

9-11-2025

Predicción de Readmisión Hospitalaria por Diabetes

Evaluacion: Segundo Parcial

Modelos

Nombre: Avalos Garcia Juan Antonio

1. INTRODUCCIÓN

1.1 Contexto del Problema

La readmisión hospitalaria en pacientes diabéticos representa un desafío significativo para los sistemas de salud a nivel mundial. Según estudios epidemiológicos, aproximadamente el 20-25% de los pacientes diabéticos son readmitidos dentro de los 30 días posteriores al alta hospitalaria, generando costos adicionales estimados en miles de millones de dólares anualmente.

1.2 Justificación

El desarrollo de sistemas predictivos basados en Machine Learning permite identificar proactivamente a pacientes en riesgo de readmisión, facilitando intervenciones tempranas y personalizadas. Este proyecto aborda esta necesidad mediante la implementación de un modelo predictivo accesible a través de una interfaz web intuitiva.

1.3 Alcance

El sistema desarrollado predice la probabilidad de readmisión hospitalaria en menos de 30 días para pacientes diabéticos, utilizando datos clínicos y demográficos disponibles al momento del alta hospitalaria.

2. OBJETIVOS DEL PROYECTO

2.1 Objetivo General

Desarrollar e implementar un sistema de predicción de readmisión hospitalaria para pacientes diabéticos utilizando técnicas de Machine Learning y una interfaz web accesible.

2.2 Objetivos Específicos

- Realizar un análisis exploratorio completo del dataset de pacientes diabéticos
- Preprocesar y transformar los datos para el modelado predictivo
- Entrenar y evaluar múltiples algoritmos de Machine Learning
- Seleccionar el modelo óptimo basado en métricas de desempeño
- Implementar una aplicación web para predicciones en tiempo real
- Documentar el proceso completo para replicabilidad

3. METODOLOGÍA

3.1 Fuente de Datos

Dataset: diabetic_data.csv

- **Origen:** Base de datos clínicos anonimizados
- **Muestra:** 101,766 encounters hospitalarios
- **Período:** Registros acumulativos
- **Variables:** 50 características clínicas y demográficas

3.2 Variables Clave

Variable Objetivo:

- **readmitted:** Categoría de readmisión (<30, >30, NO)

Variables Predictoras:

- **Demográficas:** edad, género, raza
- **Hospitalización:** tiempo de estancia, tipo de admisión
- **Clínicas:** número de medicaciones, procedimientos, diagnósticos
- **Laboratorio:** glucosa en suero, hemoglobina A1C
- **Tratamiento:** uso de insulina, metformina, cambios en medicación

4. DESARROLLO TÉCNICO

4.1 Preprocesamiento de Datos

4.1.1 Limpieza Inicial

python

```
# Eliminación de columnas no predictivas  
columns_to_drop = [  
    'encounter_id', 'patient_nbr', 'readmitted',  
    'weight', 'payer_code', 'medical_specialty'  
]
```

```
# Manejo de valores missing
```

```
df_processed = df_processed.replace('?', np.nan)
```

4.1.2 Ingeniería de Variables

- **Variable objetivo binaria:** target = (readmitted == '<30')
- **Codificación one-hot** para variables categóricas
- **Escalado estándar** para variables numéricas
- **Imputación** con mediana (numéricas) y moda (categóricas)

4.1.3 Pipeline de Preprocesamiento

python

```
preprocessor = ColumnTransformer(  
    transformers=[  
        ('num', numeric_transformer, numeric_features),
```

```
('cat', categorical_transformer, categorical_features)  
]  
)
```

4.2 Modelado de Machine Learning

4.2.1 Algoritmos Implementados

1. **Regresión Logística**
2. **Random Forest**
3. **Gradient Boosting**

4.2.2 Hiperparámetros Optimizados

python

```
param_grids = {

    'Logistic Regression': {

        'classifier__C': [0.1, 1, 10],

        'classifier__max_iter': [1000]

    },

    'Random Forest': {

        'classifier__n_estimators': [100, 200],

        'classifier__max_depth': [10, 20]

    }

}
```

4.2.3 Validación

- **Estrategia:** Validación Cruzada Estratificada (5 folds)
- **Métrica principal:** AUC-ROC
- **División:** 80% entrenamiento, 20% prueba

4.3 Implementación Web

4.3.1 Arquitectura

text

Frontend (HTML/CSS/JS) \longleftrightarrow Backend (Flask) \longleftrightarrow Modelo ML

4.3.2 Componentes

- **Frontend:** Interfaz responsive con formulario dinámico
- **Backend:** API REST con endpoint /predict
- **Modelo:** Pipeline completo de preprocesamiento y predicción

5. RESULTADOS

5.1 Desempeño del Modelo

5.1.1 Métricas de Evaluación

Modelo	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	0.852	0.643	0.234	0.344	0.640
Random Forest	0.859	0.667	0.256	0.370	0.642
Gradient Boosting	0.858	0.659	0.251	0.363	0.641

5.1.2 Modelo Seleccionado

Random Forest fue seleccionado como mejor modelo por:

- Mayor AUC-ROC (0.6425)
- Buen balance entre precisión y recall
- Robustez ante overfitting

5.2 Análisis de Características

5.2.1 Variables Más Relevantes

Durante el proceso de entrenamiento del modelo, se identificaron las variables que más influyen en la predicción de readmisiones hospitalarias en pacientes con diabetes. Las características más significativas fueron:

- Tiempo de estancia en el hospital
- Número de medicaciones prescritas
- Número de diagnósticos registrados
- Número de procedimientos de laboratorio realizados
- Edad del paciente

Estas variables fueron seleccionadas por su alta correlación con los resultados del modelo y su relevancia clínica, lo que permite una mejor interpretación de los factores que inciden en la probabilidad de readmisión.

5.3 Implementación Exitosa

5.3.1 Aplicación Web

Se desarrolló una aplicación web funcional que permite la interacción directa con el modelo predictivo. Las principales características de la plataforma incluyen:

- Interfaz intuitiva y adaptable a distintos dispositivos
- Generación de predicciones en tiempo real (menos de dos segundos por consulta)
- Visualización gráfica de los resultados obtenidos
- Recomendaciones personalizadas basadas en el perfil del paciente

Esta herramienta está diseñada para facilitar su uso por parte de profesionales de la salud, sin necesidad de conocimientos técnicos avanzados.

5.3.2 Archivos Generados

Como resultado del proceso de desarrollo, se generaron los siguientes archivos clave:

- diabetes_model.pkl: archivo que contiene el modelo entrenado
- feature_info.pkl: metadatos sobre las características utilizadas
- model_results.pkl: resultados de evaluación del modelo

Estos archivos permiten la reutilización del modelo, su análisis posterior y la integración con otros sistemas clínicos.

6. IMPLEMENTACIÓN

6.1 Requisitos del Sistema

6.1.1 Dependencias

txt

Python 3.8+

pandas>=2.0.0

scikit-learn>=1.3.0

flask>=2.3.0

joblib>=1.3.0

6.1.2 Recursos Computacionales

- **RAM mínima:** 4GB
- **Almacenamiento:** 500MB
- **Procesador:** Multicore recomendado

6.2 Instrucciones de Instalación

Paso 1: Preparación del Entorno

bash

Clonar o descargar los archivos del proyecto

git clone <repositorio>

cd diabetic-readmission-prediction

Instalar dependencias

pip install -r requirements.txt

Paso 2: Verificación de Archivos

text

Asegurar que existan:

- diabetic_data.csv
- diabetes_model.pkl

- feature_info.pkl
- app.py
- index.html, styles.css, script.js

Paso 3: Ejecución del Sistema

bash

```
# Opción A: Servidor completo
```

```
python app.py
```

6.3.1 Para Personal Médico

1. Ingresar datos demográficos del paciente
2. Completar información clínica relevante
3. Revisar predicción de riesgo
4. Seguir recomendaciones según nivel de riesgo

6.3.2 Interpretación de Resultados

-  **Alto Riesgo (>60%)**: Intervención inmediata requerida
-  **Riesgo Moderado (30-60%)**: Vigilancia aumentada
-  **Bajo Riesgo (<30%)**: Seguimiento estándar

6.4 Resultados

💡 Predicción de Readmisión

Análisis Inteligente de Riesgo Hospitalario

Tiempo en hospital (días): Número de medicamentos:

3 10

Número de diagnósticos: Insulina:

5 No

Resultado A1C: Edad:

None [0-10)

Analizar

💡 Predicción de Readmisión

Análisis Inteligente de Riesgo Hospitalario

Tiempo en hospital (días): Número de medicamentos:

3 10

Número de diagnósticos: Insulina:

5 No

Resultado A1C: Edad:

None [20-30)

Analizar

Sin Riesgo

Probabilidad: 43.38%

6.4 Resultados

Predicción de Readmisión

Análisis Inteligente de Riesgo Hospitalario

Tiempo en hospital (días): Número de medicamentos:

Número de diagnósticos: Insulina:

Resultado A1C: Edad:

 Analizar

 Sin Riesgo
Probabilidad: 45.68%

7. Conclusiones

7.1 Logros del Proyecto

7.1.1 Técnicos

El proyecto logró desarrollar un modelo predictivo con un desempeño sólido, alcanzando un valor de AUC de 0.642, superior a las métricas base utilizadas en la predicción de readmisiones hospitalarias. Se implementó un pipeline automatizado que abarca desde el preprocesamiento de datos hasta el entrenamiento y despliegue del modelo. Además, se diseñó una interfaz web intuitiva, accesible para usuarios sin formación técnica, y se elaboró documentación completa que permite la reproducción del proceso y facilita su comprensión.

7.1.2 Clínicos

Desde el punto de vista clínico, el sistema representa una herramienta de apoyo para la identificación temprana de pacientes con alto riesgo de readmisión. Ofrece recomendaciones personalizadas basadas en el perfil de riesgo de cada paciente, lo que permite implementar intervenciones preventivas y mejorar la calidad de la atención médica.

7.2 Limitaciones y Consideraciones

7.2.1 Limitaciones Técnicas

El modelo depende de la calidad de los datos clínicos disponibles, por lo que su precisión puede verse afectada por la falta de información o errores en los registros. Asimismo, existe la posibilidad de que el modelo reproduzca sesgos presentes en los datos de entrenamiento. Para garantizar su aplicabilidad, es necesario validar el sistema en diferentes contextos clínicos y poblaciones.

7.2.2 Consideraciones Éticas

Es importante destacar que el sistema está diseñado como una herramienta complementaria y no pretende sustituir el juicio clínico. Se ha puesto especial atención en el manejo responsable de los datos de salud, respetando la privacidad de los pacientes. No obstante, se reconoce que los modelos de tipo *ensemble*

pueden presentar limitaciones en cuanto a su explicabilidad, lo que representa un desafío en términos de transparencia.

7.3 Trabajo Futuro

7.3.1 Mejoras Técnicas

Se propone incorporar técnicas de inteligencia artificial explicable (XAI) para mejorar la interpretabilidad del modelo. También se plantea la integración con sistemas de historia clínica electrónica (EHR), el desarrollo de modelos de series de tiempo para seguimiento longitudinal, y la exploración de arquitecturas más complejas como redes neuronales profundas y modelos de ensamble avanzados.

7.3.2 Expansión Clínica

En el ámbito clínico, se recomienda realizar validaciones prospectivas en entornos reales, extender el sistema a otras enfermedades crónicas, integrarlo con programas de manejo de enfermedades, y desarrollar mecanismos de alerta automática y paneles de control que faciliten la toma de decisiones médicas.

8. Anexos

8.1 Estructura de Archivos del Proyecto

El proyecto se encuentra organizado en diferentes carpetas, cada una con archivos específicos que cumplen funciones esenciales para el desarrollo, implementación y documentación del sistema de predicción de readmisión hospitalaria:

- **Datos**

Contiene los archivos utilizados para el entrenamiento del modelo:

- diabetic_data.csv: conjunto de datos principal.
- data_dictionary.pdf: documento con la descripción de las variables incluidas en el conjunto de datos.

- **Código**

Incluye los scripts necesarios para el desarrollo y ejecución del modelo:

- diabetes_model_training.ipynb: cuaderno de Jupyter con el proceso de entrenamiento.
- app.py: archivo principal de la aplicación web.
- requirements.txt: lista de dependencias requeridas.
- run.bat: script para ejecutar la aplicación.

- **Web**

Contiene los archivos que conforman la interfaz web del sistema:

- index.html: estructura de la página principal.
- styles.css: hoja de estilos para la presentación visual.
- script.js: funcionalidades interactivas del sitio.

- **Modelos**

Archivos relacionados con el modelo entrenado y sus resultados:

- diabetes_model.pkl: archivo del modelo entrenado.
- feature_info.pkl: información sobre las variables utilizadas.
- model_results.pkl: resultados obtenidos durante el entrenamiento.

- **Documentación**

Incluye los documentos que explican el funcionamiento y uso del sistema:

- README.md: resumen general del proyecto.
- Technical_Report.pdf: informe técnico detallado.
- User_Manual.pdf: manual de usuario para la aplicación.

8.2 Referencias Técnicas

- Pedregosa et al. (2011). *Scikit-learn: Machine Learning in Python*.
- McKinney (2010). *Data Structures for Statistical Computing in Python*.

Grinberg (2018). *Flask Web Development*.

Declaración Final

El presente proyecto ha permitido comprobar que los sistemas de predicción de readmisión hospitalaria basados en técnicas de *Machine Learning* son viables y pueden aportar beneficios reales en el ámbito de la salud. A través del desarrollo de un modelo predictivo, se logró alcanzar un valor de AUC de 0.642, lo cual indica un desempeño aceptable en la identificación de pacientes diabéticos con riesgo de ser readmitidos.

Este tipo de herramientas puede ser de gran utilidad para los profesionales de la salud, ya que permite anticiparse a posibles complicaciones y tomar decisiones informadas que mejoren la atención médica. Además, al identificar de manera temprana a los pacientes en riesgo, se pueden implementar estrategias preventivas que contribuyan a reducir las tasas de readmisión, mejorar los resultados clínicos y optimizar el uso de los recursos hospitalarios.

En conclusión, el modelo desarrollado representa un paso importante hacia la integración de soluciones tecnológicas en el sector salud, demostrando que el uso de inteligencia artificial puede tener un impacto positivo en la gestión hospitalaria y en la calidad de vida de los pacientes.