|  |  |
| --- | --- |
| **Análisis de la Influencia de Factores Demográficos y Sociales en la Salud Mental de Diferentes Grupos Poblacionales** | |
|  | |
| Imagen que contiene Logotipo  Descripción generada automáticamente | **Nombre Estudiante**  MU Ingeniería Computacional y Matemática  Área de Inteligencia Artificial  **Nombre Tutor/a de TF**  Dr. Antonio Sarasa  **Profesor/a responsable de la asignatura**  Juan Andrés Girón Caballero |

**Informe de Seguimiento del Proyecto**

**1. Identificación del Trabajo y Fecha del Informe**

* **Título del Trabajo:** Análisis de la Influencia de Factores Demográficos y Sociales en la Salud Mental de Diferentes Grupos Poblacionales
* **Autor:** Ing. Juan Andrés Girón Caballero
* **Tutor/a:** Dr. Antonio Sarasa
* **Programa:** MU Ingeniería Computacional y Matemática
* **Área:** Inteligencia Artificial
* **Fecha:** 20/04/2025

**2. Descripción del avance del proyecto**

**2.1. Grado de cumplimiento de los objetivos y resultados previstos en el plan de trabajo**

Hasta el momento, se ha logrado:

* Seleccionar y preparar dos datasets principales:
  + Wellcome Global Monitor 2020 (percepciones de salud mental).
  + Mental Health Dataset (diagnóstico clínico real).
* Aplicar técnicas de limpieza, transformación y mapeo de variables categóricas.
* Construir modelos predictivos utilizando Random Forest, Logistic Regression y XGBoost.
* Implementar estrategias de balanceo de clases (SMOTE) para mejorar el rendimiento de los modelos.
* Evaluar el rendimiento de los modelos utilizando métricas de clasificación estándar (accuracy, recall, f1-score).
* Aplicar técnicas de interpretabilidad XAI (SHAP y LIME) para entender los factores que más influyen en las predicciones.

Estos logros cumplen con los objetivos iniciales del proyecto en cuanto a la exploración, modelización y análisis explicativo de los datos.

**2.2. Justificación de los cambios en caso necesario**

Durante el proceso de interpretación con XAI, se detectó que:

* El rendimiento predictivo en el dataset de diagnóstico real era cercano al azar (50% de accuracy), debido a limitaciones en las variables disponibles.
* Se decidió profundizar en el análisis explicativo para identificar posibles mejoras en la selección de variables y ajuste de hiperparámetros.

**Cambio justificado:**  
El enfoque del proyecto ahora incluye un proceso de refinamiento adicional, buscando identificar combinaciones de variables que permitan mejorar el rendimiento del modelo, basados en la información proporcionada por los métodos de interpretabilidad.

Además, el análisis de interpretabilidad aplicado mediante SHAP y LIME reveló que, si bien algunas variables tienen influencia en la predicción de la salud mental, su impacto individual es relativamente moderado.

Esto sugiere que las variables disponibles en los datasets actuales presentan una correlación limitada con el fenómeno objetivo, dificultando la obtención de modelos predictivos altamente precisos.

En consecuencia, se ha decidido mantener activa la fase de modelización, probando combinaciones alternativas de variables, realizando ajustes adicionales de hiperparámetros y evaluando posibles estrategias de ingeniería de características para mejorar el poder predictivo del modelo.

De esta manera, se busca maximizar la utilidad de los datos disponibles y proponer estrategias futuras de mejora basadas en la experiencia obtenida.

**3. Relación de las actividades realizadas**

**3.1. Actividades previstas en el plan de trabajo**

* Selección y procesamiento de datasets.
* Construcción de modelos predictivos básicos.
* Evaluación de métricas de rendimiento.
* Aplicación de técnicas de interpretabilidad (SHAP, LIME).

**3.2. Actividades no previstas y realizadas o programadas**

* Implementación de estrategias de balanceo de clases (SMOTE) para mejorar la capacidad de detección de clases minoritarias.
* Aplicación de ajuste de hiperparámetros en XGBoost para intentar mejorar el recall de la clase minoritaria.
* Análisis de predicciones individuales mediante LIME para validación local de los modelos.

**4. Relación de las desviaciones en la temporización y acciones de mitigación**

**Desviaciones:**

* El bajo rendimiento del modelo sobre diagnóstico real retrasó la fase de consolidación de resultados.

**Acciones de mitigación:**

* Se optó por realizar análisis explicativos detallados mediante técnicas XAI para entender mejor las limitaciones de los modelos.
* Se programó un nuevo bloque de trabajo centrado en el diseño de propuestas de mejora del dataset y de estrategias de predicción.

**Actualización del cronograma:**

* Se amplió el tiempo dedicado a la fase de Interpretabilidad y Análisis de Resultados en tres semanas adicionales.

|  |  |  |  |
| --- | --- | --- | --- |
| Fase / Entregable | Fecha límite | Estado real | Observación |
| Definición del alcance y estructura del proyecto | 6/12/2024 | Cumplido | Definí correctamente el alcance y la estructura general del proyecto. |
| Entrega PAC1 (Plan de Trabajo) | 8/12/2024 | Cumplido | Entregué el plan de trabajo en la fecha estipulada. |
| Selección y obtención del conjunto de datos | 31/12/2024 | Cumplido | Seleccioné y obtuve dos datasets sólidos para el análisis. |
| Preprocesamiento inicial de datos | 20/01/2025 | Cumplido | Realicé la limpieza, mapeo de variables y normalización de los datos. |
| Ingeniería de características y reducción de dimensionalidad | 15/02/2025 | Cumplido | Desarrollé la ingeniería de características y reduje dimensionalidad donde fue necesario. |
| Primera implementación de modelos | 25/02/2025 | Cumplido | Construí los modelos iniciales (Random Forest, Logistic Regression y XGBoost). |
| Entrega PAC2 (Estado intermedio) | 2/3/2025 | Cumplido | Entregué el avance de trabajo conforme al cronograma establecido. |
| Ajuste de hiperparámetros y entrenamiento avanzado | 20/03/2025 | Pendiente | Realice/realizo ajustes de hiperparámetros y balanceo de datos utilizando técnicas como SMOTE. |
| Evaluación de métricas y análisis preliminar | 10/4/2025 | Cumplido | Evalué los modelos mediante métricas de clasificación y realicé comparativas entre ellos. |
| Preparación de documentación intermedia para reporte | 24/04/2025 | Cumplido | Preparé los informes, análisis de resultados y gráficos de apoyo. |
| Entrega PAC3 (Resultados intermedios) | 27/04/2025 | En proceso | Actualmente estoy finalizando los entregables correspondientes al PAC3. |
| Mejora de modelos, análisis de interpretabilidad y refuerzo de conclusiones | 20/05/2025 | Iniciado | He comenzado el proceso de mejora utilizando interpretabilidad XAI (SHAP y LIME). |
| Redacción de la memoria final | 5/6/2025 | Pendiente | Redactaré la memoria final después de concluir PAC3. |
| Entrega PAC4 (Memoria Final) | 11/6/2025 | Pendiente | Programado para ser entregado en la fecha establecida. |
| Preparación presentación virtual | 20/06/2025 | Pendiente | La prepararé tras la finalización de la memoria final. |
| Defensa final | 30/06/2025 | Pendiente | Queda pendiente la asignación de la fecha oficial de defensa. |

**5. Listado de los resultados parciales obtenidos hasta el momento**

**Entregables adjuntos:**

* Dataset limpiado y transformado (df\_clean y df\_model).
* Modelos entrenados: Random Forest, Logistic Regression, XGBoost (tuned).
* Reportes de métricas de clasificación para ambos datasets.
* Gráficos de SHAP (summary plot, waterfall plot) para ambos modelos.
* Explicaciones locales mediante LIME para ejemplos representativos.
* Documento de interpretación y comparativa entre datasets (percepciones vs diagnóstico real).