

Librerías Python Usadas

Pandas	Pandas es una librería de Python que se utiliza para la manipulación y análisis de datos.
Matplotlib	Sirve para crear visualizaciones y gráficos de datos.
Seaborn	Proporciona varias funciones para personalizar los gráficos.
Lightgbm	Utilizado para la categorización, clasificación y muchas otras tareas de aprendizaje automático, en las que es necesario maximizar o minimizar una función objetivo mediante la técnica de gradient boosting, que consiste en combinar clasificadores sencillos.
Scikit Learn	Herramientas para preprocesar datos, las funciones para evaluar modelos y los mecanismos para ajustar los parámetros de cada modelo.

Funciones Usadas de Pandas

head()	Devuelve las n primeras filas del DataFrame df.
info()	Devuelve información (número de filas, número de columnas, índices, tipo de las columnas y memoria usado) sobre el DataFrame df.
shape()	Devuelve una tupla con el número de filas y columnas del DataFrame df.
isnull()	Devuelve una estructura con las mismas dimensiones que la que se cede como argumento sustituyendo cada valor por el booleano True si el correspondiente elemento es un valor nulo, y por el booleano False en caso contrario.
value_counts()	Obtener cuantos datos tenemos de algo en específico.
map()	Se utiliza para asignar valores de dos series que tienen una columna similar.

Funciones Usadas de Seaborn

countplot()	Utilizado para mostrar los recuentos de observaciones en cada contenedor categórico mediante barras.
heatmap()	Permite crear mapas de calor o gráficos de correlación en Python con seaborn.

Funciones Usadas de Lightgbm

dataset()	Carga data en la librería.
train()	Seteo de data para entrenamiento.
predict()	Seteo de data para predicción.

Funciones Usadas de Matplotlib

figure()	Elemento base sobre el que se construyen todos los gráficos
title()	Seteo de un título para la gráfica a mostrar.
show()	Mostrar figura.

Funciones Usadas de Scikit Learn

fit_transform()	Ajustar a los datos y luego transformarlos.
train_test_split()	Divida matrices o matrices en subconjuntos de pruebas y trenes aleatorios.



By StefPanchana

cheatography.com/stefpanchana/

Not published yet.

Last updated 26th April, 2024.

Page 1 of 2.

Sponsored by [Readable.com](https://readable.com)

Measure your website readability!

<https://readable.com>

Funciones Usadas de Scikit Learn (cont)

- Random Forest Classifier** Un bosque aleatorio es un metaestimador que ajusta una serie de clasificadores de árboles de decisión en varias submuestras del conjunto de datos y utiliza promedios para mejorar la precisión predictiva y controlar el sobreajuste.
- predict_proba()** Proporciona un método predict_proba para regresión logística, que devuelve las probabilidades predichas para cada clase.

Análisis del Dataset

Cantidad de Registros:	7905
Cantidad de Columnas:	20
Tipos de datos de las columnas:	float64(10), int64(3), object(7)
Información Nula Encontrada:	0 registros.

Tareas de EDA aplicadas al Dataset

- 1.- Verificación de Columnas de Información Existentes
- 2.- Validación de tipos de datos de los registros.
- 3.- Comprobación del set de datos global.
- 4.- Análisis de la distribución de datos por la columna Estado.
- 5.- Verificación de datos nulos en el dataset.
- 6.- Transformación de datos categóricos a binarios.
- 7.- Crear visualización de correlación de data.
- 8.- Limpieza de columna de id.
- 9.- Configuración de data para entrenamiento y predicción.
- 10.- Regresión de datos.
- 11.- Aplicación de entrenamiento y predicción de datos.
- 12.- Creación de un archivo plano de los resultados obtenidos.

Distribución de Estados - Matplotlib

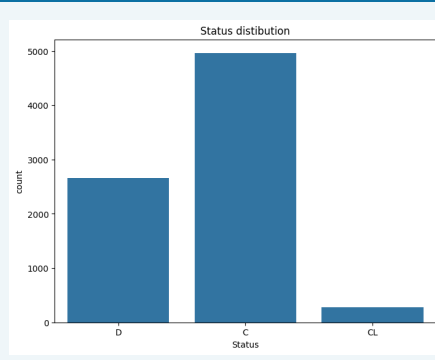
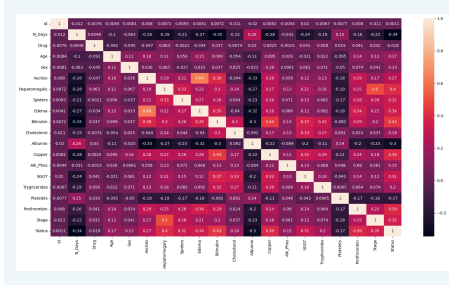


Gráfico de Correlación de Datos - Seaborn



Link de los Recursos

- Kaggle Link:** <https://www.kaggle.com/competitions/playground-series-s3e-26/data>
- Información General del Dataset:** El conjunto de datos se generó a partir de un modelo de aprendizaje profundo entrenado en el conjunto de datos de Predicción de supervivencia del paciente con cirrosis.