

Primera Entrega de Proyecto

POR:

Iván Daniel Salazar Alarcón

MATERIA

Fundamentos de Deep Learning

PROFESOR:

Raúl Ramos Pollan



1 8 0 3

UNIVERSIDAD DE ANTIOQUIA

Facultad de Ingeniería

Medellín 2023

1. Contexto de aplicación

El reingreso hospitalario es el retorno no programado de un paciente dentro de un preespecificado período de tiempo después del alta hospitalaria,, internacionalmente se asumen los siguientes 30 días como dicha ventana de tiempo [1].

Las causas detrás de los reingresos hospitalarios son diversas y muchas son evitables, por este motivo es importante abordar este tema de estudio, además que esta medida podría ser utilizada como indicador de calidad de atención al paciente y de otros factores considerables como su impacto económico en el sistema de salud.

2. Dataset

Este dataset se encuentra disponible en www.hindawi.com en uno de sus artículos [2], el conjunto de datos representa 5 MB de información recolectada entre 1999 y 2008 de 130 hospitales de EE. UU. Incluye más de 100 mil registros con 50 características clínicas.

El archivo diabetic_data.csv contienen atributos tales como:

- **patient_nbr:** Identificador único de un paciente.
- **age:** Edad intervalos de 10 años: (0-10), (10-20), ..., (90-100).
- **time_in_hospital:** Número de días de hospitalización del paciente.
- **num_lab_procedures:** Número de pruebas de laboratorio realizadas durante el encuentro.
- **num_procedures:** Número de procedimientos realizados durante la hospitalización.
- **num_medications:** Número de medicamentos distintos administrados durante la hospitalización.
- **number_outpatient:** Número de visitas ambulatorias del paciente en el año anterior a la hospitalización.
- **number_emergency:** Número de visitas a urgencias del paciente en el año anterior a la hospitalización.
- **number_inpatient:** Número de visitas hospitalarias del paciente en el año anterior a la hospitalización.
- **number_diagnoses:** Número de diagnósticos ingresados al sistema.

3. Objetivo de machine learning

Esta problemática la viven los hospitales y sus pacientes, en donde se hace la valoración del alta de una persona debido a temas cualitativos, subjetivos y según la experiencia del personal de salud. Sin embargo, se espera poder cuantificar la detección de los pacientes con mayor riesgo de reingreso a través de técnicas de Machine Learning. Por lo tanto, en esta base de datos se consideran 3 salidas diferentes dadas por el campo *readmitted*:

- '**NO**': Sin readmisión.
- '<**30**': Un reingreso en menos de 30 días.
- '>**30**': Un reingreso en más de 30 días.

4. Métricas de desempeño (de machine learning y negocio)

La métrica de evaluación principal para el modelo será la función de perdida Cross-entropía Categórica, la cual se define de la siguiente manera:

$$\mathcal{L}(\hat{y}, y) = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C 1_{y_i \in C_j} \log (P_{model}[y_i \in C_j])$$

Donde:

$1_{y_i \in C_j}$: Función indicadora de la i -ésima observación perteneciente a la j -ésima categoría.

$P_{model}[y_i \in C_j]$: Probabilidad predicha por el modelo para que la i -ésima observación pertenezca a la j -ésima categoría

Se decide hacer uso de la Cross-entropía Categórica ya que la predicción se realizará para más de dos clases, por lo tanto, la salida es la probabilidad de que la entrada de la red se clasifique como perteneciente a la categoría respectiva.

De esta manera se espera que con el diseño de un modelo de Machine Learning logremos alcanzar por lo menos un acierto en el 70% de las predicciones al calificar a una persona con mayor riesgo de readmisión.

5. Referencias y resultados previos

Desde hace varios años se han desarrollado modelos basados en datos para identificar a los pacientes con alto riesgo de reingreso hospitalario, han demostrado un mejor potencial de los modelos de aprendizaje automático que las puntuaciones tradicionales (LACE y HOSPITAL) [3].

Se evidencia que cerca del 80% del esfuerzo en un modelo analítico es el preprocesamiento, fusión, personalización y limpieza de conjuntos de datos, la predicción precisa de los reingresos hospitalarios es difícil de lograr, debido a su compleja relación con las condiciones de salud de los pacientes, especialmente en las enfermedades crónicas [4].

BIBLIOGRAFÍA

- [1] Wang, Shuwen and Zhu and Xingquan, "Predictive modeling of hospital readmission: challenges and solutions," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, pp. 2975--2995, 2021.
- [2] Strack, Beata and DeShazo, Jonathan P and Gennings, Chris and Olmo, Juan L and Ventura, Sebastian and Cios, Krzysztof J and Clore y John N, «Impact of HbA1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records,» *BioMed research international*, 2014.
- [3] Van Walraven, Carl and Dhalla, Irfan A and Bell, Chaim and Etchells, Edward and Stiell, Ian G and Zarnke, Kelly and Austin, Peter C and Forster y Alan J, «Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community,» *Can Med Assoc*, vol. 182, pp. 551-557, 2010.
- [4] Rajkomar, Alvin and Oren, Eyal and Chen, Kai and Dai, Andrew M and Hajaj, Nissan and Hardt, Michaela and Liu, Peter J and Liu, Xiaobing and Marcus, Jake and Sun y Mimi and others, «Scalable and accurate deep learning with electronic health records,» *Nature Publishing Group UK London*, p. 18, 2018.