

Prevalencia de trastornos psiquiátricos en familiares de pacientes con trastorno bipolar

Presentan:

- Luis Alfonso Sabanero Esquivel
- Carlos Daniel Mendoza Paz
- Guillermo Alfonso Muñoz Hermosillo

Asesora: Grettel Barceló Alonso

Patrocinador: Instituto HOOB



Patrocinador del proyecto

El Instituto Hoob es una institución dedicada a la investigación y promoción del conocimiento sobre Salud Mental. En colaboración con el Tecnológico de Monterrey, hemos obtenido acceso a una base de datos de pacientes con trastorno bipolar para el proyecto "Prevalencia de trastornos psiquiátricos en familiares de pacientes con trastorno bipolar".

Misión

Ofrecer a la comunidad atención accesible, integral y de alta calidad para los problemas de conducta y ansiedad.

Visión

Realizar proyectos de investigación y promover el conocimiento sobre los problemas de Salud Mental, desarrollados con conocimiento científico, tanto a profesionales de la salud como al público general.

Agradecemos sinceramente al Instituto Hoob por su invaluable apoyo, acceso a la información y el continuo respaldo que nos permitieron completar este proyecto con éxito.

Definición del problema

En el área de la medicina moderna, especialmente en el ámbito de la salud mental, se observó una necesidad creciente en la mejora de los métodos de diagnóstico y tratamiento de los trastornos psiquiátricos.

El trastorno bipolar es una condición mental que tiene un profundo impacto en la vida de los individuos y sus familias, provocando cambios extremos en el estado de ánimo, niveles de energía y capacidad para realizar actividades diarias.

La identificación precisa y el tratamiento temprano de esta enfermedad a través del uso de la tecnología moderna juegan un papel crucial en la calidad de vida de las personas.

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.

Metodología Utilizada

En este proyecto, se utilizó la metodología CRISP-DM (Cross Industry Standard Process for Data Mining) para estructurar el proyecto. Esta metodología se compone de seis fases:

Comprensión del Negocio

- El objetivo principal del proyecto es mejorar la detección temprana y el manejo del trastorno bipolar. Así como determinar la prevalencia del trastorno bipolar en ambientes familiares variados.

Comprensión de los Datos

- El primer paso del EDA consistió en la importación y agrupación de los datos de los pacientes con sus familiares. Dicho archivo contiene información detallada sociodemográfica y clínica de pacientes y sus familiares diagnosticados con bipolaridad u otros trastornos psiquiátricos.

Preparación de los Datos

- Se incluyeron varias tareas esenciales para asegurar la calidad y consistencia de los datos antes del análisis: manejo de valores nulos, agrupamiento de los datos, creación de nuevas variables.

Modelado

- Se creó un nuevo archivo llamado "final_dataframe.xlsx". Este archivo se utiliza para extraer datos y entrenar los modelos siguientes. En base a este archivo se llevan a cabo el entrenamiento de diferentes modelos de IA.

Evaluación

- Se evaluaron los modelos para validar que cumplan con los objetivos del negocio y proporcionen valor agregado al modelo. Se presenta la evaluación de la métrica clave definida el Recall, así como otras métricas importantes como la precisión o el F1 Score.

Despliegue

- Los resultados del análisis y modelos desarrollados se integran en el entorno clínico del Instituto HOOB. La implementación efectiva y la capacitación del personal aseguran que los modelos sean utilizados correctamente para mejorar la atención al paciente.

Metodología Base de datos

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.

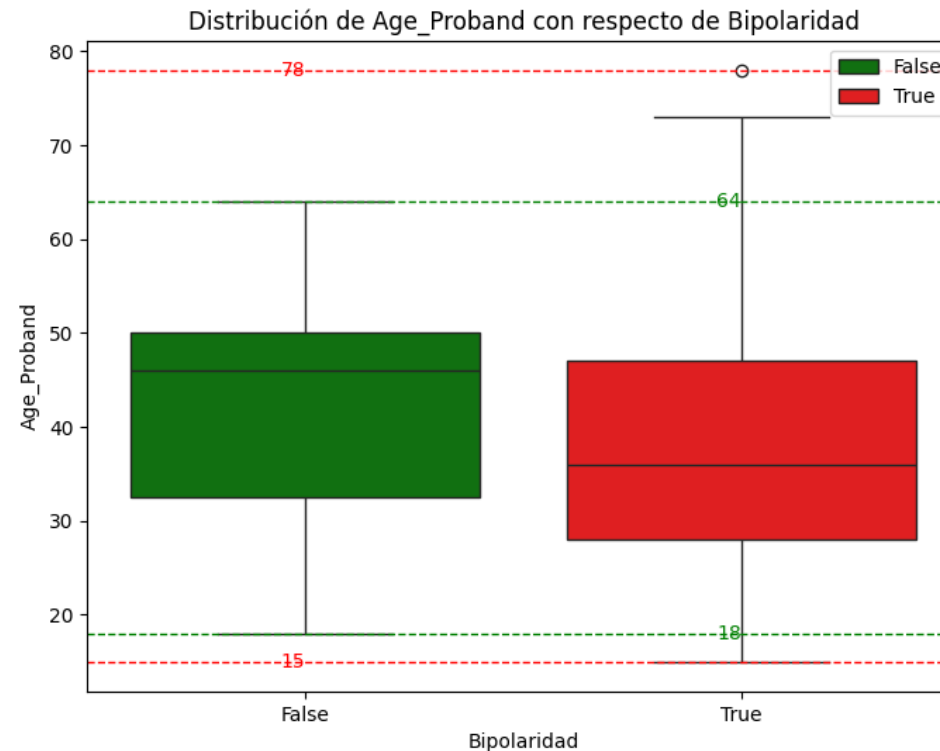
3. Creación de Nuevas Variables: Se crearon nuevas variables derivadas de los datos existentes para proporcionar perspectivas adicionales y enriquecer el análisis. Por ejemplo, se agregaron como base a los 2 familiares con un parentesco más cercano al paciente (probando).

El orden establecido para la creación de estas características fue el siguiente:

- 1. Mamá
- 2. Papá
- 3. Hermanos
- 4. Abuelos
- 5. Otros

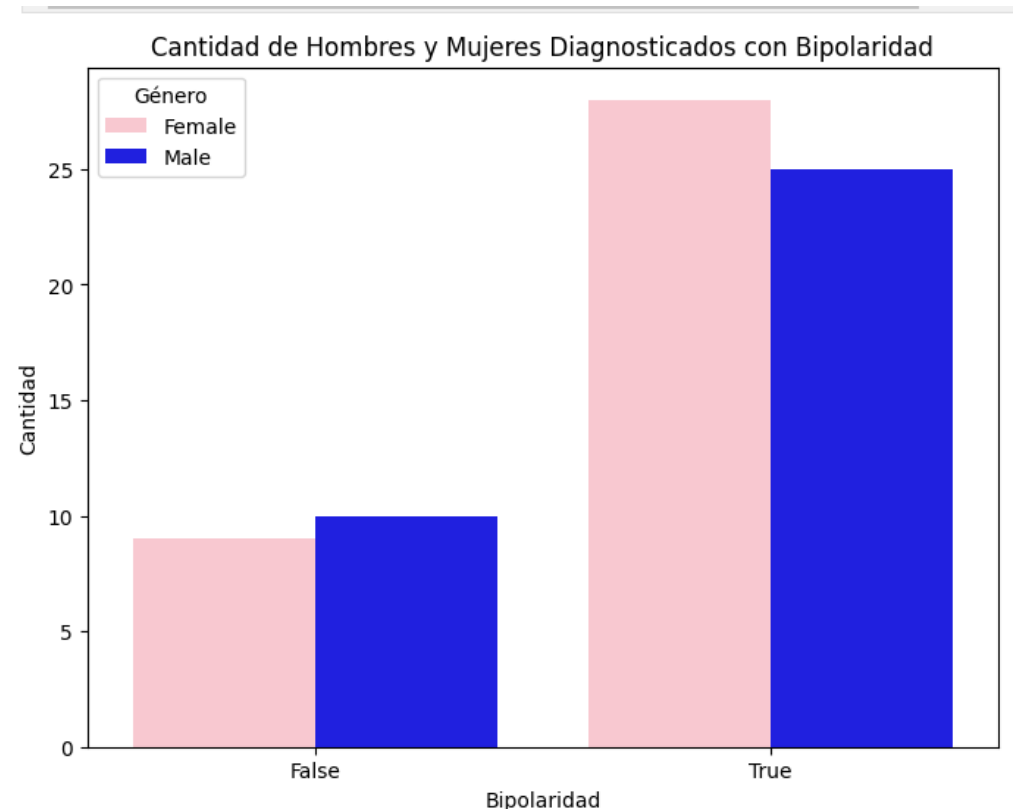
Hallazgos encontrados en el análisis de datos

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



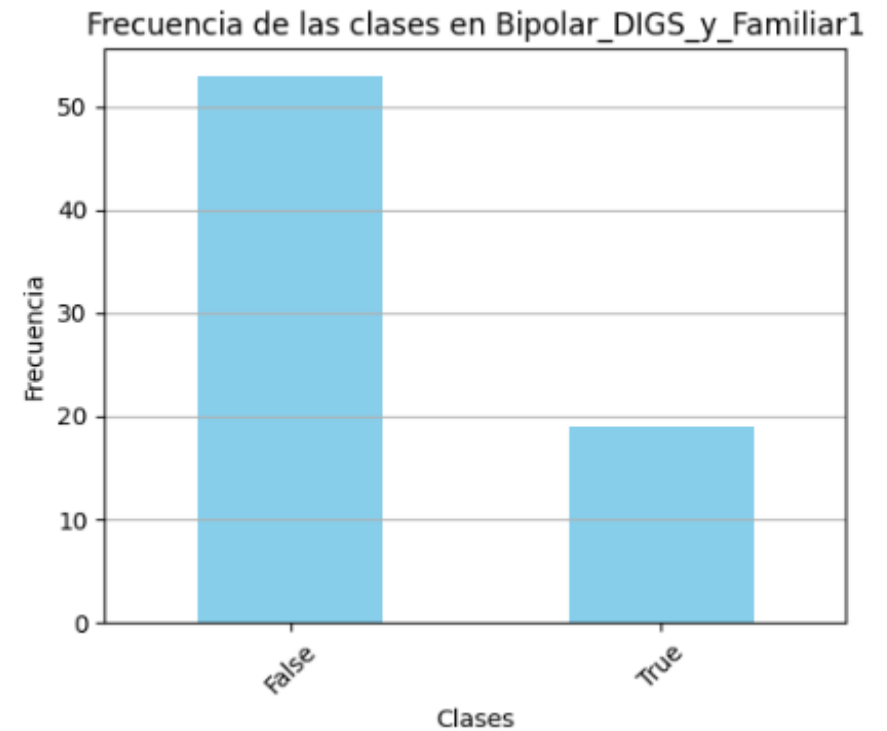
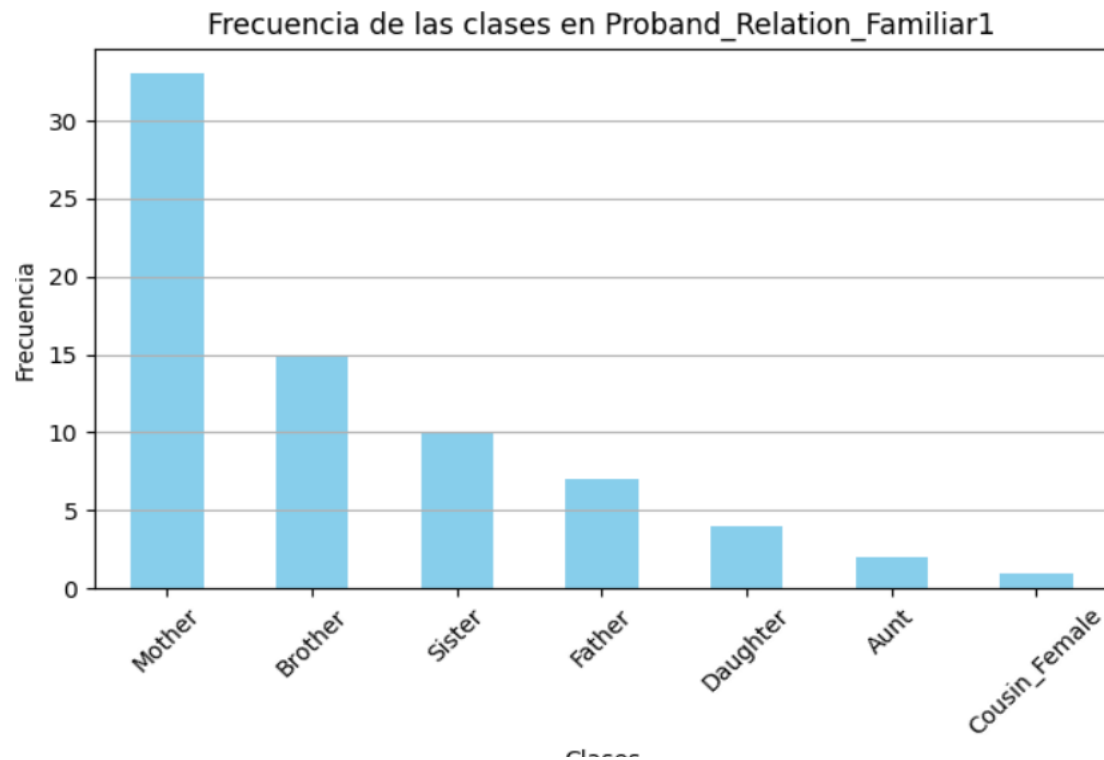
Hallazgos encontrados en el análisis de datos

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



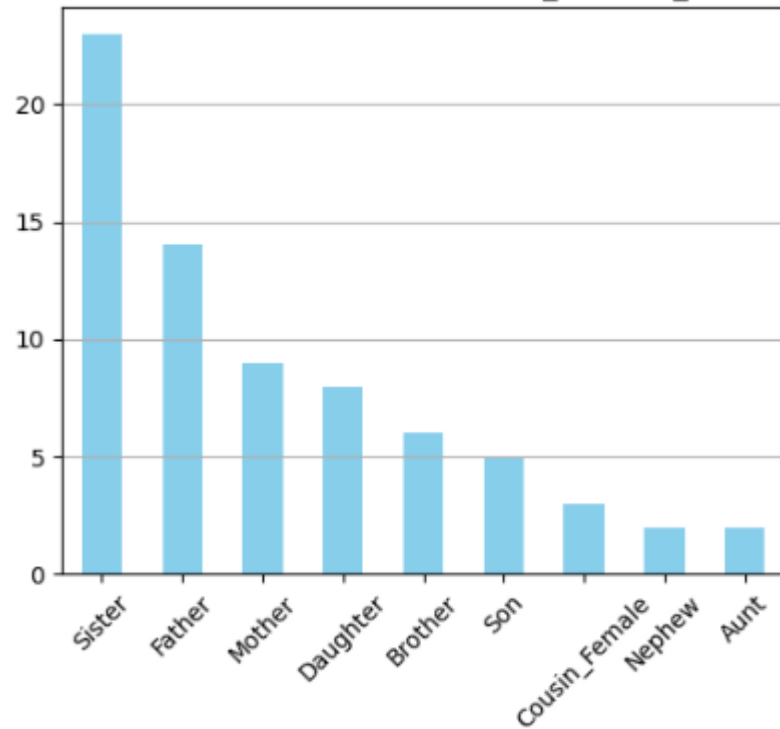
Hallazgos encontrados en el análisis de datos

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).

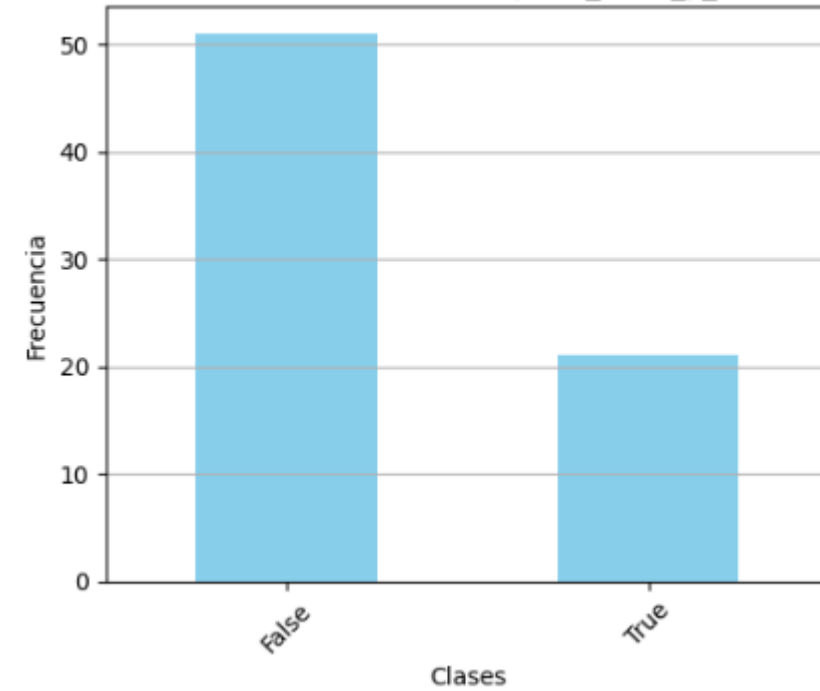


Hallazgos encontrados en el análisis de datos

Frecuencia de las clases en Proband_Relation_Familiar2

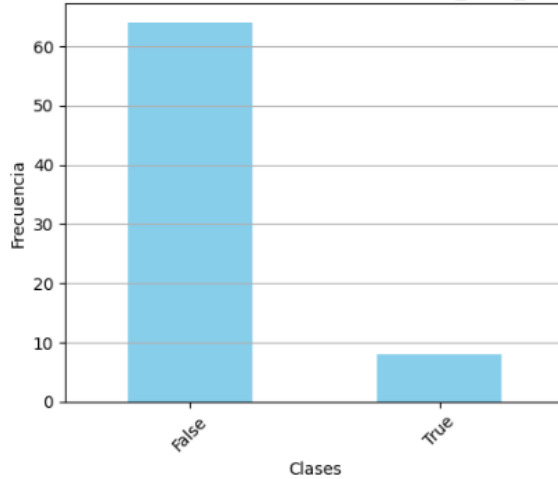


Frecuencia de las clases en Bipolar_DIGS_y_Familiar2

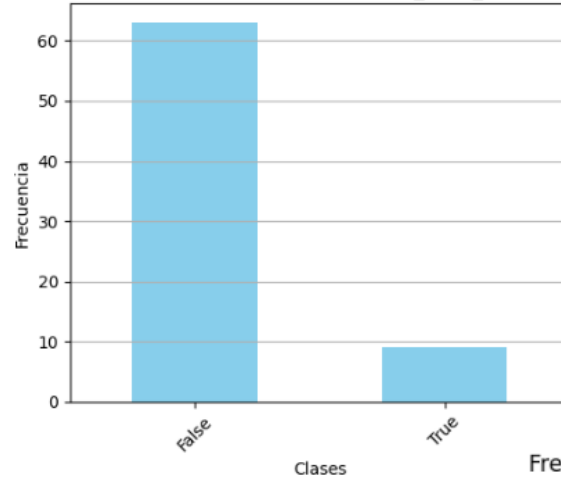


Hallazgos encontrados en el análisis de datos

Frecuencia de las clases en D+AH1epression_DIGS_Cercanos



Frecuencia de las clases en Anxiety_DIGS_Cercanos

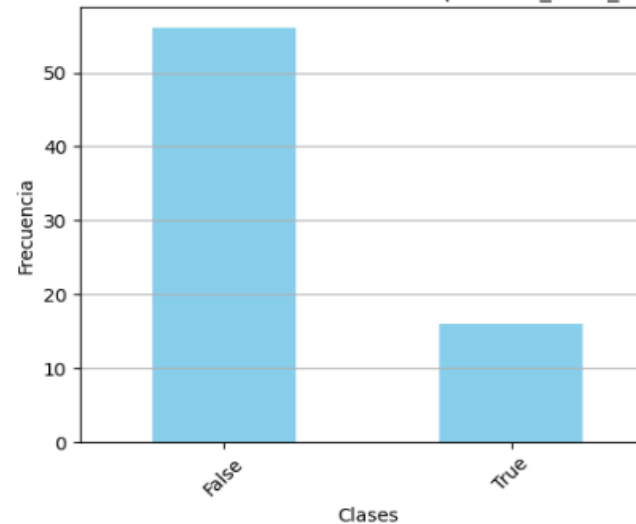


La **depresión** y la **ansiedad** fueron los trastornos más comúnmente diagnosticados en ambos grupos de familiares, seguido por **trastornos de personalidad y abuso de sustancias**.

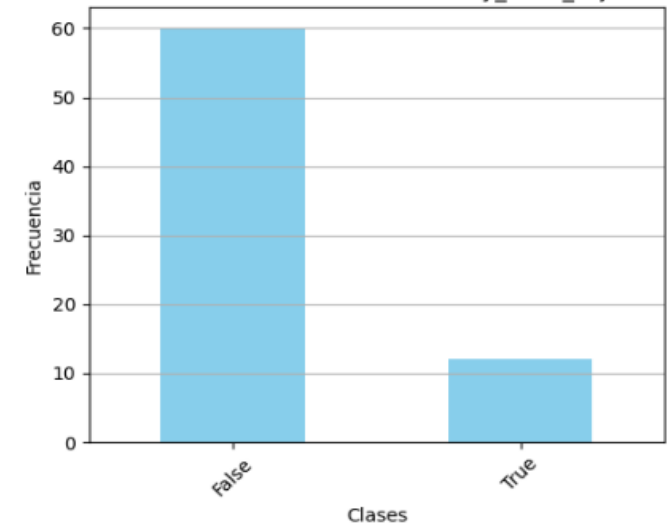
¿Relacional o Causal?

¿Cuál es el nivel de convivencia con el paciente?

Frecuencia de las clases en D+AH1epression_DIGS_Lejanos



Frecuencia de las clases en Anxiety_DIGS_Lejanos




Modelo Final Random Forest

Para cumplir con el objetivo de implementar un modelo de clasificación que apoye al equipo de especialistas se propone como solución el emplear un modelo de tipo 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 este es uno de los más altos y por lo tanto de los mejores modelos para trabajar con la base de datos.



	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

Modelo Final Random Forest

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.

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.

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.

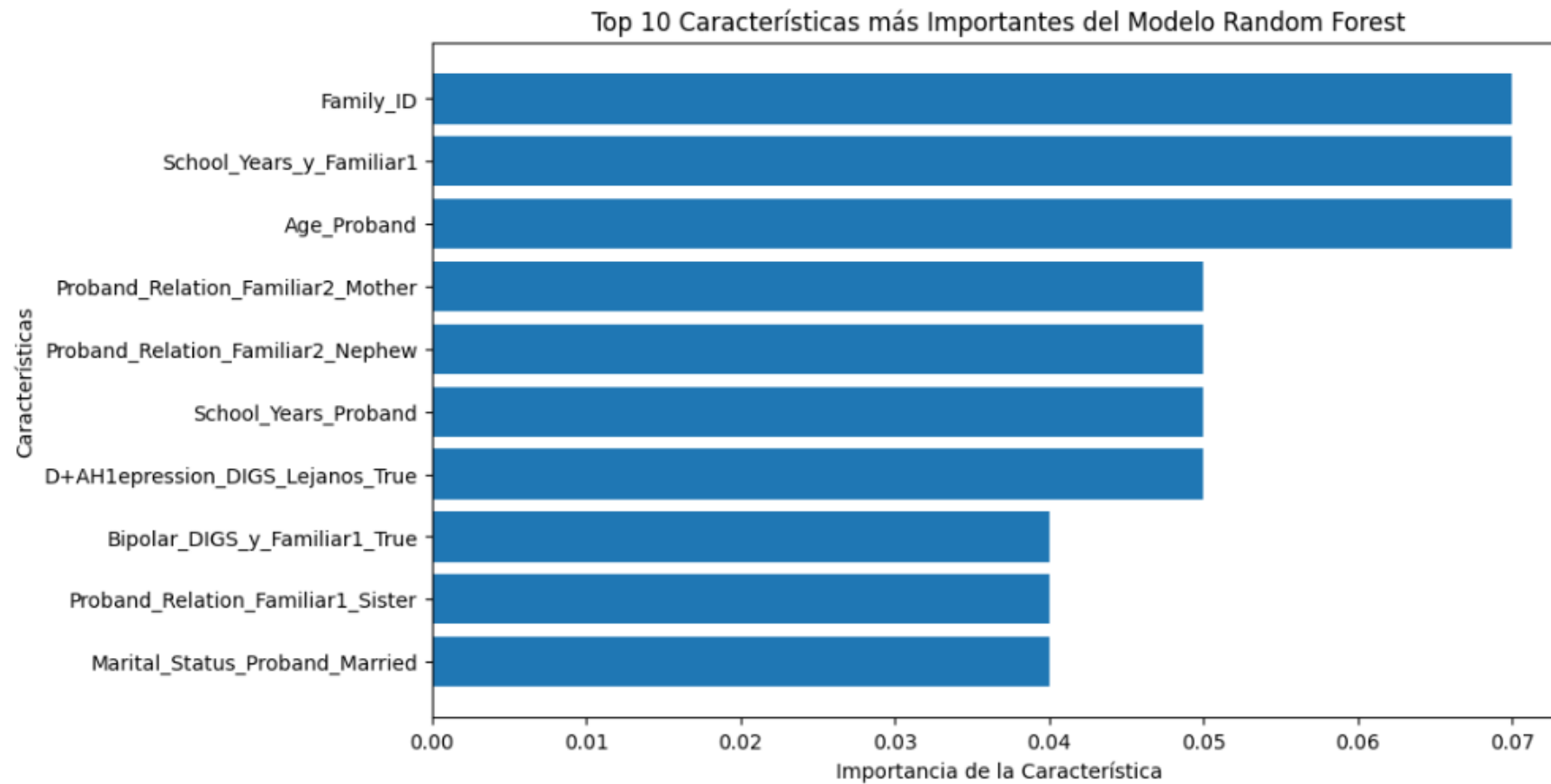
Interpretabilidad Relativa

- 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.

Tiempo de Entrenamiento

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

Características importantes del modelo Random Forest



Beneficios

- 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
- 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
- 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.

Recomendaciones

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.

Recomendaciones

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.

Trabajos futuros



Desarrollar nuevos modelos que analicen otros padecimientos presentes en el estudio de pacientes, basándose en los datos familiares.

Incrementar la cantidad de información en los registros familiares de la base de datos, para que los modelos puedan aprender y generalizar con mayor eficiencia.

Aplicar el estudio utilizando grupos de control para evaluar el rendimiento del modelo y analizar su eficacia.

Conclusiones

En conclusión, este proyecto investiga la prevalencia del trastorno bipolar en familias con historial de trastornos psiquiátricos, explorando los factores socioeconómicos asociados. Nuestra colaboración con el Instituto Hoob busca mejorar la detección temprana y el manejo de esta condición.

Durante el desarrollo del proyecto, enfrentamos desafíos como la limpieza de datos y la definición de la agrupación de familiares según su cercanía sanguínea y trastornos psicológicos. Este proceso nos hizo reflexionar sobre el potencial que representa la investigación en enfermedades mentales en México.

La aplicación de inteligencia artificial en este campo será fundamental para futuras investigaciones del Instituto Hoob, fortaleciendo el cuidado de la salud mental y mejorando continuamente el tratamiento de estas condiciones.

Q & A



Tecnológico
de Monterrey