capa oculta se calculan como funciones de las características de entrada X_1, \dots, X_p ,

$$A_k = h_k(X) = g(w_{k0} + \sum_{j=1}^p w_{kj} X_j) \quad (11)$$

Donde $g(z)$ es la función de activación no lineal que se especifica en avance. Se puede considerar cada A_k como una transformación diferente $h_k(X)$ de las características originales. Estas K funciones de activación de la capa oculta después alimentan la capa de salida teniendo como resultado un modelo de regresión lineal es las $K = 5$ activaciones:

$$f(X) = \beta_0 + \sum_{k=1}^K \beta_k A_k \quad (12)$$

Todos los parámetros β_0, \dots, β_k y w_{10}, \dots, w_{Kp} necesitan ser estimados desde los datos. En los primeros casos de la red neuronal, la función sigmoide de activación estaba dada por:

$$g(z) = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}} \quad (13)$$

Esta es la misma función utilizada en la regresión logística para convertir una función lineal en probabilidades con valores entre 0 y 1. La opción preferida en las redes neuronales modernas es la unidad lineal rectificadora que tiene la forma: (Gareth James, Deep Learning, 2021)

$$g(z) = (z)_+ = \begin{cases} 0 & \text{if } z < 0 \\ z & \text{otherwise} \end{cases} \quad (14)$$

La activación de la unidad lineal rectificada puede ser calculada y guardada más eficientemente que la función de activación. Aunque su umbral es cero, debido a que se aplica a una función lineal, el término constante w_{k0} desplazará este punto de inflexión. (Gareth James, Deep Learning, 2021)

En otras palabras, el modelo de la red neuronal simple representado la ilustración 2 deriva 5 nuevas características calculando 5 diferentes combinaciones lineales de X , y después concentra a cada uno por medio de una función de activación $g(\cdot)$ para transformarla. El modelo final es lineal en estas variables derivadas. (Gareth James, Deep Learning, 2021)

El nombre red neuronal se deriva originalmente de pensar en estas unidades ocultas como análogas a las neuronas en el cerebro: los valores de las activaciones $A_k = h_k(X)$ cercanos a uno se activan, mientras que los cercanos a cero son silenciosos (usando la función de activación sigmoidea). (Gareth James, Deep Learning, 2021)

La no linealidad en la función de activación $g(\cdot)$ es esencial, ya que sin esto el modelo $f(x)$ en el modelo de la red neuronal colapsaría a un modelo lineal simple en $X = (X_1, X_2, \dots, X_p)$. Además, una función de activación no lineal permite que el modelo capture efectos de interacción y no linealidades complejas. (Gareth James, Deep Learning, 2021)

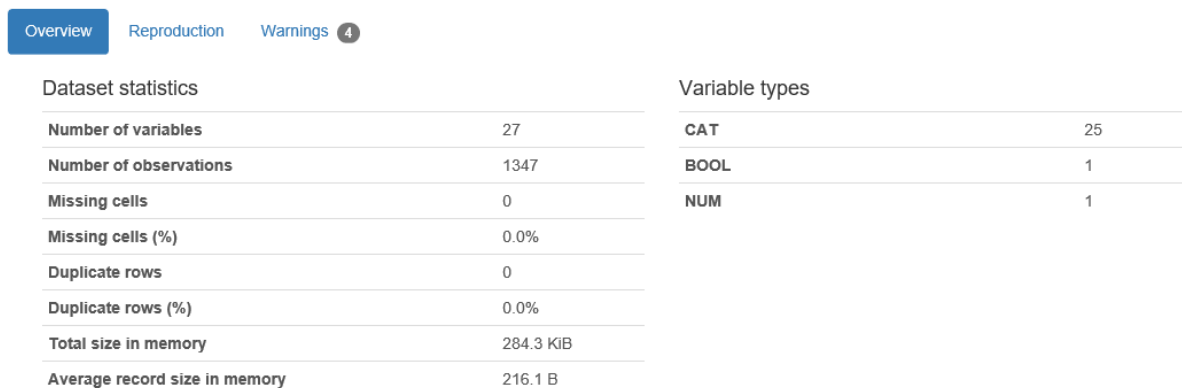
Ajustar una red neuronal requiere estimar los parámetros desconocidos en el modelo de la red neuronal. Para una respuesta cuantitativa, se utiliza típicamente la pérdida de error al cuadrado, de modo que los parámetros se eligen para minimizar:

$$J = \sum_{i=1}^n (y_i - f(x_i))^2 \quad (15)$$

Las redes neuronales modernas normalmente tienen más de una capa oculta y muchas unidades por capa. Para la presente tesis el modelo utilizado solo contiene una capa oculta por lo que no se entrará en detalle en redes neuronales con multicapa.

4. RESULTADOS Y DISCUSIÓN

La base de datos utilizada para el desarrollo de la tesis cuenta con 27 variables, 1347 observaciones y ningún dato faltante, lo que facilita la preparación de los datos, por otro lado, casi todas las variables son categóricas. En la siguiente imagen se puede observar este resumen obtenido con la ayuda de pandas, que es una herramienta de manipulación y análisis de datos de código abierto construida sobre el lenguaje de programación Python. (Pandas, 2021)



Dataset statistics		Variable types	
Number of variables	27	CAT	25
Number of observations	1347	BOOL	1
Missing cells	0	NUM	1
Missing cells (%)	0.0%		
Duplicate rows	0		
Duplicate rows (%)	0.0%		
Total size in memory	284.3 KiB		
Average record size in memory	216.1 B		

Ilustración 3 - Resumen base de datos de Pandas

La primera variable demográfica que se tiene es la del sexo, como se puede observar en la siguiente imagen la mayor parte de respuestas, un 76.8% fueron dadas por mujeres; a pesar de que el cuestionario fue lanzado a grupos donde había un balance de hombres y mujeres estas últimas se caracterizan por ser más abiertas con sus emociones y sentimientos, lo cual puede verse reflejado en la cantidad de respuestas obtenidas por mujeres.

Value	Count	Frequency (%)
Mujer	1035	76.8%
Hombre	304	22.6%
Prefiero no decirlo	7	0.5%
Binario	1	0.1%

Ilustración 4 – Frecuencia de sexo de la muestra tomada

La segunda variable demográfica es la edad de las personas que realizaron el cuestionario, como se mencionó anteriormente el rango de edad disponible en las respuestas era de 12 a 61 años o más; como se puede ver en el siguiente histograma la mayor parte de las respuestas fueron obtenidas de personas entre 20 y 45 años, teniendo la mayor concentración de personas con edades entre 30 y 35 años.

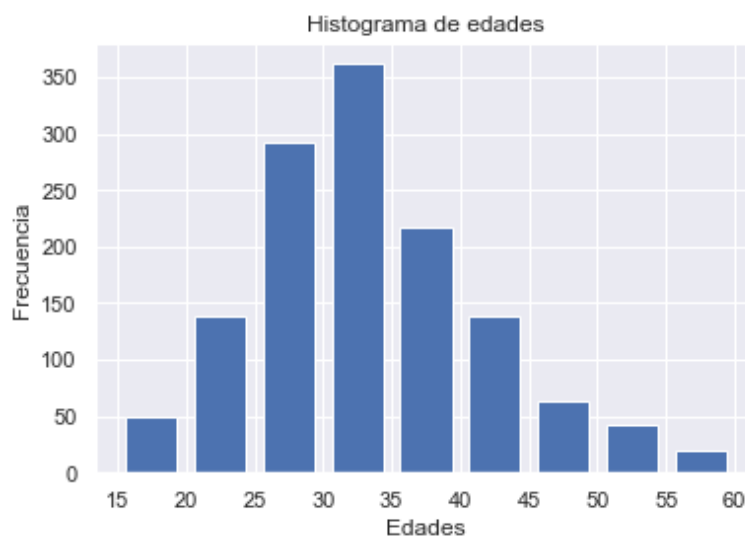


Ilustración 5 - Histograma de edades

La siguiente variable es la de religión, esta variable es la que tiene la mayor cantidad de respuestas diferentes (69 en total), para la cual la mayor parte (67.3%) pertenece a la religión católica, seguida de personas ateas (9.1%) y personas agnósticas (8.5%). En la siguiente imagen se puede observar claramente como la mayoría de las personas se puede dividir en las primeras 5 opciones (Catolicismo, Ateísmo, Agnosticismo, Cristianismo o Ninguna, la cual podría traducirse en Ateísmo), al ser una pregunta abierta se tienen respuestas muy diferentes que podrían ser reagrupadas en una menor cantidad de variables, sin embargo,

este dato no es relevante para el estudio y es solamente una referencia. Debido a la gran variedad que contiene se decidió que esta variable fuera excluida de las conclusiones.

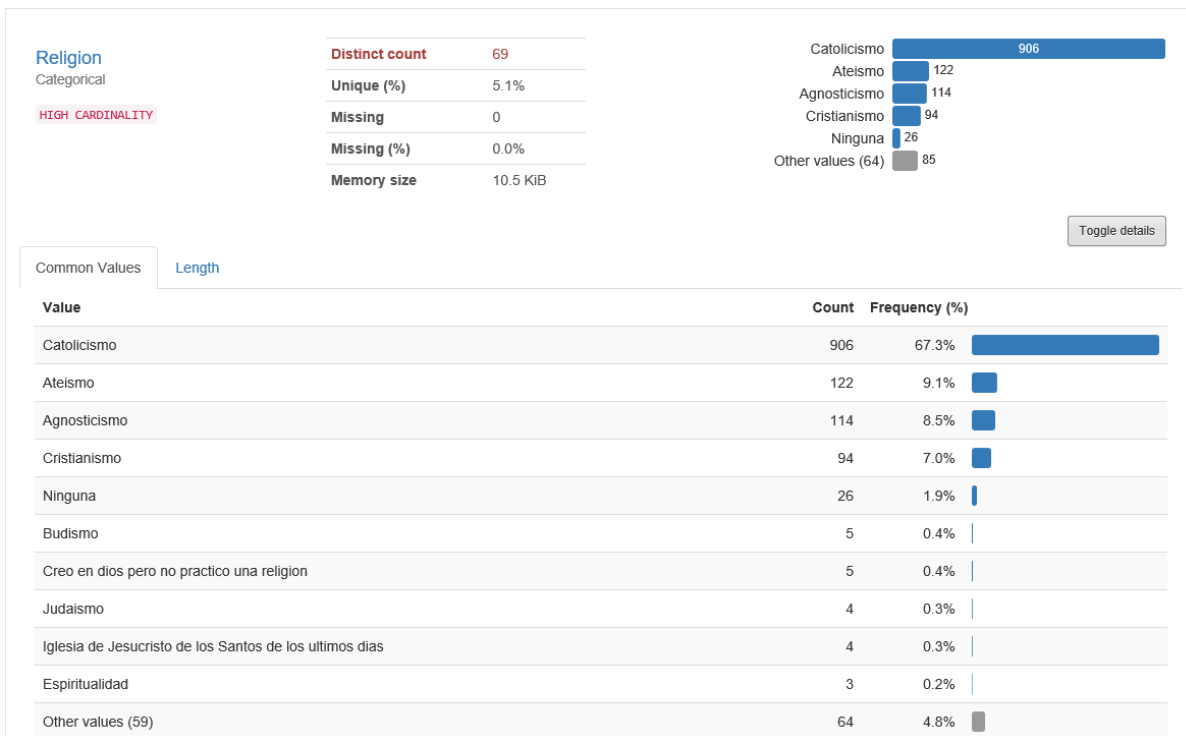


Ilustración 6 - Frecuencia de religión de la muestra tomada

La cuarta variable demográfica es la escolaridad terminada, como se puede ver la mayoría de las personas que contestaron el cuestionario tienen la universidad terminada. Si se observa el histograma se puede ver que la muestra tomada se compone mayormente por personas que terminaron la preparatoria, universidad o maestría.

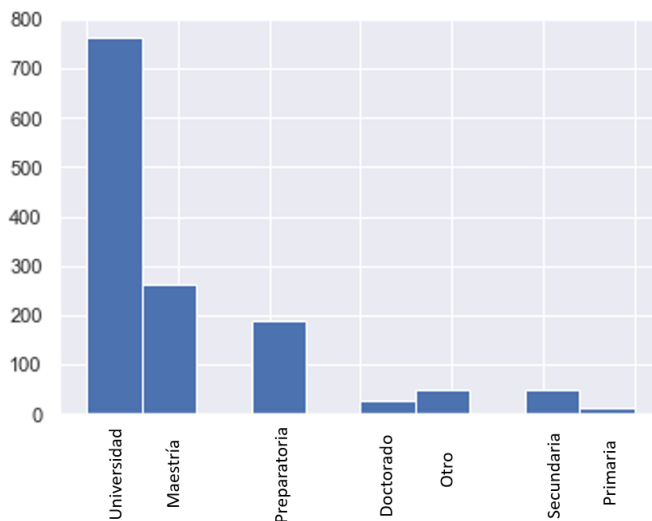


Ilustración 7 - Histograma de profesiones

La quinta variable representa el estado civil de la muestra tomada, en la cual se puede ver que la mayoría son solteros (56.6%), después sigue un 25.5% de personas casadas y un 13.3% de personas en unión libre; por último, se tiene a un 3.9% de personas divorciadas y solo un 0.7% (9 personas) viudas.

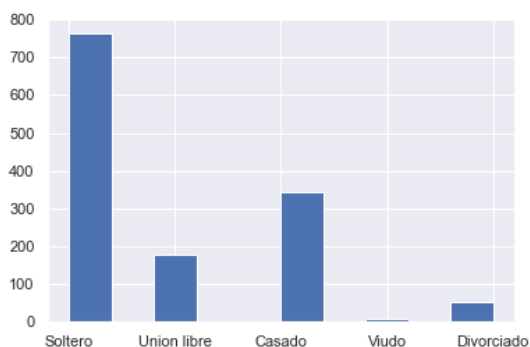


Ilustración 8 - Histograma de estado civil

La siguiente pregunta habla sobre el lugar de residencia de la persona que contestó el cuestionario, por la cantidad de opciones muestro la siguiente tabla de frecuencia donde se puede observar que la mayoría se encuentran viviendo entre Jalisco, Querétaro, Distrito Federal, Michoacán y Nuevo León, mientras el resto de la muestra está distribuida entre los otros estados de la república.

Value	Count	Frequency (%)
Jalisco	464	34.4%
Querétaro	175	13.0%
Distrito Federal	95	7.1%
Michoacan	82	6.1%
Nuevo Leon	81	6.0%
Estado de Mexico	72	5.3%
Guanajuato	43	3.2%
Vivo fuera de México	39	2.9%
Sonora	31	2.3%
Sinaloa	30	2.2%
Other values (23)	235	17.4%

Ilustración 9 - Frecuencia de lugar de residencia

Por último, se tiene la variable “a” que contiene las respuestas a la pregunta ¿has estado o estás siendo atendido por un profesional de la salud mental? Como se puede observar en el histograma de esta pregunta las respuestas están bastante balanceadas teniendo a un 50.3% de personas que están siendo o fueron atendidas por un profesional de la salud mental y a un 49.7% que no.



Ilustración 10 - Histograma de personas que han sido atendidas por un profesional de la salud mental

De todos los datos anteriores y basándonos en lo que más se repite se podría decir que la moda entre la muestra tomada está representada por una mujer católica que tiene entre 30 y 35 años, la universidad terminada, es soltera, vive en Jalisco y ha sido atendida por un profesional de la salud mental.

Cumpliendo con el segundo objetivo específico de identificar las características demográficas de las personas que tienen mayor probabilidad de tener TLP en la muestra tomada, podemos encontrar que en su mayoría son mujeres, 203 contra 31 hombres; el 65% de las personas están entre los 21 y 32 años de edad; el 52% tiene un título universitario, seguido por un 26% que tiene la preparatoria terminada y por último, el 63% son personas solteras.

El resto de las preguntas son utilizadas para poder hacer el cribado inicial que permite saber si una persona tiene o no TLP, como se mencionó antes estas están basadas en los 9 criterios clínicos que, de acuerdo con el DSM-V, contienen las características de una persona con este trastorno; para poder hacer el diagnóstico las personas deben cumplir al menos con 5 de estos criterios, dentro de los cuales destacan 3 criterios cardinales que indican una alta probabilidad de tenerlo.

A continuación, se muestran los 9 criterios clínicos determinantes, las preguntas diseñadas para cada uno se pueden encontrar en la parte de Anexos del presente trabajo, esta lista de preguntas fue cuidadosamente seleccionada y adaptada para que pudiera ser respondida por cualquier persona y aprobada por profesionales de la salud mental.

Criterios clínicos:

1. Esfuerzos desesperados para evitar el desamparo real o imaginado. (No incluir suicida)
2. Patrón de relaciones interpersonales inestables e intensas que se caracteriza por una alternancia entre los extremos de idealización y de devaluación.
3. Alteración de la identidad: inestabilidad intensa y persistente de la autoimagen y del sentido del yo.
4. Impulsividad en dos o más áreas que son potencialmente autolesivas (p. ej., gastos, sexo, drogas, conducción temeraria, atracones alimentarios. (No incluir suicida)
5. Comportamiento, actitud o amenazas recurrentes de suicidio, o comportamiento de automutilación.
6. Inestabilidad afectiva debida a una reactividad notable del estado de ánimo (p. ej., episodios intensos de disforia, irritabilidad o ansiedad que generalmente duran unas horas y, rara vez, más de unos días).
7. Sensación crónica de vacío.
8. Enfado inapropiado e intenso, o dificultad para controlar la ira (p. ej., exhibición frecuente de genio, enfado constante, peleas físicas recurrentes).
9. Ideas paranoides transitorias relacionadas con el estrés o síntomas disociativos graves.

Para la parte descriptiva de estos criterios clínicos solo se presentarán los cardinales, ya que la cantidad de preguntas es muy extensa.

El primer factor cardinal es el de la alteración de la identidad, para este factor se utilizaron 2 preguntas:

1. A menudo me cuesta saber quién soy, me siento perdido dentro de mi y me pregunto quién soy realmente (x_1).

2. Me siento desorientado, sin objetivos, no sé dónde voy en la vida y con facilidad cambio de intereses, creencia (x_5).

El segundo factor habla sobre la impulsividad en dos o más áreas que son potencialmente autolesivas, para el cual se tienen las siguientes preguntas:

1. Cuando me siento mal y/o frustrado, mi relación con el sexo, el uso de drogas, las compras o la comida no lo puedo controlar, aunque después me arrepiento (x_{10}).
2. A veces tengo explosiones de ira o de cólera que no puedo controlar (x_3).

El último factor cardinal hace referencia a la sensación de vacío y se propusieron las siguientes preguntas:

1. Con frecuencia me siento vacío que me angustia y no puedo quitarme, no le encuentro sentido a mi vida y no sé a dónde ir ni que quiero (x_{12}).
2. Me aburro fácilmente de las cosas que previamente me interesaban (x_{17}).

En la siguiente gráfica se pueden ver estas 6 preguntas representadas en gráficas de barras; para esta parte solamente se graficó a las personas totales que contestaron afirmativamente cada una de las preguntas; la parte oscura de cada una de las barras representa la cantidad de personas que, además de haber contestado a cada una de las preguntas tienen altas probabilidades de tener TLP, mientras que en la parte blanca está la cantidad de personas que, a pesar de haber contestado a la pregunta de forma positiva no tienen altas probabilidades de tener TLP.

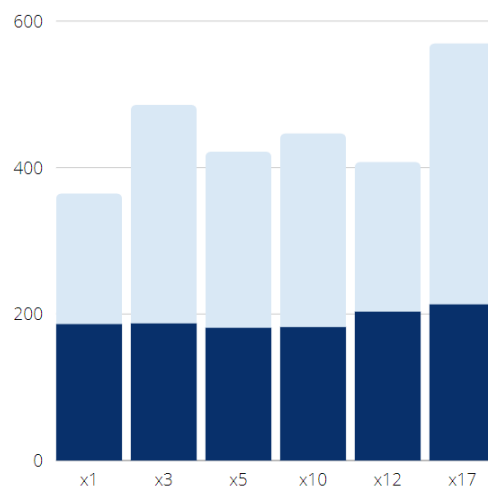


Ilustración 11 - Variables significativas

Con este análisis se puede conocer los rasgos de personalidad más comunes en la muestra tomada, el cuál es el tercer objetivo específico del presente trabajo. Como se puede observar en las variables x_1 y x_{12} la porción de personas con altas probabilidades de tener TLP es casi la mitad de la cantidad que contestaron positivamente a estas preguntas, mientras que para el resto, aunque no es la mitad, la porción de personas con altas probabilidades de tener TLP también es alta.

Por último, se tiene la variable de respuesta y , la cual indica si una persona tiene alta probabilidad de tener TLP (1), o no la tiene (0), y fue calculada para cada una de las personas que contestaron el cuestionario con ayuda de la psicóloga especialista en el tema.

En el siguiente histograma se puede ver que son poco más de 200 personas de las 1300 con alta probabilidad de tener TLP, consultando la tabla de frecuencias se puede ver con exactitud la cantidad de personas que tienen altas probabilidades de tener TLP la cuál es de 238, estas personas representan un 17.7% de la muestra total tomada el cual es un porcentaje muy alto considerando que, de acuerdo con las cifras oficiales no debería pasar del 1.5%. Incluso si se quisiera empatar con las cifras oficiales se tendrían que descartar al más del 90% de las personas en el diagnóstico oficial lo cual es muy poco probable que suceda. Con esto se cumple el cuarto objetivo específico propuesto para este proyecto.

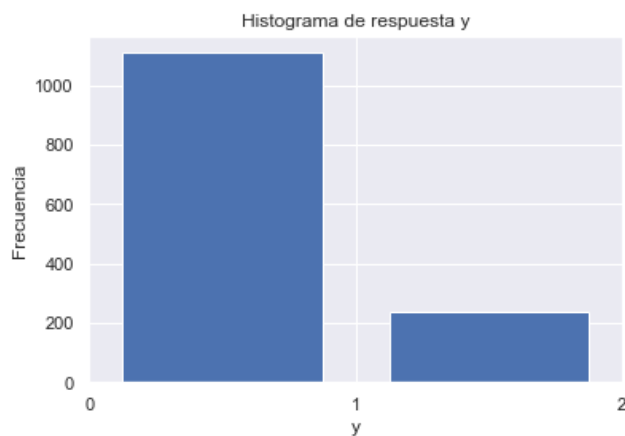


Ilustración 12 - Histograma de respuesta y

Frequency Table			
Value	Count	Frequency (%)	
0	1109	82.3%	<div style="width: 82.3%; height: 10px; background-color: #1f77b4;"></div>
1	238	17.7%	<div style="width: 17.7%; height: 10px; background-color: #1f77b4;"></div>

Ilustración 13 - Frecuencia de personas con y si probabilidades de tener TLP

Como se mencionó en un inicio, uno de los objetivos es implementar al menos 3 modelos de clasificación para determinar, en el cribado inicial, si una persona tiene altas probabilidades de tener trastorno límite de la personalidad.

Estos 3 modelos son una regresión logística, un árbol de decisión para clasificación y una red neuronal. Para poder hacer el ejercicio se tomaron, en todos los modelos, las variables x_1 a x_{17} como variables de entrada y la variable y como la variable de salida. Asimismo, se destinó un 80% de los datos al entrenamiento de los diferentes modelos y un 20% para pruebas.

Para poder comparar los modelos y ver cuál es el mejor para cumplir con el objetivo principal del proyecto se evaluaron con las métricas de “accuracy”, “precision” y “recall” y por otro lado también se obtuvo la matriz de confusión para cada uno de los modelos. Estas métricas nos ayudan a evaluar el porcentaje de datos que el modelo clasificó de forma correcta y serán explicadas a más detalle en la siguiente sección.

El primer modelo que se evaluó fue el de la regresión logística. Este modelo es el que se aplica con mayor frecuencia para resolver problemas de clasificación ya que es un modelo muy sencillo y del que normalmente se obtienen muy buenos resultados.

Para poder hacer este modelo el único preprocesamiento que se le hizo aplicó a los datos fue la conversión de las variables independientes de “si” y “no” a 1 y 0 respectivamente, así como la conversión de estas variables de texto a números enteros.

Por otro lado, todas las variables x_1 a x_{17} fueron utilizadas resultando el siguiente polinomio de primer grado:

$$y = -15.50 + 1.44x_1 + 1.55x_2 + 1.33x_3 + 1.69x_4 + 1.28x_5 + 2.08x_6 + 1.60x_7 + 1.56x_8 + 1.16x_9 + 1.60x_{10} + 2.35x_{11} + 1.37x_{12} + 1.52x_{13} + 1.57x_{14} + 1.79x_{15} + 1.80x_{16} + 1.55x_{17} \quad (16)$$

En la siguiente tabla se pueden ver los valores obtenidos para cada una de las métricas, como se puede observar en el accuracy se obtuvo un 98.1%, en la precision un 100% y en el recall un 89.4%, lo cual nos indica que es un buen modelo lo cual queda reafirmado con un score del modelo de 98.15% y un score de la curva ROC-AUC del modelo de un 94.68%

Accuracy	Precision	Recall
0.981	1.000	0.894

Ilustración 14 - Métricas para la regresión logística

Por otro lado, se muestra la matriz de confusión, en esta se puede ver la cantidad de datos que se clasificaron de forma adecuada, de igual manera se va a explicar más a fondo en la siguiente sección.

Como se puede ver en la siguiente matriz, la regresión logística clasificó correctamente a 223 personas con bajas probabilidades de tener TLP lo cual representa un 82.59%, y a un 15.56% o 42 personas con altas probabilidades de tener TLP; sin embargo, clasificó incorrectamente a 5 personas como personas con bajas probabilidades de tener TLP cuando en realidad tienen altas probabilidades, pero no identificó a ninguna con altas probabilidades de tener TLP cuando en realidad no las tienen.

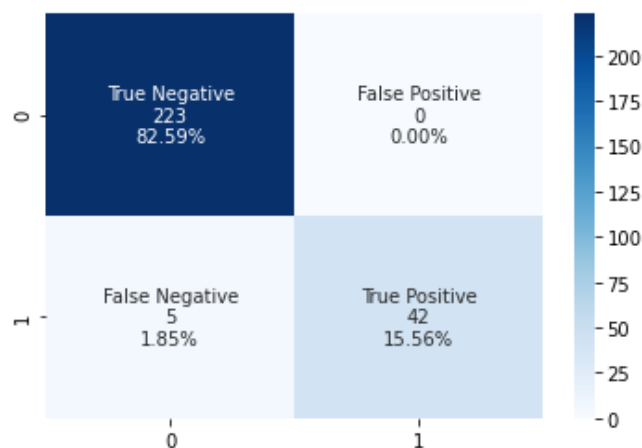


Ilustración 15 - Matriz de confusión para la regresión logística

El siguiente modelo que se evaluó fue el árbol de decisión para clasificación, para encontrar los mejores parámetros se utilizó la herramienta GridSearchCV la cual hace la combinación de todos los parámetros especificados para encontrar la que de como resultado el mejor modelo. En este caso la mejor combinación resultó de la siguiente forma:

- Medición de pureza: Gini
- Divisor: Best
- Estado aleatorio: 8
- Máxima profundidad: 12
- Máxima característica: Auto

A continuación, se describe lo que significa cada uno de estos parámetros:

Medición de pureza: La función para medir la calidad de una división. Los criterios admitidos son "gini" para la impureza de Gini y "entropía" para la ganancia de información.

Divisor: La estrategia utilizada para elegir la división en cada nodo. Las estrategias admitidas son "mejores" para elegir la mejor división y "aleatorias" para elegir la mejor división aleatoria.

Estado aleatorio: Controla la aleatoriedad del estimador. Las características siempre se permutan aleatoriamente en cada división, incluso si el divisor está configurado como "mejor".

Máxima profundidad: La profundidad máxima del árbol. Si es ninguno, los nodos se expanden hasta que todas las hojas sean puras o hasta que todas las hojas contengan menos de `min_samples_split` muestras.

Máxima característica: La cantidad de características a considerar al buscar la mejor división. (scikit learn, 2021)

Para este segundo modelo se obtuvieron los resultados mostrados en la siguiente imagen. Como se puede ver la métrica que salió mejor fue la del accuracy para la cual se obtuvo un valor de 94.4%, en el caso de la precision y el recall se obtuvieron valores más bajos de 79.6% y 88.6% respectivamente; sin embargo no se puede aún concluir que sea un mejor o peor modelo que la regresión logística.

```
Árbol de decisión:  
Accuracy    Precision    Recall  
0.944    0.796    0.886
```

Ilustración 16 - Métricas para el árbol de decisión

Si se analiza la matriz de confusión generada para el árbol de decisión se puede observar que el modelo clasificó correctamente a 216 personas con altas probabilidades de tener TLP (80%) y a 39 personas con altas probabilidades de tener TLP (14.44%), mientras que clasificó erróneamente a 10 personas con altas probabilidades de tener TLP cuando realmente no las tienen y a 5 personas con bajas probabilidades de tener TLP cuando en realidad si las tienen.

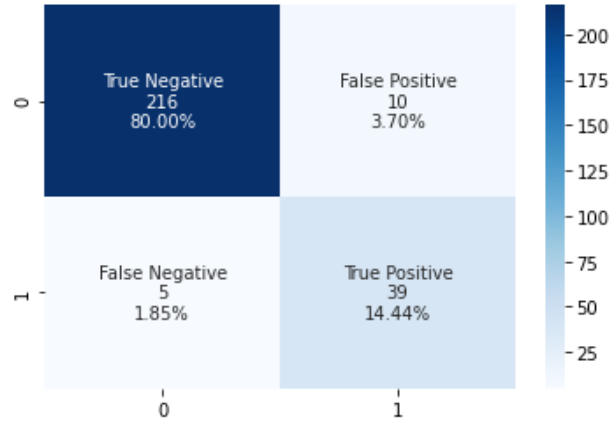


Ilustración 17 - Matriz de confusión para el árbol de decisión

Por último, se implementó una red neuronal secuencia con una sola capa oculta con ayuda de la cual se obtuvieron los resultados mostrados en la imagen que se muestra abajo, como se puede ver los porcentajes de todas las métricas están por encima del 95% siendo la precision la más baja con 95.3% y el accuracy la más alta con un 98.9%.

Red Neuronal			
	Accuracy	Precision	Recall
Test	0.989	0.953	0.976

Ilustración 18 - Métricas para la red neuronal

Al observar la matriz de confusión para la red neuronal se puede apreciar como prácticamente todos los datos fueron clasificados correctamente y sólo se tienen 2 personas clasificadas con altas probabilidades de tener TLP cuando en realidad no lo tienen y 1 persona con bajas probabilidades de tener TLP cuando en realidad si lo tienen.

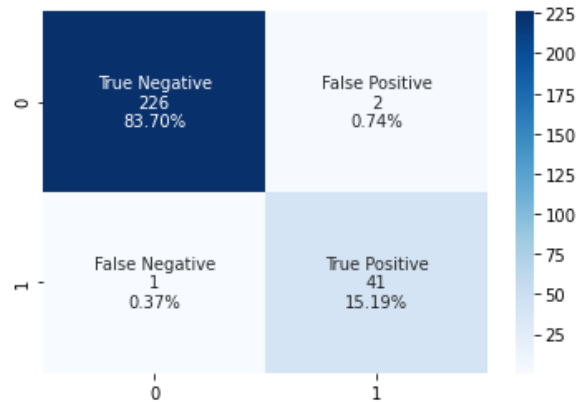


Ilustración 19 - Matriz de confusión para la red neuronal

La estructura de la red neuronal consiste en una red secuencial con una capa oculta en la cual se tienen 2 neuronas y se utiliza una función de activación sigmoide utilizando las 17 variables $x_1 - x_{17}$.

Una vez obtenidos los resultados de los 3 modelos se puede proceder a elegir cuál es el mejor para resolver este problema en particular.

5. CONCLUSIONES Y TRABAJO FUTURO

5.1 Conclusiones

Como se mencionaba en la sección anterior se utilizaron 3 métricas diferentes para evaluar cada uno de los modelos, así como la matriz de confusión de estos. Antes de mostrar cual es el mejor modelo es importante entender lo que cada una de estas métricas significa:

Accuracy: Es el porcentaje total de elementos clasificados correctamente.

Precision: Es el número de elementos identificados correctamente como positivo de un total de elementos identificados como positivos.

Recall: Es el número de elementos identificados correctamente como positivos del total de positivos verdaderos. (Sitio big data, 2019)

La forma de calcular cada una de estas métricas se puede observar en la siguiente figura:

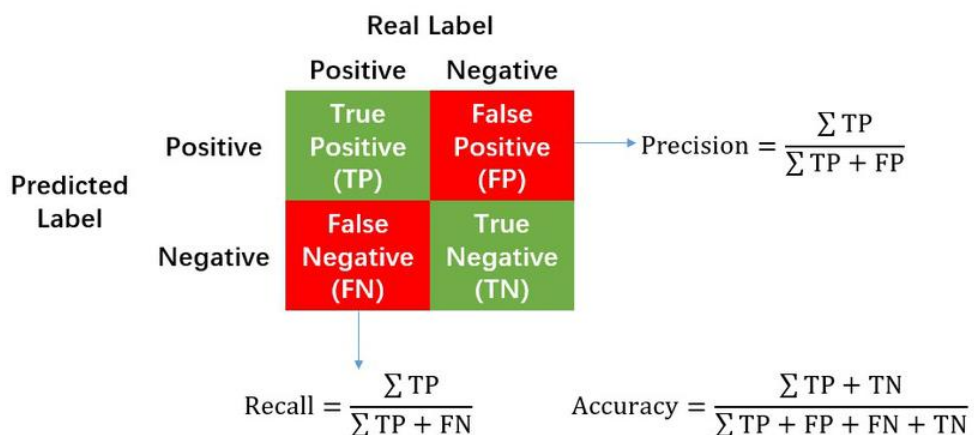


Ilustración 20 - Métricas para medir los modelos de clasificación

Recordando un poco el objetivo principal del proyecto, que es identificar a las personas con altas probabilidades de tener TLP, lo que se quiere reducir es la cantidad de falsos

negativos que el modelo pueda identificar. Realmente no afecta mucho tener Falsos Positivos identificados ya que estos pueden ser descartados en las sesiones de terapia con el psicólogo o psiquiatra, pero si desde un inicio se descarta que una persona con TLP lo pueda tener será más difícil y tardado para el especialista de la salud mental regresar e identificar este diagnóstico lo cual afectará directamente al paciente en su tratamiento.

Con base en esto se puede hacer la comparación de los resultados obtenidos en cada uno de los modelos. Como se puede ver en la siguiente tabla la regresión logística y el árbol de decisión tienen resultados muy parecidos; el árbol de decisión tiene un porcentaje de recall muy similar que la regresión logística, pero ambos tienen 5 falsos negativos, por lo que se puede complicar la elección entre estos dos modelos.

Por otro lado, la red neuronal solamente tiene un falso negativo y un recall del 97.6%, con base en esto se podría decir que este es el mejor modelo para cumplir con el objetivo principal de este trabajo.

Tabla 3 - Comparación de modelos

Modelo	Regresión Logística	Árbol de decisión	Red neuronal
Recall	89.4%	88.6%	97.6
Falsos Negativos	5	5	1

5.2 Trabajo futuro

Lo siguiente por hacer es validar el modelo en aplicaciones reales; actualmente se está trabajando en hacer la base de datos más grande para la validación del modelo, conformada principalmente con personas ya diagnosticadas con TLP.

Por otro lado, se puede ver si este modelo puede adaptarse al resto de los trastornos de personalidad, al igual que se hizo con este trastorno, todo se basará en los criterios clínicos ya establecidos. De ser posible esta adaptación, se explorará la posibilidad de tener una sola herramienta capaz de hacer un cribado inicial que te diga que trastorno o trastornos puede tener una persona.

Por último se probarán otros modelos como bagging trees y SVM para clasificación para ver si se tienen mejores resultados y se revisará si los registros que fueron mal clasificados son los mismos en los 3 modelos obtenidos.

6. BIBLIOGRAFÍA

- American Journal Psychiatry. (1985). Research on mental illness and addictive disorders. *The American Journal of Psychiatry*, 142.
- American Psychological Association. (2020). *APA Dictionary of Psychology*. Obtenido de American Psychological Association: <https://dictionary.apa.org/borderline-personality-disorder>
- Balgañón, P. (Febrero de 1990). Servicio de Salud Mental de la Consejería de Salud. *Comunicación personal*. Madrid.
- Bergeret, J. (1990). Los estados límites 20 años después. (D. Wintrebert, Entrevistador)
- Cavan, N. (19 de Diciembre de 2018). *7 Datos sorprendentes que no conocías de la psicología en México*. Obtenido de Deep mind Marketing: <https://deepmindmarketing.com/la-psicologia-en-mexico/#:~:text=Para%20que%20te%20des%20una,%25%20y%20el%2028%25%20hombres>.
- Corsaro, B. (1992). *¿Los pacientes borderline son bien diagnosticados?* Buenos Aires.
- Departamento de salud y servicios humanos de Estados Unidos. (Junio de 2018). *Publicaciones: Trastorno límite de la personalidad*. Obtenido de Instituto nacional de salud mental: https://www.nimh.nih.gov/health/publications/espanol/trastorno-limite-de-la-personalidad/19-mh-8087s-trastornolimitedelapersonalidad_158250.pdf
- Edwards, J. (4 de December de 2020). *A machine learning approach identifies unique predictors of borderline personality disorder*. Obtenido de The Association for Child and Adolescent Mental Health: <https://www.acamh.org/research-digest/a-machine-learning-approach-identifies-unique-predictors-of-borderline-personality-disorder/>
- Emerging Technology From The Arxiv. (3 de Agosto de 2017). *La inteligencia artificial, la realidad virtual y el "big data" están revolucionando la psiquiatría*. Obtenido de MIT Technology Review: <https://www.technologyreview.es/s/8474/la-inteligencia-artificial-la-realidad-virtual-y-el-big-data-estan-revolucionando-la>
- Eurecat. (10 de Octubre de 2019). *Aplican Inteligencia Artificial para gestionar crisis emocionales en pacientes con trastorno límite de personalidad*. Obtenido de Eurecat: <https://eurecat.org/es/inteligencia-artificial-gestionar-crisis-emocionales-pacientes-trastorno-limite-personalidad/>

- Fedak, V. (8 de November de 2018). *Top 10 Most Popular AI Models*. Obtenido de DZone:
<https://dzone.com/articles/top-10-most-popular-ai-models>
- Fernández-Guerrero, M. J. (26 de Junio de 2017). *Los confusos límites del trastorno límite*. Obtenido de Scientific Electronic Library Online:
<http://scielo.isciii.es/pdf/neuropsi/v37n132/0211-5735-raen-37-132-0399.pdf>
- Fundación Pública Andaluza para la Integración Social de Personas con Enfermedad Mental. (14 de Abril de 2021). *Conoce los problemas de salud mental*. Obtenido de 1decada4: <https://www.1decada4.es/mod/page/view.php?id=30>
- Gareth James, D. W. (2021). Classification. En D. W. Gareth James, *An Introduction to Statistical Learning with Applications with R* (págs. 133-139).
- Gareth James, D. W. (2021). Deep Learning. En D. W. Gareth James, *An Introduction to Statistical Learning with Applications in R* (págs. 404-410).
- Gareth James, D. W. (2021). Tree-Based Methods. En D. W. Gareth James, *An Introduction to Statistical Learning with Applications in R* (págs. 327-325).
- Gerard Heinze, G. d.-H. (2016). *Los especialistas en psiquiatría en México*. Obtenido de Scientific Electronic Library Online:
http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-33252016000200069#:~:text=Existen%204393%20psiquiatras%20en%20M%C3%A9xico,por%20cada%20100%20000%20habitantes.
- IBM. (2020). *El modelo de redes neuronales*. Obtenido de IBM Documentation - SPSS Modeler: <https://www.ibm.com/docs/es/spss-modeler/SaaS?topic=networks-neural-model>
- Imanol Perez, G. G. (2018). *A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder*. Oxford: Creative Commons Attribution 4.0.
- INEGI. (2020). *INEGI*. Obtenido de Población: <https://www.inegi.org.mx/temas/estructura/>
- Kohonen, T. (1989). *Self-organization and associative memory*. New York: Springer Verlag.
- Matich, D. J. (Marzo de 2001). *Informática aplicada a la ingeniería de procesos*. Obtenido de Facultad Regional Rosario:
https://www.frrro.utn.edu.ar/repositorio/catedras/quimica/5_ano/orientadora1/monografias/matich-redesneuronales.pdf
- Millon, T. y. (1998). *Trastornos de la personalidad. Más allá del DSM-IV*. Barcelona: Masson, S. A.
- Oracle. (2021). *¿Qué es la Inteligencia Artificial?* Obtenido de Oracle Mexico:
<https://www.oracle.com/mx/artificial-intelligence/what-is-ai/>
- Oracle. (2021). *Definición de aprendizaje automático*. Obtenido de Oracle Mexico:
<https://www.oracle.com/mx/data-science/machine-learning/what-is-machine-learning/>
- Pandas. (17 de October de 2021). *pandas*. Obtenido de pandas: <https://pandas.pydata.org/>
- Pitts, W. S. (1943). A logical calculus of the ideas immanent in nervous activity. *Society for Mathematical Biology*, 115-133.
- Poma, C. M. (s.f.). *Universidad Nacioal de San Marcos*. Obtenido de Biblioteca digital de la Universidad Nacioal de San Marcos:

https://sisbib.unmsm.edu.pe/bibvirtualdata/Tesis/Basic/Salcedo_pc/enPDF/Cap2.PDF

- Pomenta, S. (1991). El borderline o la manera narcisista de vivir. (B. Corsario, Entrevistador) Caracas-Barcelona: Científico-Medicina Editora CA.
- Pose, M. G. (Agosto de 2013). Introducción a las redes de neuronas artificiales. Coruña: Depto. Tecnologías de la Información y las comunicaciones.
- Secretaria de Salud MX. (09 de Julio de 2018). *Alrededor de 1.5 % de la población padece trastorno límite de la personalidad*. Obtenido de Gobierno de México: <https://www.gob.mx/salud/prensa/270-alrededor-de-1-5-de-la-poblacion-padece-trastorno-limite-de-la-personalidad>
- Selby, E. A. (2021). *Temporal Bayesian Network modeling approach to evaluating the emotional cascade model of borderline personality disorder*. Obtenido de American Psychological Association: <https://psycnet.apa.org/record/2020-26461-001>
- Sitio big data. (19 de January de 2019). *Machine Learning: Selección Métricas de clasificación*. Obtenido de sitiobigdata: <https://sitiobigdata.com/2019/01/19/machine-learning-metrica-clasificacion-parte-3/#>
- Stuart Russel, P. N. (2004). *Inteligencia artificial: un enfoque moderno*. Madrid: Pearson.
- Szalai, J. (2020). *The potential use of artificial intelligence in the therapy of borderline personality disorder*. Budapest: Hungarian Ministry of Innovation and Technology.
- Tarnopolsky A, B. M. (1987). Borderline personality. A review of recent research. *British Journal of Psychiatry*, 724-734.
- Wasserman, P. D. (1989). *Neural computing: theory and practice*. New York: Van Nostrand Reinhold Co.

APÉNDICE A. Cuestionario para recopilar base de datos

Prueba encuesta de cribado de TLP

Sexo:

- Mujer
- Hombre
- Prefiero no decirlo

Edad:

Religión:

- Catolicismo
- Cristianismo
- Budismo
- Judaísmo
- Islamismo
- Agnosticismo
- Ateísmo
- Otro

Escolaridad terminada:

- Primaria
- Secundaria
- Preparatoria
- Universidad
- Maestría
- Doctorado
- Otra

Estado civil:

- Soltero
- Casado
- Divorciado
- Unión libre
- Viudo

Lugar de residencia:

¿Ha sido o está siendo tratado por un especialista de la salud mental?

- Si
- No

A menudo me cuesta saber quién soy, me siento perdido dentro de mi y me pregunto quién soy realmente.

- Si
- No

Tengo miedo a quedarme solo

- Si
- No

A veces tengo explosiones de ira o de cólera que no puedo controlar

- Si
- No

Me cuesta mucho tener relaciones sociales y afectivas estables. Suelo acabar teniendo conflictos.

- Si
- No

Me siento desorientado, sin objetivos, no sé a dónde voy en la vida y con facilidad cambio de intereses y/o creencias.

- Si
- No

Alguna vez he planeado, amenazado y/o intentado suicidarme.

- Si
- No

Mis sentimientos hacia personas importantes de mi vida a veces cambian rápidamente de un extremo a otro (de admiración a decepción, de amor a odio)

- Si
- No

Suelo estar irritable y me enfado con facilidad

- Si
- No

A veces, cuando las cosas van mal, siento que lo que me rodea no es real, que estoy viviendo un sueño o veo pasar la vida delante de mi como si fuera una película.

- Si
- No

Cuando me siento mal y/o frustrado, mi relación con el sexo, el uso de drogas, las compras o la comida no lo puedo controlar, aunque después me arrepiento.

- Si
- No

Mi ánimo no suele ser estable, cambiando de un extremo a otro en un mismo día, puedo dar saltos de alegría y enseguida sentirme profundamente triste o deprimido/a.

- Si
- No

Con frecuencia siento un vacío que me angustia y no puedo quitarme, no le encuentro sentido a mi vida y no sé a dónde ir ni que quiero.

- Si
- No

Con tal de que no me abandonen puedo halagar o complacer a la persona, o agredir con el objetivo de que no me dejen.

- Si
- No

Tengo problemas con el control y manejo de mis impulsos, lo que me lleva a actuar sin pensar.

- Si
- No

Cuando estoy bajo estrés o en una situación de intensa emoción me he sentido observado, criticado o juzgado por quienes me rodean.

- Si
- No

Algunas veces me he causado daño o me he herido intencionalmente (cortarme, quemarme, morderme, jalarme el cabello,...)

- Si
- No

Me aburro fácilmente de las cosas que previamente me interesaban.

- Si
- No