Contenido

[Entender el problema y definir el objetivo 1](#_Toc194819302)

[Obtener los datos (cargar librerías y dataset) 2](#_Toc194819303)

[Explorar el dataset (Análisis exploratorio de datos - EDA) 2](#_Toc194819304)

[Preparar los datos (transformación y limpieza de datos) 2](#_Toc194819305)

[Visualización inicial 2](#_Toc194819306)

[Modelado y evaluación inicial 2](#_Toc194819307)

[Optimización del modelo 2](#_Toc194819308)

[Predicciones finales e insights 3](#_Toc194819309)

[Presentación de resultados y recomendaciones 3](#_Toc194819310)

[(Opcional) Despliegue del modelo 3](#_Toc194819311)

**Parte 1: Analítica de Datos**

# Entender el problema y definir el objetivo

* **person\_age**: Edad del solicitante del préstamo.
* **person\_income**: Ingreso anual del solicitante.
* **person\_home\_ownership**: Tipo de propiedad de vivienda (ej. alquilada, propia, hipotecada).
* **person\_emp\_length**: Tiempo de empleo del solicitante en años.
* **loan\_intent**: Propósito del préstamo (ej. educación, negocio, automóvil).
* **loan\_grade**: Calificación del préstamo según su riesgo.
* **loan\_amnt**: Monto del préstamo solicitado.
* **loan\_int\_rate**: Tasa de interés aplicada al préstamo.
* **loan\_status**: Estado del préstamo (0 = no incumplimiento, 1 = incumplimiento).
* **loan\_percent\_income**: Porcentaje del ingreso anual destinado al pago del préstamo.
* **cb\_person\_default\_on\_file**: Historial de incumplimiento de pagos (sí o no).
* **cb\_person\_cred\_hist\_length**: Duración del historial crediticio del solicitante en años.

Algunos de los objetivos que podemos plantear son:

* Objetivo: Predecir la probabilidad de incumplimiento del préstamo (loan\_status) Clasificacion
* Objetivo: Predecir la tasa de interés (loan\_int\_rate) basada en las características del cliente y el préstamo. regresion
* Objetivo: agrupar a los solicitantes según características como ingresos, historial crediticio y porcentaje de ingresos destinado al préstamo. k means clustering

Resumen

En esta parte entendimos las características del data set y definimos objetivos principales para implementación de modelos de aprendizaje automático. En lo sucesivo se extraerá información sobre el análisis y exploración de datos

# Obtener los datos (cargar librerías y dataset)

En esta parte simplemente cargamos los datos y no hubo ninguna dificultad ya que estamos hablando de 32000 filas y un archivo csv

# Explorar el dataset (Análisis exploratorio de datos - EDA)

* 1. Inspección de datos
  2. Detección de valores nulos y duplicados
  3. Análisis de datos categóricos
  4. Análisis de datos numéricos
  5. Correlación y covarianza

# Preparar los datos (transformación y limpieza de datos)

En esta parte a partir del paso previo nos dimos cuenta que había que hacer el siguiente tipo de limpieza en los datos:

* 1 Estudiar los valores atipicos en las

En esta parte nos encontramos que efectivamente había valores satíficos y los redujimos imputando el valor de la mediana

* 2 Estudiar la data nula e imputarla

En este punto dos columnas tenían valores nulos por lo que investigamos el tipo de Variable que era cada columna graficamos su distribución y comprobamos que tras imputar un valor no cambie su distribución

* 3 Transformacion de datos categoricos con variables dummy

Se hizo lo correspondiente con las variables categóricas

* 4 Considerar escalar variables

En la parte de regression se escalaron los datos

* 5 Ver correlaciones y evitar la multicolinealidad

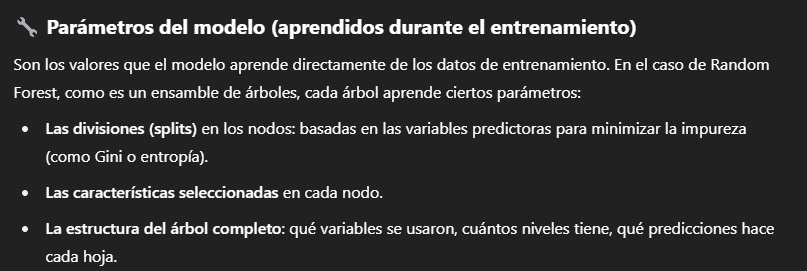
# Visualización inicial

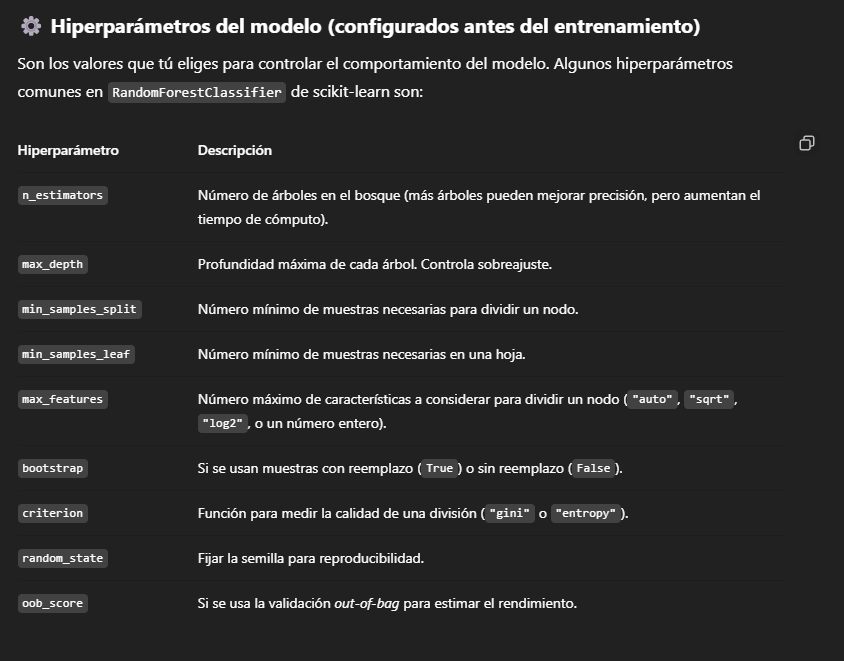
**Parte 2: Científico de Datos**

# Modelado y evaluación inicial

# Optimización del modelo

## Parámetros de hiperparámetros de un bosque aleatorio





# Predicciones finales e insights

# Presentación de resultados y recomendaciones

# (Opcional) Despliegue del modelo