



Instituto Tecnológico de Monterrey

Resumen Ejecutivo

Proyecto Integrador: Detección
de trastorno bipolar

Profesora: Dra. Grettel Barceló

Alumnos

- Luis Alfonso Sabanero Esquivel – A01273286
- Carlos D. Mendoza Paz – A01793101
- Guillermo A. Muñiz Hermosillo – A01793027

Índice

Índice de figuras.....	2
Índice de tablas.....	2
Introducción.....	3
Definición del problema	3
Hallazgos del Análisis Exploratorio de Datos	4
Importación y Agrupación de Datos	4
Estadísticas Clave y Tendencias Descubiertas.....	4
Distribución de Diagnósticos Psiquiátricos:	4
Edad y Género	6
Relaciones Familiares y Trastornos.....	7
Modelos Generados y Elección del Modelo Final	8
Modelos Generados.....	8
Regresión Logística	8
Arboles de Decisión (DT).....	9
Blending NN-DT (Neural Network-Decision Tree)	9
Bosques Aleatorios (Random Forest)	9
Red Neuronal (NN)	9
Elección del modelo final	9
Recomendaciones Clave para Implementar la Solución	11
Recomendaciones Específicas	11
Validación Adicional.....	11
Monitorización Continua	11
Optimización de Hiperparámetros	11
Actualización Periódica del Modelo.....	11
Estandarización del Número de Familiares Encuestados	12
Capacitación del Personal.....	12
Infraestructura Tecnológica.....	12
Políticas de Privacidad y Seguridad	13
Comunicación Efectiva.....	13
Recursos Necesarios	13
Datos de Entrenamiento y Validación	13
Equipo Técnico.....	13
Infraestructura Tecnológica.....	13
Capacitación	13
Herramientas de Monitorización:.....	14
Análisis Costo-Beneficio	14
Costos.....	14
Recolección de los datos.....	14
Infraestructura tecnológica	14
Desarrollo del modelo	14
Implementación y capacitación	14

Beneficios potenciales	16
Beneficios cuantificables	16
Beneficios intangibles	16
<i>Riesgos y Desafíos de la Solución.....</i>	16
Identificación y clasificación de Riesgos	16
Mitigación de Riesgos	17
<i>Conclusión.....</i>	18
<i>Referencias</i>	18

Índice de figuras

Figura 1 Distribución de diagnósticos psiquiátricos familiar 1	5
Figura 2 Distribución de diagnósticos psiquiátricos familiar 2	6
Figura 3 Grafico de edad y genero	6
Figura 4 Gráfico de bipolaridad en hombres y mujeres	7
Figura 5 Grafico de frecuencias por parentesco familiar 1	7
Figura 6 Grafico de frecuencias por parentesco familiar 2	8
Figura 7 Top 10 de características más importantes del modelo final	10

Índice de tablas

Tabla 1 Tabla de costos estimados.....	15
Tabla 2 Tabla de riesgos y desafíos	16
Tabla 3 Tabla de mitigación de riesgos	17

Introducción

Este proyecto se lleva a cabo en colaboración con el Hospital Universitario UANL y el Instituto HOOB, con el objetivo principal de mejorar la detección temprana y el manejo del trastorno bipolar. El objetivo es proporcionar una atención más efectiva y oportuna a los pacientes y sus familias, algo crucial en el entorno de salud actual.

El trabajo presentado utiliza la Inteligencia Artificial para abordar una problemática real en el ámbito médico: el trastorno bipolar. Este trastorno mental se caracteriza por cambios extremos en el estado de ánimo, incluyendo episodios de manía (estado de ánimo anormalmente elevado o irritable) y depresión, así como variaciones en la energía, actividad y la capacidad de llevar a cabo tareas diarias.

Como parte de la colaboración con el Instituto HOOB, se llevó a cabo una investigación sobre el trastorno bipolar mediante encuestas a pacientes diagnosticados y a personas sin este padecimiento. Se analizaron antecedentes familiares de trastornos psiquiátricos y factores socioeconómicos comúnmente asociados a esta condición. El objetivo es identificar el riesgo intrafamiliar de desarrollar trastorno bipolar a partir del análisis de las relaciones de parentesco entre los pacientes diagnosticados y otros miembros de su familia con condiciones psiquiátricas o psicológicas.

En el contexto de la medicina actual, especialmente en el ámbito de la salud mental, existe una creciente necesidad de mejorar los métodos de diagnóstico y tratamiento de los trastornos psiquiátricos. Esta necesidad se intensificó con la pandemia, durante la cual muchas personas experimentaron problemas de salud mental, a menudo sin ser conscientes de ello. Promover la salud mental en la población es, por tanto, de suma importancia.

El trastorno bipolar afecta profundamente la vida de los individuos y sus familias, causando cambios extremos en el estado de ánimo, niveles de energía y la capacidad para realizar actividades diarias. Por ello, una correcta identificación y tratamiento temprano de este padecimiento es fundamental para mejorar la calidad de vida de los pacientes y reducir la carga sobre el sistema de salud a través del uso de un modelo de Inteligencia Artificial que nos permita hacer un prediagnostico.

Definición del problema

Nuestro proyecto busca abordar la prevalencia del trastorno bipolar en familiares que cuentan con antecedentes de este u otros trastornos psiquiátricos, así como algunos factores socioeconómicos que normalmente se asocian a este trastorno. Intentamos centrar nuestra investigación en identificar el riesgo intrafamiliar de desarrollar trastorno bipolar, analizando la relación de parentesco entre los pacientes diagnosticados con trastorno bipolar y otros miembros de su familia que padecen dichas condiciones u otros padecimientos psiquiátricos o psicológicos.

Creemos que, en la industria de la medicina actual, en especial en el área de la salud mental, existe la necesidad creciente de mejorar los métodos de diagnóstico y tratamiento de trastornos psiquiátricos. El trastorno bipolar es una condición mental que afecta profundamente la vida de los individuos y sus familias, causando cambios extremos en el estado de ánimo, los niveles de energía y la capacidad de realizar actividades diarias. Es así como la correcta identificación y el tratamiento temprano de esta enfermedad toma un papel preponderante para poder mejorar la calidad de vida de los pacientes y reducir la carga sobre el sistema de salud.

Hallazgos del Análisis Exploratorio de Datos

El análisis exploratorio de datos (EDA) realizado en este proyecto de investigación reveló varios hallazgos importantes. A continuación, presentamos un resumen de las estadísticas clave, tendencias descubiertas y puntos clave que respaldan la necesidad de la solución propuesta.

Importación y Agrupación de Datos

El primer paso del análisis exploratorio de datos (EDA) consistió en la importación y agrupación de los datos de los pacientes y sus familiares. Los datos fueron importados desde el archivo de base de datos originales, provisto por la UANL y el instituto HOOB en formato CSV. Dicho archivo contiene información sobre los pacientes diagnosticados con trastorno bipolar y sus familiares, incluyendo datos demográficos, diagnósticos psiquiátricos y relaciones familiares.

Antes de proceder con el análisis, se realizaron varias tareas de limpieza de datos para asegurar la calidad y consistencia de estos:

1. **Manejo de Valores Nulos:** Se identificaron y trataron los valores nulos presentes en el conjunto de datos. Dependiendo del contexto, estos valores fueron eliminados o imputados utilizando técnicas adecuadas.
2. **Agrupamiento de los datos:** Para un análisis más eficiente, los datos fueron agrupados por familias utilizando el campo: Family_ID y las relaciones familiares como métodos de agrupación. Esto permitió una comprensión más clara de los patrones intrafamiliares los cuales pueden ser cruciales para entender la herencia y la influencia de factores genéticos en los trastornos psiquiátricos.
3. **Creación de Nuevas Variables:** Se crearon nuevas variables derivadas de los datos existentes para proporcionar perspectivas adicionales y enriquecer el análisis. Por ejemplo, se agregaron como base a los 2 familiares con un parentesco más cercano al paciente (probando). El orden establecido para la creación de estas características fue el siguiente:
 1. Mamá
 2. Papá
 3. Hermanos
 4. Abuelos
 5. Otros

Así pues, cada Familia está asociada a estas nuevas características con los 2 familiares más cercanos genéticamente hablando al probando. También al existir pacientes con más de 2 familiares en la base de datos y evitar la pérdida de datos, se crearon otro conjunto de características las cuales actúan como un XOR lógico, en el cual la columna contendrá un valor positivo (True) Si existen familiares que padezcan la enfermedad psiquiátrica referida en la columna y Falso (False) en el caso contrario.

Después de completar este paso, se creó un nuevo archivo llamado "final_dataframe.xlsx". Este archivo se utiliza para extraer datos y entrenar los modelos siguientes. Además, se emplea este archivo para generar un resumen que incluye estadísticas clave, tendencias descubiertas e información relevante que respalda la necesidad de la solución propuesta.

Estadísticas Clave y Tendencias Descubiertas

Distribución de Diagnósticos Psiquiátricos:

- Se identificó una prevalencia significativa de trastornos psiquiátricos entre los familiares de pacientes con trastorno bipolar.
- La **depresión** y la **ansiedad** fueron los trastornos más comúnmente diagnosticados en estos familiares, seguido por **trastornos de personalidad y abuso de sustancias**.

Familiar 1

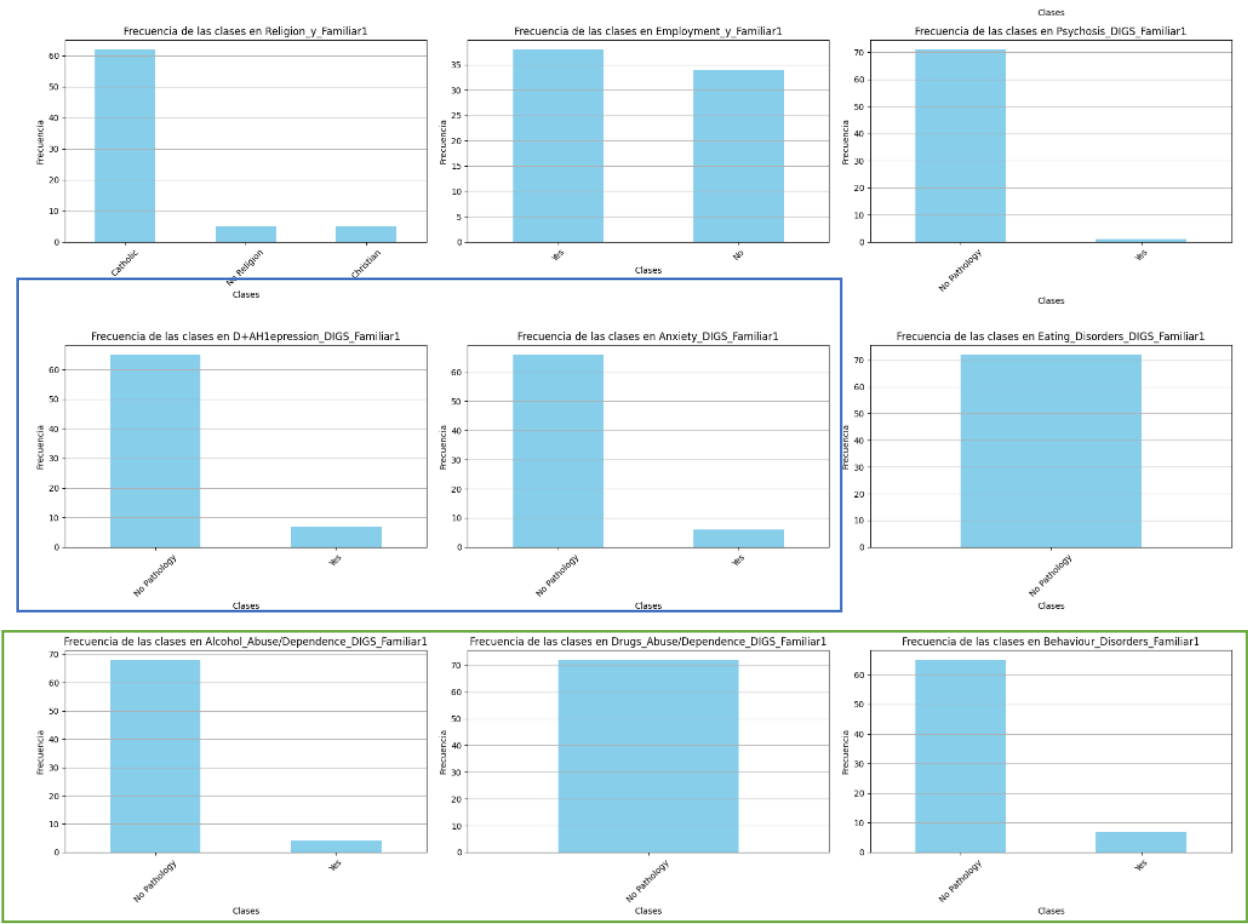


Figura 1 Distribución de diagnósticos psiquiátricos familiar 1

Familiar 2

Azul: depresión y la ansiedad

Verde: trastornos de personalidad y abuso de sustancias.

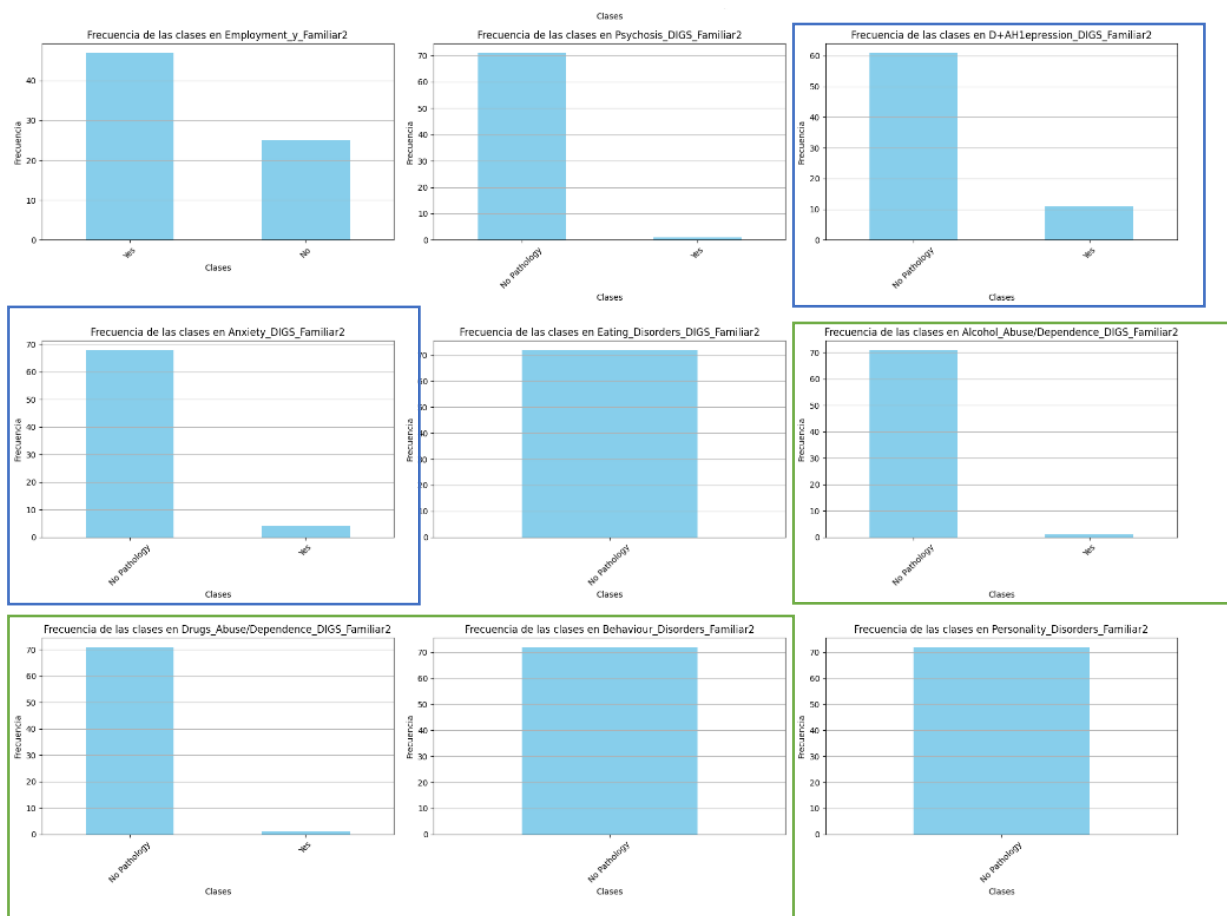


Figura 2 Distribución de diagnósticos psiquiátricos familiar 2

Edad y Género

- La mayoría de los pacientes diagnosticados con trastorno bipolar tenían edades comprendidas entre los 20 y 40 años.

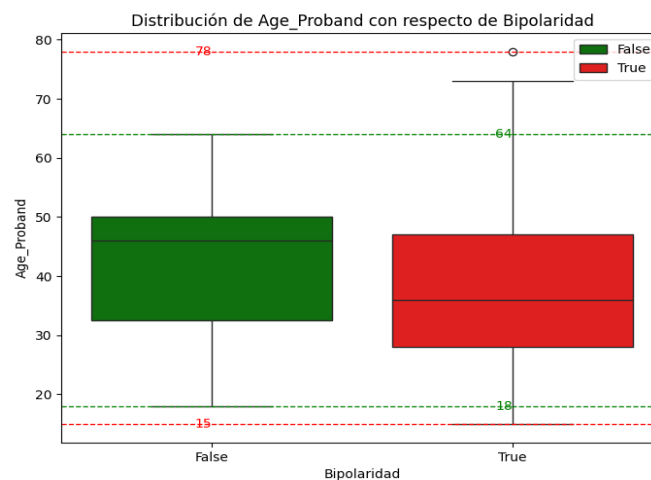


Figura 3 Grafico de edad y genero

- Se observó una ligera predominancia del género femenino en los diagnósticos de trastorno bipolar.

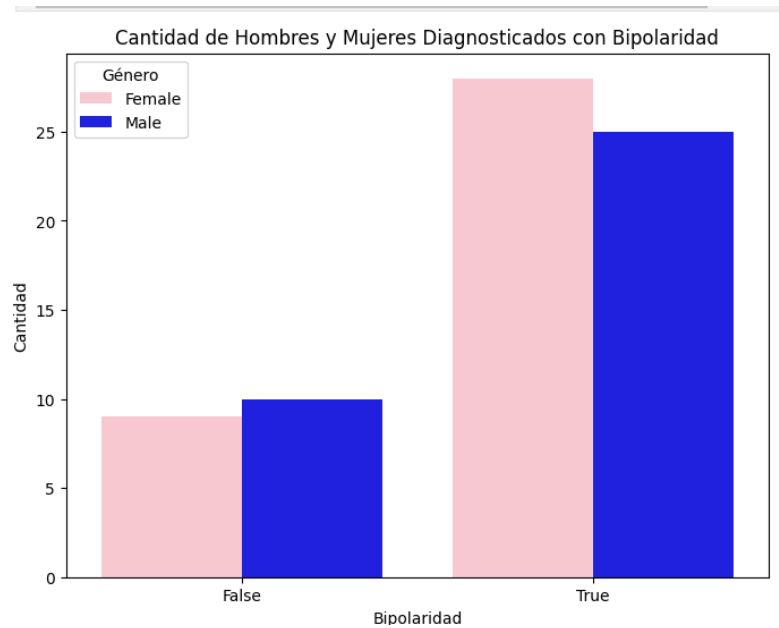


Figura 4 Gráfico de bipolaridad en hombres y mujeres

Relaciones Familiares y Trastornos

- Los datos mostraron que los familiares de primer grado (padres, hermanos, hijos) tienen una mayor prevalencia de trastornos bipolares y otros trastornos psiquiátricos en comparación con los familiares de segundo grado (tíos, abuelos, primos).

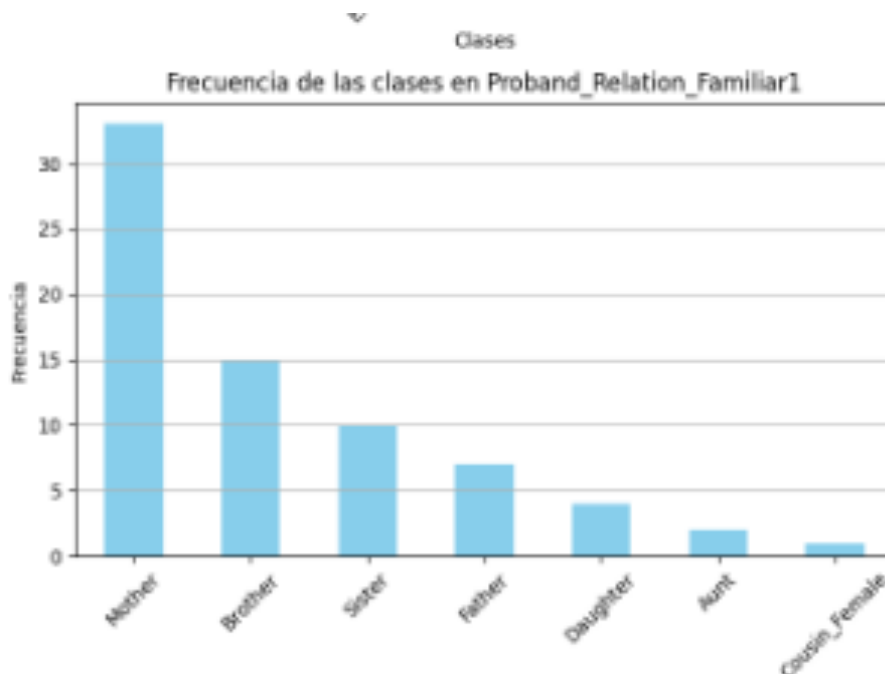


Figura 5 Grafico de frecuencias por parentesco familiar 1

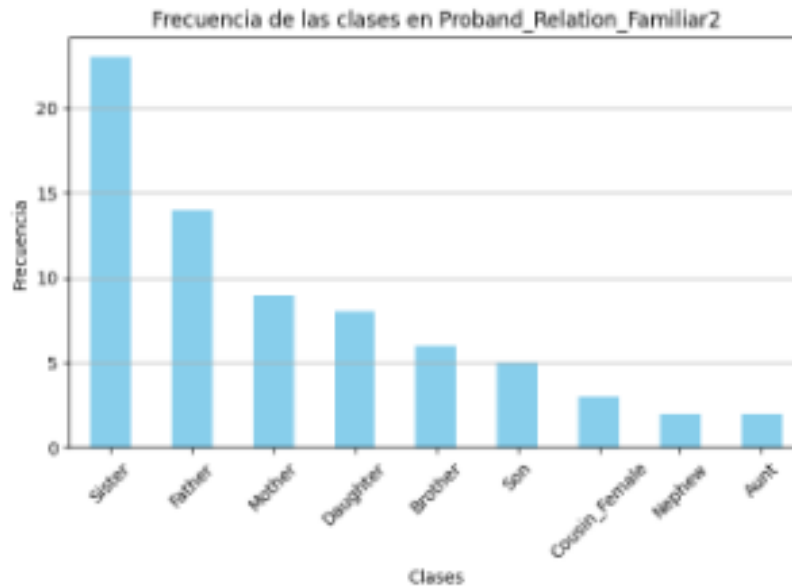


Figura 6 Grafico de frecuencias por parentesco familiar 2

- Este hallazgo sugiere una fuerte influencia genética en la manifestación de trastornos psiquiátricos dentro de las familias.

Modelos Generados y Elección del Modelo Final

Durante el proyecto, desarrollamos varios modelos de aprendizaje automático para predecir la probabilidad de desarrollar trastorno bipolar en pacientes con antecedentes familiares de trastornos psiquiátricos y considerando otros factores socioambientales.

Modelos Generados

A continuación, se describen brevemente algunos de los modelos considerados:

	Modelo/Ensamble	Test_Accuracy	Test_Precision	Test_Recall	Test_F1score	Train_Recall	Tiempo_de_ejecución
0	LR	0.736364	0.736364	1.000000	0.848120	1.000000	1.345279
1	DT	0.736364	0.736364	1.000000	0.848120	1.000000	1.345279
2	Blending NN-DT	0.736364	0.736364	1.000000	0.848120	1.000000	1.383171
3	RF	0.724242	0.732323	0.983333	0.839039	0.998039	2.369419
4	NN	0.702020	0.742299	0.912037	0.815431	0.976233	1.345279

Regresión Logística

Descripción: Un modelo de clasificación lineal que estima la probabilidad de que una instancia pertenezca a una clase particular.

Ventajas: Sencillo de interpretar, rápido de entrenar, adecuado para problemas binarios.

Desempeño: Se utilizó como línea base para comparar con otros modelos. La precisión (73%) fue moderados y el recall (100%) mostro gran proporción de sobreajuste.

Arboles de Decisión (DT)

Descripción: Un modelo basado en árboles que divide los datos en subconjuntos homogéneos basados en características.

Ventajas: Fácil de interpretar, no requiere normalización de datos, captura relaciones no lineales.

Desempeño: Mostró un desempeño similar que la regresión logística en términos de precisión (73%) y recall (100%).

Blending NN-DT (Neural Network-Decision Tree)

Descripción: Es una técnica de ensamblado de modelos que combina los puntos fuertes de redes neuronales artificiales (NN) y árboles de decisión (DT) para mejorar el rendimiento general del modelo. Este enfoque se basa en la idea de que diferentes tipos de modelos pueden capturar diferentes aspectos de los datos y, al combinarlos, se puede obtener un modelo más robusto y preciso.

Ventajas: Capacidad para capturar relaciones complejas y no lineales en los datos, Eficiencia en la detección de patrones intrincados gracias a múltiples capas y neuronas. Fáciles de interpretar y visualizar. No requieren normalización de datos y pueden manejar tanto variables categóricas como continuas.

Desempeño: Mostró una precisión (73%) y recall (100%) similar a los modelos anteriores lo que indica los mismos problemas de sobreajuste ya mencionados.

Bosques Aleatorios (Random Forest)

Descripción: Un conjunto de árboles de decisión entrenados en diferentes subconjuntos de datos y características.

Ventajas: Reduce el sobreajuste, maneja grandes cantidades de datos y características, robusto ante ruido.

Desempeño: Mostró una precisión similar a los modelos anteriores (73%) sin embargo el recall (98%) en comparación con dichos modelos reduce el sobreajuste existente. Al ser esta nuestra métrica principal, este es uno de los modelos candidatos a ser el modelo final.

Red Neuronal (NN)

Descripción: Modelos inspirados en la estructura del cerebro humano, capaces de capturar relaciones complejas y no lineales.

Ventajas: Gran capacidad de modelado, adecuado para grandes volúmenes de datos y características complejas.

Desempeño: Presentaron una mejor precisión que el modelo de Random Forest (94%) sin embargo el recall(91%) fue significativamente más bajo.

Elección del modelo final

Modelo Final Elegido: Bosques Aleatorios (Random Forest)

Las razones del equipo para la elección del modelo de Bosques aleatorios se resumen en los siguientes puntos:

1. Rendimiento General:

- El modelo de bosques aleatorios ofreció un balance óptimo entre precisión, recall y robustez frente al sobreajuste. Superó consistentemente a otros modelos en las métricas de desempeño clave elegidas por el equipo al inicio del proyecto.

2. Manejo de Datos Complejos:

- Su capacidad para manejar grandes volúmenes de datos y características complejas lo hizo adecuado para el conjunto de datos utilizado, que incluye múltiples variables demográficas.

3. Robustez y Flexibilidad:

- Los bosques aleatorios son menos susceptibles al ruido y a las variaciones en los datos, lo que los hace ideales para aplicaciones en entornos clínicos donde los datos pueden ser inconsistentes.

4. Interpretabilidad Relativa:

- Aunque no tan interpretables como los árboles de decisión individuales, los bosques aleatorios permiten la extracción de importancia de características, proporcionando perspectivas valiosas sobre las variables más influyentes.

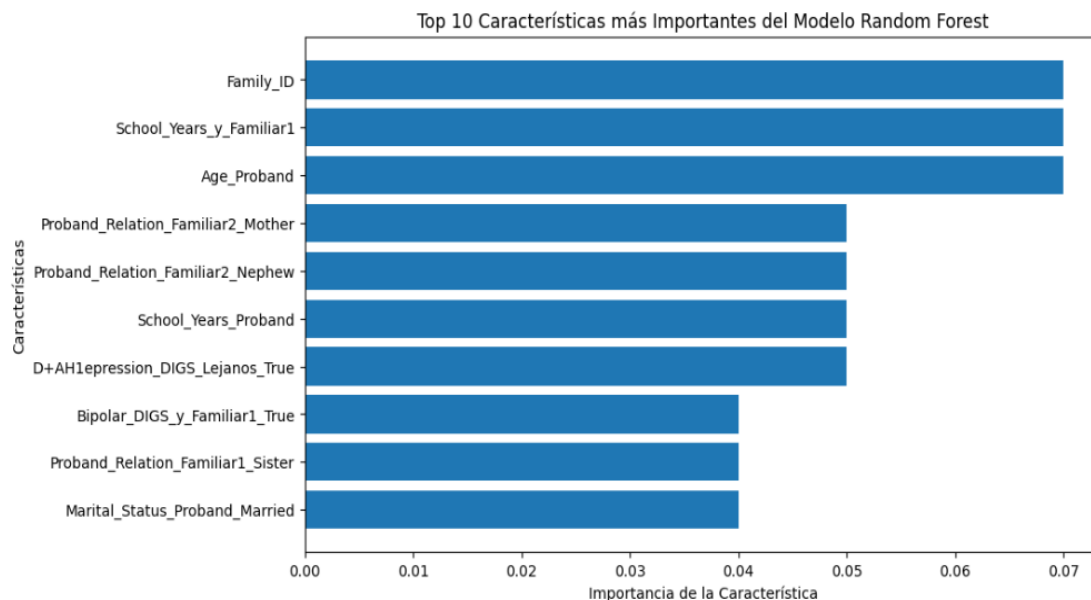


Figura 7 Top 10 de características más importantes del modelo final

5. Tiempo de Entrenamiento:

- Ofrecieron un buen compromiso entre tiempo de entrenamiento y precisión. que, aunque ligeramente superior al resto de los modelos, consideramos que es un poco más eficiente que las redes neuronales en este aspecto dado los resultados de la métrica de recall.

Desempeño del Modelo Final

- **Precisión: 73%**
- **Recall: 98%**
- **F1 Score: 83%**

Estos resultados reflejan la capacidad del modelo para identificar correctamente los casos de riesgo de desarrollar trastorno bipolar, minimizando tanto los falsos positivos como los falsos negativos. Creemos pues que estas métricas combinadas demuestran que el modelo Random Forest ofrece un rendimiento robusto y fiable, lo que lo convierte en una opción excelente para nuestro proyecto.

Recomendaciones Clave para Implementar la Solución

Recomendaciones Específicas

A continuación, se presenta una lista de recomendaciones específicas, recursos requeridos y plazos sugeridos para la implementación de nuestro modelo en un ambiente de producción.

Validación Adicional

- **Descripción:** Realizar validaciones adicionales con conjuntos de datos externos para garantizar la robustez y la capacidad de generalización del modelo. Esto implica la inclusión de datos de múltiples fuentes y la verificación de la consistencia de los resultados obtenidos.
- **Acciones Específicas:**
 - Identificar y obtener conjuntos de datos externos relevantes.
 - Ejecutar el modelo con estos nuevos datos y comparar los resultados con los obtenidos durante la fase de desarrollo.
 - Ajustar el modelo según sea necesario para mejorar su rendimiento y robustez.
- **Plazo:** 1-2 meses.

Monitorización Continua

- **Descripción:** Implementar un sistema de monitorización continua para evaluar el rendimiento del modelo en tiempo real. Esto permitirá detectar cualquier degradación en las métricas clave y tomar medidas correctivas oportunas.
- **Acciones Específicas:**
 - Desarrollar e implementar un sistema de monitorización que registre métricas de rendimiento como recall, precisión y F1 score.
 - Establecer alertas para detectar desviaciones significativas en el rendimiento del modelo.
 - Realizar análisis periódicos de los datos recopilados para identificar posibles mejoras.
- **Plazo:** 2-3 meses para la configuración inicial y luego continua.

Optimización de Hiperparámetros

- **Descripción:** Ajustar continuamente los hiperparámetros del modelo para maximizar su rendimiento. Esto implica un análisis detallado de cómo diferentes configuraciones afectan el desempeño del modelo en la predicción de trastorno bipolar.
- **Acciones Específicas:**
 - Realizar experimentos sistemáticos para probar diferentes configuraciones de hiperparámetros.
 - Utilizar técnicas de optimización como búsqueda en cuadrícula o búsqueda aleatoria.
 - Implementar un sistema automatizado para ajustar y validar los hiperparámetros de manera continua.
- **Plazo:** Proceso continuo, con revisiones mensuales.

Actualización Periódica del Modelo

- **Descripción:** Planificar actualizaciones periódicas del modelo con nuevos datos para mantener su relevancia y precisión. Esto es crucial para asegurar que el modelo se mantenga actualizado con la información más reciente disponible.

- **Acciones Específicas:**
 - Establecer un cronograma para la recolección de nuevos datos y la actualización del modelo.
 - Reentrenar el modelo con los datos nuevos y realizar validaciones exhaustivas.
 - Implementar las nuevas versiones del modelo en el entorno de producción de manera controlada.
- **Plazo:** Cada 3-6 meses.

Estandarización del Número de Familiares Encuestados

- **Descripción:** Establecer un protocolo claro para la recolección de datos a través de encuestas con pacientes y familiares, asegurando la consistencia en la información recopilada.
- **Acciones Específicas:**
 - Definir el número estándar de familiares que deben ser encuestados por cada paciente.
 - Desarrollar un protocolo de encuestas detallado que incluya preguntas específicas y métodos de recolección de datos.
 - Capacitar al personal encargado de realizar las encuestas para asegurar la calidad y **consistencia de los datos**.
- **Plazo:** 1 mes.

Capacitación del Personal

- **Descripción:** Proporcionar capacitación específica sobre la realización de encuestas, la introducción de datos y el tratamiento de la información, asegurando que todo el personal involucrado esté preparado para utilizar el modelo.
- **Acciones Específicas:**
 - Desarrollar programas de capacitación detallados que cubran aspectos técnicos y prácticos.
 - Realizar sesiones de entrenamiento para todo el personal médico y técnico involucrado en el proyecto.
 - Proporcionar materiales de apoyo y recursos continuos para facilitar el aprendizaje continuo.
- **Plazo:** 1-2 meses.

Infraestructura Tecnológica

- **Descripción:** Garantizar que la infraestructura tecnológica esté preparada para soportar la implementación del modelo en producción. Esto incluye la capacidad de procesamiento y almacenamiento de datos, así como la integración con los sistemas existentes.
- **Acciones Específicas:**
 - Seleccionar y configurar la plataforma de servicios en la nube más adecuada (preferentemente AWS).
 - Implementar sistemas de almacenamiento y procesamiento de datos escalables.
 - Asegurar la integración fluida del modelo con los sistemas médicos existentes.
- **Plazo:** 2-3 meses para la configuración inicial.

Políticas de Privacidad y Seguridad

- **Descripción:** Implementar medidas robustas de protección de datos para asegurar la confidencialidad, especialmente en un contexto médico.
- **Acciones Específicas:**
 - Desarrollar y documentar políticas de privacidad y seguridad de datos.
 - Implementar tecnologías de cifrado y control de acceso para proteger los datos sensibles.
 - Realizar auditorías periódicas para asegurar el cumplimiento de las políticas de privacidad y seguridad.
- **Plazo:** 1-2 meses para la configuración inicial y luego continua.

Comunicación Efectiva

- **Descripción:** Establecer canales claros de comunicación para compartir actualizaciones, resolver inquietudes y recopilar retroalimentación de todas las partes interesadas.
- **Acciones Específicas:**
 - Crear un plan de comunicación detallado que incluya todos los stakeholders (equipo de desarrollo, médicos, encuestadores, pacientes, familiares).
 - Utilizar plataformas de comunicación colaborativas (como Slack o Microsoft Teams) para mantener a todos informados y coordinados.
 - Organizar reuniones regulares para revisar el progreso del proyecto y abordar cualquier problema emergente.
- **Plazo:** Proceso continuo.

Recursos Necesarios

Esta es una lista de los recursos que consideramos necesarios para la implementación de las recomendaciones anteriores

Datos de Entrenamiento y Validación

Acceso a bases de datos relevantes y actualizadas que contengan información detallada sobre antecedentes familiares y diagnósticos de trastornos psiquiátricos.

Equipo Técnico

Desarrolladores de machine learning para modelado y optimización.

Ingenieros de datos para la recolección, limpieza y procesamiento de datos.

Personal de TI para la gestión de infraestructura y seguridad.

Infraestructura Tecnológica

Servicios en la nube, preferentemente AWS, debido a su facilidad de uso, escalabilidad y amplio soporte para servicios de machine learning.

Sistemas de almacenamiento y procesamiento de datos escalables y seguros.

Capacitación

Programas de capacitación para el personal médico y técnico involucrado en el proyecto.

Herramientas de Monitorización:

Software y sistemas de monitorización en tiempo real para evaluar el rendimiento del modelo.

Análisis Costo-Beneficio

Costos

A continuación, se muestra un análisis detallado de los costos incurridos para el desarrollo de este proyecto.

Recolección de los datos

En primera instancia tenemos los costos de adquisición de datos, los cuales incluyen la recolección de todos y cada uno de los datos, como tal el instituto HOB fue quien se encargó de recolectar uno a uno cada uno de estos datos. Sin embargo, aquí depende mucho de la forma en que se requieran recolectar estos datos, como tal fue a través de la aplicación de encuestas, por ende, se debe incluir el número de personas que aplicaron estas encuestas y el salario que estas percibían.

Infraestructura tecnológica

Este tipo de costos van relacionados a toda la tecnología que se va a utilizar para el desarrollo e implementación del proyecto, lo cual puede incluir el uso de servidores o servicios de computación en la nube, licencias de software específico para la realización del tratamiento y análisis de los datos de igual forma se debe considerar el hardware, costos de equipos de cómputo especializados que se hayan ocupado.

Para nuestro caso en específico y por la naturaleza de los datos y su volumen se utilizaron herramientas de software de uso libre, como lo fue el uso de Google Colab, sin embargo, sí se quisiera expandir este proyecto a un mayor volumen de datos, sería necesario evaluar las herramientas que pudieran dar el ancho de banda para el entrenamiento del modelo y por ende todas las licencias necesarias y costos incurridos.

Desarrollo del modelo

En esta parte se deben considerar todos los costos relacionados a las personas que trabajaran en el desarrollo del modelo de IA, el tiempo que les tomará ejecutarlo e implementarlo.

Para fines de este proyecto este costo es nulo ya que se realiza como parte de una actividad académica y como una colaboración de empresa institución académica.

Implementación y capacitación

Para esta parte se deben incluir todos aquellos costos incurridos a la hora de capacitar a las personas que harán uso del modelo en producción para los fines que este fue creado.

Para este caso práctico estos costos serían nulos también ya que este proyecto es académico, y la capacitación del personal que hará uso del modelo es a través de una presentación en conjunto con la institución académica.

Tabla 1 Tabla de costos estimados

Concepto	Costo estimado	Justificación
Adquisición de datos	\$10,000/anual	La adquisición de datos es fundamental para entrenar el modelo de IA. Se requiere una gran cantidad de datos precisos de encuestas a pacientes con trastorno bipolar y personas sin este padecimiento, además de antecedentes familiares y factores socioeconómicos. Estos datos son cruciales para identificar patrones y relaciones que ayuden a predecir el riesgo de desarrollar trastorno bipolar, mejorando la detección temprana y el manejo de esta condición.
Infraestructura tecnológica	\$20,000/anual	El desarrollo y la implementación del modelo de IA requieren una infraestructura tecnológica robusta, incluyendo servidores para almacenamiento y procesamiento de datos y software especializado. Esta infraestructura permite manejar grandes volúmenes de datos y ejecutar algoritmos complejos, asegurando la eficiencia y precisión del modelo para proporcionar una herramienta efectiva en el diagnóstico temprano y tratamiento del trastorno bipolar.
Desarrollo del modelo	\$15,000/anual	El desarrollo del modelo de IA es el núcleo del proyecto, involucrando investigación, diseño, programación y pruebas. Expertos en IA y salud mental crearán un modelo que analice datos recolectados y prediga el riesgo de trastorno bipolar. Este modelo será entrenado y validado para asegurar su precisión y fiabilidad, con el objetivo de integrarse en el entorno clínico y asistir a los profesionales de salud en mejorar la atención al paciente.
Implementación y capacitación	\$5,000/anual	La implementación del modelo en el entorno clínico y la capacitación de los profesionales de salud son cruciales. Esta fase incluye la integración del modelo en los sistemas del hospital y la formación del personal en su uso, asegurando que comprendan cómo interpretar y utilizar los resultados. Una implementación efectiva mejorará significativamente la detección temprana del trastorno bipolar y permitirá una intervención oportuna.
Mantenimiento y actualización	\$2,000/anual	El mantenimiento y actualización continuos del modelo de IA son esenciales para asegurar su eficacia a largo plazo. A medida que se recopilan más datos y se realizan nuevas investigaciones, el modelo necesitará ser actualizado para mejorar su precisión. Además, el mantenimiento técnico asegura el correcto funcionamiento del sistema. Esta inversión anual garantiza que el modelo permanezca actualizado y relevante, apoyando continuamente a los profesionales de salud en el diagnóstico y tratamiento del trastorno bipolar.

Beneficios potenciales

Beneficios cuantificables

A continuación, se describen los beneficios cuantificables obtenidos en el desarrollo de este proyecto:

- Mejora en la precisión de diagnóstico de trastorno bipolar
- Reducción de costos en atención médica a la hora de identificar y tratar este tipo de padecimiento
- Aumento en la eficiencia operativa al optimizar recursos médicos
- Reducción en los costos de diagnóstico de padecimientos de trastorno bipolar

Beneficios intangibles

A continuación, se describen los beneficios intangibles obtenidos en el desarrollo del proyecto:

- Mejora en la experiencia del paciente a la hora de recibir diagnósticos más precisos y tratamiento oportuno
- Mejora en la toma de decisiones clínicas basadas en el uso de herramientas tecnológicas de vanguardia como lo es la IA
- Mejora en la precisión de diagnóstico de trastorno bipolar
- Oportunidades de investigación adicionales al contar con datos estructurados y analizados
- Promoción de la salud mental a través de la disminución de costos y promoción de la salud mental usando las nuevas tecnologías.

Riesgos y Desafíos de la Solución

Identificación y clasificación de Riesgos

Durante la realización del proyecto, se han identificado los siguientes riesgos y desafíos para el modelo de ML.

Tabla 2 Tabla de riesgos y desafíos

Riesgo/Desafíos	Descripción	Tipo
Requiere estandarizar la base de datos de pacientes	La base de datos propuesta de pacientes y sus familias cuenta con una gran cantidad de registros, sin embargo, es importante asegurarse de no llenar erróneamente la base.	Datos
Requiere cuidar la calidad de datos	Debido a la información sensible de los pacientes, requiere evitar la interpolación de registros y evitar la eliminación de registros.	Prueba/Confianza
Datos desbalanceados	Algunos registros de enfermedades tienen a hacia una clase negativa, por lo que es importante alimentar los registros con diagnósticos positivos para evitar sesgos en la clasificación	Datos

Mantener privacidad de los pacientes	La base de datos propuesta no requiere se llenarla con datos privados como nombre, dirección, cuentas o teléfonos. Por lo que los datos serán anónimos	Cumplimiento
Exceso de variables categóricas	La base actual contiene una gran cantidad de variables de tipo categóricas, acción que puede afectar negativamente el entrenamiento del modelo.	Datos
Requiere mayor cantidad de registros	La base de pacientes refinada cuenta con una cantidad que no supera los 70 registros, por lo que se requiere más registros que permitan un mejor entrenamiento.	Prueba/Confianza
Salidas incorrectas	Debido a la cantidad de registros se corre el riesgo que el modelo prediga falsos negativos/falsos positivos.	Prueba
Requiere tener atención en el consumo de recursos en la nube	Debido a que las computadoras de AWS generan un costo por procesamiento, es importante monitorizar los recursos que se consumen.	Cumplimiento

Mitigación de Riesgos

Con el objetivo de garantizar el correcto funcionamiento del modelo y de la información de los registros se recomienda realizar las siguientes estrategias:

Tabla 3 Tabla de mitigación de riesgos

Estrategia	Descripción
Buscar nuevos pacientes o alimentar los registros actuales con más familiares	Debido a la falta de registros se recomienda alimentar la base de datos con nuevos registros de familiares/pacientes de diagnóstico positivo que den más información que aprender al modelo. De esta manera se puede evitar el entrenar con datos desbalanceados y mejorar el rendimiento del modelo.
Asignar nuevas variables de tipo numéricas a la base de pacientes	Se ha detectado que la mayoría de las variables usadas para el entrenamiento son del tipo categóricas, con el objetivo de maximizar la generalización de la información.

Para mejorar el procesamiento de información se propone una nueva estructura de base de datos	Para poder trabajar con el modelo de ML, se ha refinado la base original manteniendo únicamente las variables de los familiares más cercanos y una simplificación de familiares extras.
Configurar maquinas AWS para consumir recursos sobre demanda del usuario.	Para mejorar el consumo de recursos de AWS se propone configurar las máquinas para consumir únicamente cuando un usuario solicite información del modelo.

Conclusión

Para concluir, este proyecto se centra en investigar la prevalencia del trastorno bipolar en familias con historial de trastornos psiquiátricos, analizando los factores socioeconómicos asociados. Hemos explorado la relación entre pacientes diagnosticados con trastorno bipolar y otros familiares con condiciones psiquiátricas, con el propósito de comprender mejor el riesgo intrafamiliar y sus implicaciones. En el ámbito de la medicina moderna, y particularmente en el campo de la salud mental, existe una creciente necesidad de mejorar los métodos de diagnóstico y tratamiento de los trastornos psiquiátricos. El trastorno bipolar, dada su considerable repercusión en la vida de los afectados y sus familias, requiere una identificación y tratamiento tempranos para mejorar la calidad de vida de los pacientes y aliviar la carga sobre el sistema de salud.

Nuestra colaboración con el Hospital Universitario UANL y el Instituto HOOB se enfoca en mejorar la detección temprana y el manejo del trastorno bipolar. Proponemos ofrecer una atención más efectiva y oportuna para los pacientes y sus familias, lo cual es esencial en el actual contexto de salud pública.

Por lo tanto, hacemos un llamado a los tomadores de decisiones para que respalden esta iniciativa. Invertir en la investigación y desarrollo de métodos avanzados para la detección temprana y tratamiento del trastorno bipolar no solo mejorará significativamente la calidad de vida de los pacientes, sino que también optimizará el uso de los recursos del sistema de salud, generando un impacto positivo a largo plazo en la comunidad. Su apoyo es fundamental para alcanzar estos objetivos y avanzar hacia un cuidado integral de la salud mental.

Referencias

- Angst, J. (2006). Do many patients with depression suffer from bipolar disorder? Canadian Journal of Psychiatry. *Revue Canadienne de Psychiatrie*, 51(1), 3–5.
<https://doi.org/10.1177/070674370605100102>
- Drancourt, N., Etain, B., Lajnef, M., Henry, C., Raust, A., Cochet, B., Mathieu, F., Gard, S., MBailara, K., Zanouy, L., Kahn, J. P., Cohen, R. F., Wajsbrot-Elgrabli, O., Leboyer, M., Scott, J., & Bellivier, F. (2013). Duration of untreated bipolar disorder: missed opportunities on the long road to optimal treatment. *Acta Psychiatrica Scandinavica*, 127(2), 136–144.
<https://doi.org/10.1111/j.1600-0447.2012.01917.x>

- Montes, J, 30. (2015, marzo 30). *La mitad de las personas con trastorno bipolar no están diagnosticadas*. Adamed TV. <https://www.adamedtv.com/trastorno-bipolar/la-mitad-de-las-personas-con-trastorno-bipolar-no-estan-diagnosticadas/>
- McCraw, S., Parker, G., Graham, R., Synnott, H., & Mitchell, P. B. (2014). The duration of undiagnosed bipolar disorder: Effect on outcomes and treatment response. *Journal of Affective Disorders*, 168, 422–429. <https://doi.org/10.1016/j.jad.2014.07.025>
- Parker, G., McCraw, S., Hadzi-Pavlovic, D., & Fletcher, K. (2013). Costs of the principal mood disorders: A study of comparative direct and indirect costs incurred by those with bipolar I, bipolar II and unipolar disorders. *Journal of Affective Disorders*, 149(1–3), 46–55. <https://doi.org/10.1016/j.jad.2012.10.002>
- Post, R. M., Leverich, G. S., Kupka, R. W., Keck, P. E., Jr, McElroy, S. L., Altshuler, L. L., Frye, M. A., Luckenbaugh, D. A., Rowe, M., Grunze, H., Suppes, T., & Nolen, W. A. (2010). Early-onset bipolar disorder and treatment delay are risk factors for poor outcome in adulthood. *The journal of clinical psychiatry*, 71(07), 864–872. <https://doi.org/10.4088/jcp.08m04994ye>