

Comparación de Modelos de Machine Learning y Deep Learning para la Predicción de Accidentes Cerebrovasculares

Grupo 09

Fernielis Vicioso
Julián Núñez
Michael Ventura



MODELO DE PREDICCIÓN DE PADECIMIENTO DE UN STROKE

ACCIDENTE CARDIO VASCULAR (ACV)

Enlace a GitHub:

```
https://github.com/michaelventura01/healthcare-dataset-stroke-data/blob/main/healthcare\_dataset\_stroke\_data.ipynb
```

1

Resumen Teorico

A

Objetivo

C

Procesamiento y
Preparación del Dataset

B

Tipo de Modelado

D

Pipeline sugerido
(EDA y Modelado)

Objetivo

Predecir la probabilidad de que un paciente sufra un stroke (ACV) a partir de variables demográficas, clínicas y de estilo de vida.



1

Resumen Teorico

A

Objetivo

C

Procesamiento y
Preparación del Dataset

B

Tipo de Modelado

D

Pipeline sugerido
(EDA y Modelado)

Tipo de Modelado

Tipo de problema:

Clasificación binaria (stroke: 0 o 1)

Modelos a aplicar:

- **3 modelos tradicionales:** Logistic Regression, Random Forest y Support Vector Machine (SVM).
- **Red Neuronal con 1 capa oculta.**
- **Red Neuronal con múltiples capas ocultas (DNN).**



1

Resumen Teorico

A

Objetivo

C

Procesamiento y
Preparación del Dataset

B

Tipo de Modelado

D

Pipeline sugerido
(EDA y Modelado)

Procesamiento y Preparación del Dataset

Variable	Tipo	Tratamiento
id	Identificador	Eliminar (no aporta valor predictivo)
gender	Categórica	One-hot encoding o Label encoding
age	Numérica	Normalizar o escalar
hypertension	Binaria (0/1)	Usar directamente
heart_disease	Binaria (0/1)	Usar directamente
ever_married	Categórica	Convertir a 0 (No), 1 (Yes)
work_type	Categórica	One-hot encoding
Residence_type	Categórica	One-hot encoding (Urban/Rural)
avg_glucose_level	Numérica	Normalizar o escalar
bmi	Numérica	Imputar valores nulos (media o regresión)
smoking_status	Categórica	One-hot encoding, considerar categoría
stroke	Binaria	Variable objetivo

1

Resumen Teorico

A

Objetivo

C

Procesamiento y
Preparación del Dataset

B

Tipo de Modelado

D

Pipeline sugerido
(EDA y Modelado)

Pipeline sugerido (EDA y Modelado)

EDA (Exploratory Data Analysis)

- Distribución del target (stroke).
- Distribuciones por age, bmi, glucose_level.
- Gráficas por categoría (work_type vs stroke, gender vs stroke).
- Heatmap de correlaciones.
- Outliers (en bmi, glucose, etc).

Preprocesamiento

- Encoding de variables categóricas.
- Escalado de variables numéricas.
- Imputación de valores nulos (bmi).
- Balanceo de clases con SMOTE o class_weight.

Pipeline sugerido (EDA y Modelado)

Modelos Tradicionales

- Logistic Regression (baseline simple, interpretable).
- Random Forest (para ver importancia de variables).
- SVM (bueno para espacios de alta dimensión).

Evalúa con

- accuracy, precision, recall, f1-score, AUC
- Confusion matrix
- Curva ROC

