

Briefing del Proyecto: Análisis de Discriminación en Servicios de Salud Mediante Modelos Predictivos

Introducción:

El proyecto busca determinar si la admisión hospitalaria de los pacientes está influenciada por la compañía de seguros que poseen. Esto permitirá analizar posibles patrones de discriminación en los servicios de salud.

Contexto:

En muchas instituciones de salud, factores económicos y administrativos pueden afectar el acceso y la calidad de los servicios. Este proyecto explora esta problemática desde la perspectiva de los seguros, destacando su impacto en la equidad del sistema.

Relevancia:

Detectar desigualdades para impulsar mejoras en políticas de salud.

Objetivo principal:

Desarrollar un modelo predictivo capaz de identificar patrones que vinculen la admisión de pacientes con su compañía de seguros, evaluando así la posibilidad de discriminación.

Funcionalidades Propuestas

1. Esencial:

- Construcción de un modelo de clasificación básico que prediga la admisión hospitalaria en función de atributos clave.
- Preprocesamiento de datos: limpieza, normalización y manejo de valores faltantes.
- Realización de un Análisis Exploratorio de Datos (EDA) para comprender las relaciones entre las variables.
- Generación de un informe con métricas como precisión, recall, F1-score y matriz de confusión.
- Creación de un dashboard interactivo para visualizar los datos.

2. Medio:

- Desarrollar una interfaz web simple que permita a los usuarios ingresar datos y

obtener predicciones sobre admisión hospitalaria.

- Implementación de técnicas avanzadas de clasificación como ensemble methods (e.g., Random Forest, Gradient Boosting).
- Validación cruzada para evaluar el rendimiento del modelo de manera robusta.
- Ajuste de hiperparámetros utilizando herramientas como GridSearchCV u Optuna.
- Incorporación de métricas avanzadas como AUC-ROC para evaluar la capacidad discriminadora del modelo.
- Visualizaciones detalladas para evidenciar correlaciones y patrones discriminatorios.

3. Avanzado:

- Dockerizar el programa para facilitar su despliegue en diferentes entornos.
- Implementación de técnicas de interpretación de modelos (e.g., SHAP o LIME) para analizar la influencia de las variables en las predicciones.
- Almacenamiento de predicciones en una base de datos para monitoreo continuo y posibles reentrenamientos.

4. Experto:

- Desarrollo de un modelo basado en redes neuronales.
- Implementación de un sistema de MLOps para gestionar el ciclo de vida del modelo (entrenamiento, validación y despliegue continuo).
- Uso de MLFlow para rastrear experimentos, métricas y artefactos del modelo.
- Despliegue del modelo como servicio en la nube.
- Propuestas basadas en hallazgos para mitigar desigualdades y promover la equidad en el sistema de salud.

5. Tecnologías Propuestas

- Lenguajes de Programación: Python para el desarrollo principal.
- Librerías y Frameworks de IA/ML: Scikit-learn, XGBoost, etc
- Herramientas de Análisis de Datos: Pandas, NumPy, Matplotlib, Seaborn.
- Infraestructura y Desarrollo: Jupyter Notebooks para experimentación.
- Git para control de versiones.
- Bases de Datos: MySQL, PostgreSQL para almacenar resultados y registros si es necesario.
- Power BI o Tableau
- Streamlit
- Servicios en la Nube:
- Despliegue en la nube usando recursos gratuitos

6. Datos a Utilizar

Datos

Descripción del Dataset:

- El conjunto de datos consta de 10.000 registros, cada uno de los cuales representa un historial clínico sintético del paciente.
- Incluye varios atributos, como datos demográficos del paciente, condiciones médicas, detalles de admisión y más.
- El conjunto de datos está destinado a un uso educativo y no comercial. Es totalmente sintético y no contiene datos reales de pacientes.

Formato de los Datos:

- Archivos CSV con columnas representativas como: tipo de seguro, tasa de admisión, edad, género y diagnósticos.
- Consideraciones Éticas y de Privacidad:
- Garantizar que no se expongan datos personales sensibles.