Proyecto Final

INTELIGENCIA ARTIFICIAL

Estiven Liscano
Juan Lizcano

Índice

1 Introducción

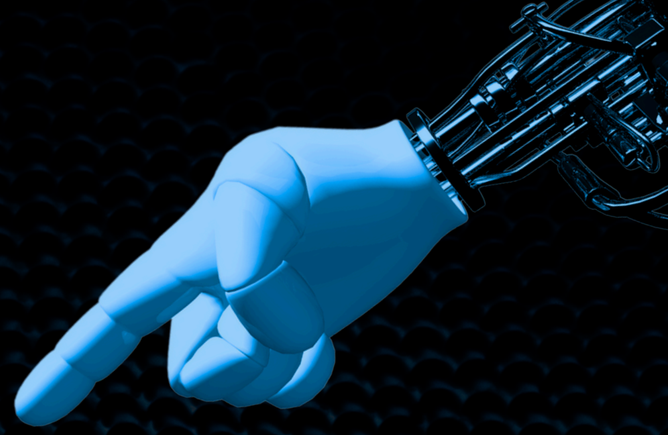
2 Modelos Implementados

3 Justificación de los Modelos

4 Métricas de Evaluación

5 Clonclusiones

6 Proyecto en GitHub



Introducción

Este es un sistema web desarrollado en Django que implementa tres modelos de machine learning diferentes para la predicción de diabetes. El sistema permite comparar el rendimiento de estos modelos de manera visual e interactiva.

Modelos Implementados

1 Árbol de Decisión (Decision Tree)

- Utiliza GridSearchCV para optimización de hiperparámetros
- Parámetros optimizados: profundidad máxima, muestras mínimas para división y hojas
- Ofrece buena interpretabilidad de resultados

2 Naive Bayes (NB)

- Implementa el clasificador Gaussiano de Naive Bayes
- Modelo más simple pero efectivo
- No requiere optimización de hiperparámetros

3 Máquina de Vectores de Soporte (SVM)

- También utiliza GridSearchCV para optimización
- Optimiza parámetros como C, kernel y gamma
- Conocido por su efectividad en problemas de clasificación

Justificación de los Modelos

1

Árbol de Decisión (Decision Tree)

El Árbol de Decisión se seleccionó por su capacidad de manejar tanto datos numéricos como categóricos, y especialmente por su interpretabilidad. En el contexto médico de predicción de diabetes, es crucial que los profesionales de la salud puedan entender cómo se llegó a una determinada predicción, y los árboles de decisión proporcionan un camino claro y visual de la toma de decisiones.

2

Naive Bayes (NB)

Se implementó el clasificador Naive Bayes debido a su eficacia con conjuntos de datos médicos donde las características pueden ser relativamente independientes entre sí. Es particularmente útil cuando trabajamos con múltiples variables predictoras (como niveles de glucosa, presión arterial, edad, etc.), y su simplicidad lo hace computacionalmente eficiente y menos propenso al sobreajuste.

3

Máquina de Vectores de Soporte (SVM)

El SVM se eligió por su robustez en la clasificación binaria (diabético/no diabético) y su capacidad para manejar datos no lineales a través de diferentes kernels. Es especialmente efectivo en espacios de alta dimensionalidad como los datos médicos, donde las relaciones entre variables pueden ser complejas y no lineales.

Métricas de Evaluación

Accuracy (Precisión general)

Mide la proporción total de predicciones correctas. Se incluye para dar una visión general del rendimiento del modelo, aunque en contextos médicos no debe ser la única métrica considerada.

Precision (Precisión específica)

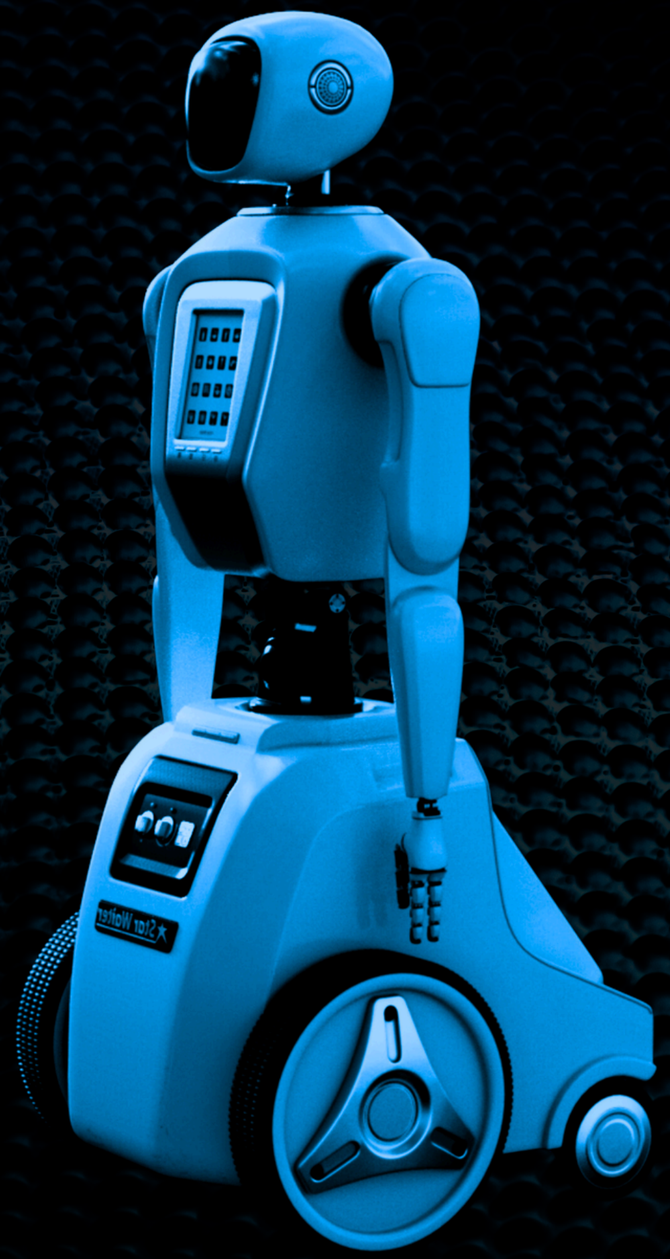
Indica qué proporción de las predicciones positivas fueron correctas. Es crucial en diagnóstico médico para minimizar falsos positivos, evitando así preocupaciones innecesarias en pacientes sanos.

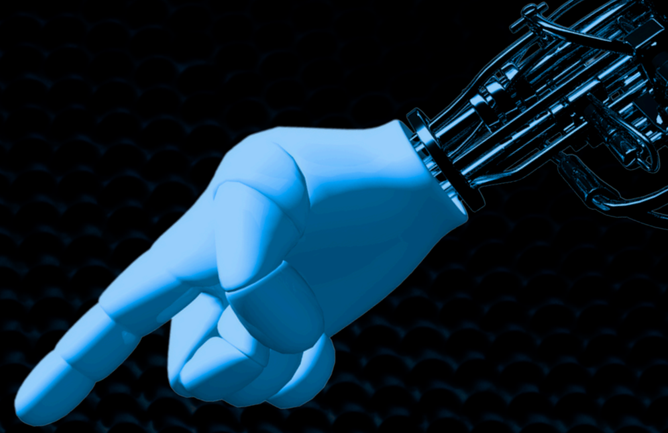
Recall (Sensibilidad)

Mide la capacidad del modelo para identificar correctamente los casos positivos reales. En el contexto de diabetes, es vital para no pasar por alto pacientes que realmente tienen la condición.

F1-Score (Media armónica entre precisión y recall)

Proporciona un balance entre Precision y Recall, siendo especialmente útil en conjuntos de datos médicos donde tanto los falsos positivos como los falsos negativos tienen consecuencias significativas.





Conclusión

Este proyecto demuestra la aplicación práctica de la inteligencia artificial en el campo médico, específicamente en la predicción de diabetes. La implementación de tres modelos diferentes, cada uno con sus propias fortalezas, junto con un sistema de métricas comprehensivo, proporciona una herramienta robusta para el apoyo en la toma de decisiones médicas. La interfaz web intuitiva y las visualizaciones detalladas hacen que la herramienta sea accesible tanto para profesionales de la salud como para investigadores. Los resultados comparativos entre los modelos permiten una evaluación crítica de cada enfoque, mientras que el preprocesamiento cuidadoso de los datos y la optimización de parámetros aseguran la confiabilidad de las predicciones. Este sistema no solo sirve como herramienta de predicción, sino también como plataforma educativa para entender mejor la aplicación de machine learning en el diagnóstico médico.

Proyecto GitHub

Gracias