**Proyecto Integrador**

**Avance 6**

**Conclusiones Claves**

Dra. Grettel Barcelo Alonso

**EQUIPO 21**

|  |  |
| --- | --- |
|  |  |
| **Bernardo Mijango Flores**  **Dalia Isabel López Tapia**  **David Valles Canedo** | **A01793654**  **A01245026**  **A01191310** |
|  |  |

**9 de junio de 2024**

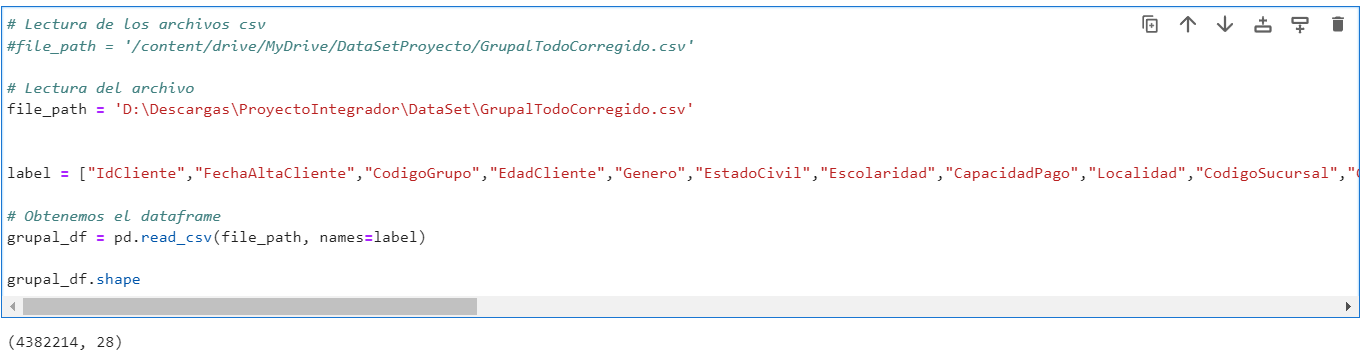
**Objetivo General:**

El objetivo principal de este estudio es desarrollar y evaluar modelos de aprendizaje automático que permitan predecir el estado del crédito de los clientes del producto “Grupal” de la empresa Alsol Contigo utilizando un conjunto de datos que contiene información demográfica y financiera de los clientes, como la edad, estado civil, número de créditos, y diversas métricas crediticias, se busca construir modelos robustos que puedan diferenciar de manera efectiva entre créditos buenos y malos.

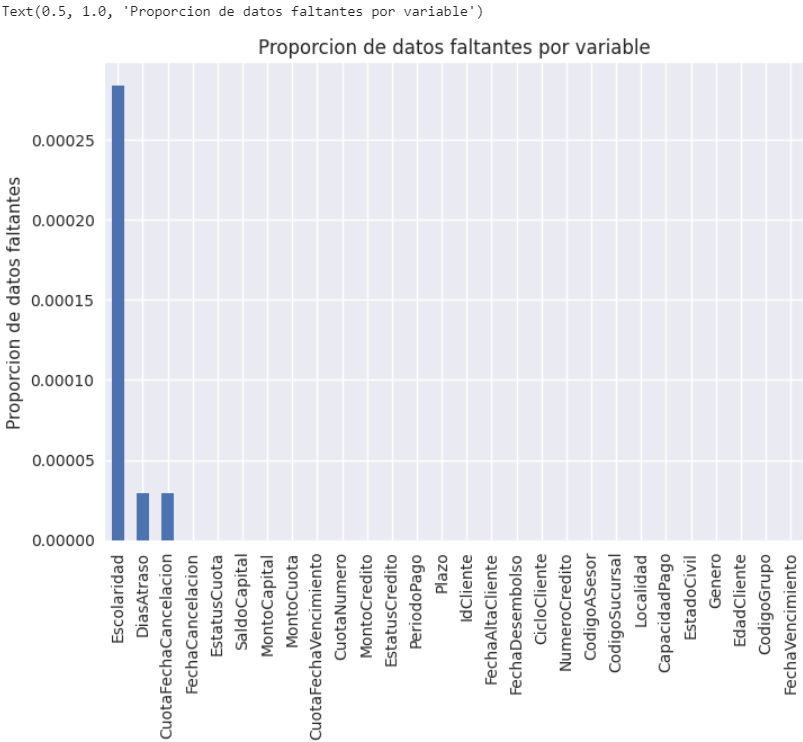
Se busca identificar los modelos con el mejor rendimiento en términos de precisión y eficiencia computacional, con el fin de ofrecer una herramienta confiable y eficaz para la evaluación del riesgo crediticio, utilizando los resultados en la gestión de cobranza temprana con la implementación de un Chat Bot inteligente. El análisis incluye la selección de los modelos más adecuados mediante la comparación de sus puntajes y tiempos de entrenamiento, así como el ajuste destallado de los hiperparámetros para optimizar su desempeño. Además, se evalúa la capacidad predictiva de los modelos seleccionados mediante métricas de evaluación de clasificación y se proporciona una interpretación detallada de los resultados.

**Análisis de los datos**

En la exploración inicial de los datos se comenzó con la carga de información que contiene datos sobre clientes, créditos, características socioeconómicas, historiales crediticios, entre otros. la carga de datos se realiza utilizando bibliotecas como Pandas en Python, que permite la manipulación eficiente de datos tabulares. Se realiza una exploración de estadística descriptiva obteniendo un análisis completo de la información.



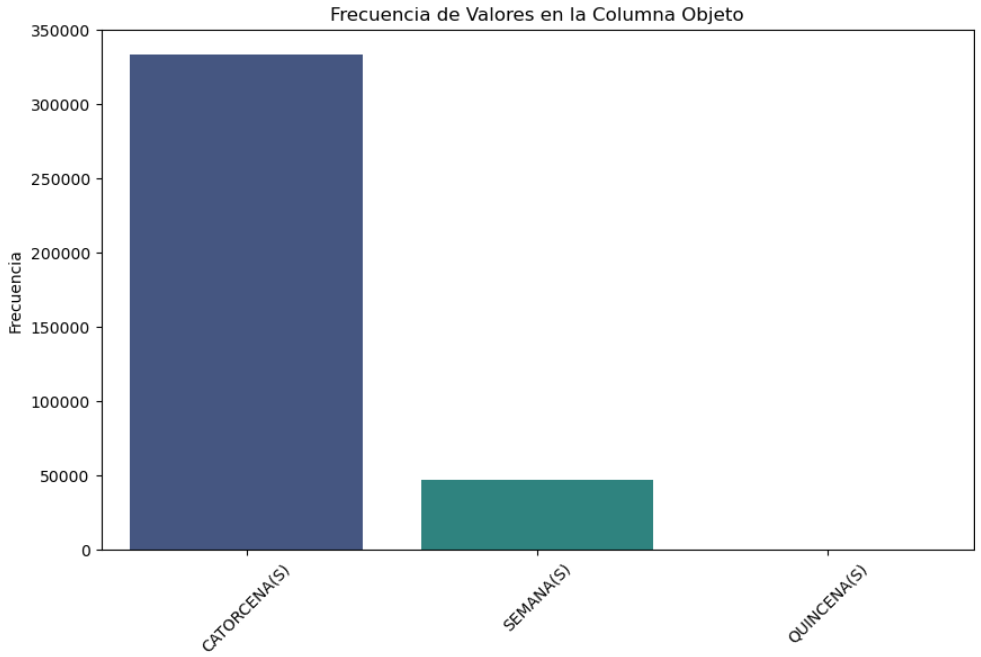
Se determina quitar los valores nulos de la base de datos ya que representa un porcentaje del 0.028% del total de la base de datos.



Al tener el detalle del comportamiento de cada crédito por cliente se determina realizar un análisis de los datos por crédito agregando una nueva variable para obtener el Máximo días de atraso, la Media de los días de atraso y la Mediana de los días de atraso por cada crédito, es decir, se genera una nueva base de datos por crédito agregando las variables “DAMax”, “DAMean” y “DAMedian”



El producto Grupal cuenta con tres periodos de pagos de pagos “CATORCENA(S)”, “SEMANA(S)” y “QUINCENA(S)”, procedemos a obtener la frecuencia de la variable “PeriodoPago”, consideramos que el comportamiento de los créditos puede ser diferentes entre ellos, por lo que se decide trabajar con los créditos de mayor frecuencia resultando ser los de pagos “CATORCENA(S)”



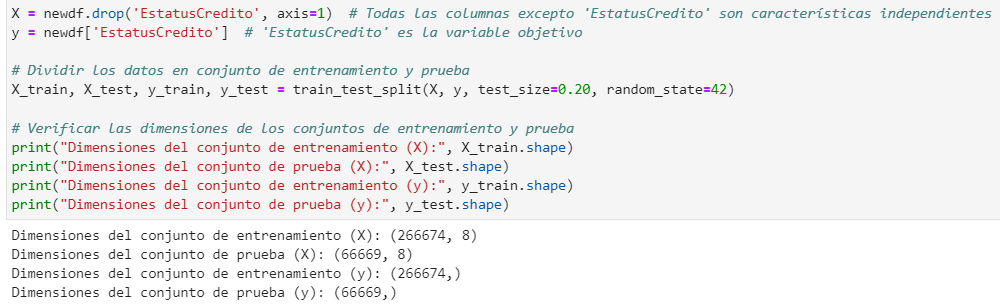
**Variables**

* IdCliente: Número de identificación del cliente
* NumeroCredito: Numero de crédito
* EdadCliente: Edad del cliente al momento del otorgamiento del crédito
* CicloCliente: Numero de créditos contratados por el cliente
* Plazo: Numero de pagos del crédito en periodo CATORCENA(S)
* EstadoCivil: Código del estado civil (1=SOLTERO(A), 2=CASADO(A), 3=VIUDO(A), 4=DIVORCIADO(A), 5=UNION LIBRE, 6=SEPARADO, 7=SE IGNORA)
* DAMax: Dias máximo de atraso
* DAMean: Media de los dias de atraso
* EstatusCredito: Estatus del credito (CAN=PAGADO, CAI=NO PAGADO)

Esta ultima variable se convierte en variable objetivo de valor numérico donde CAN=1 y CAI=0

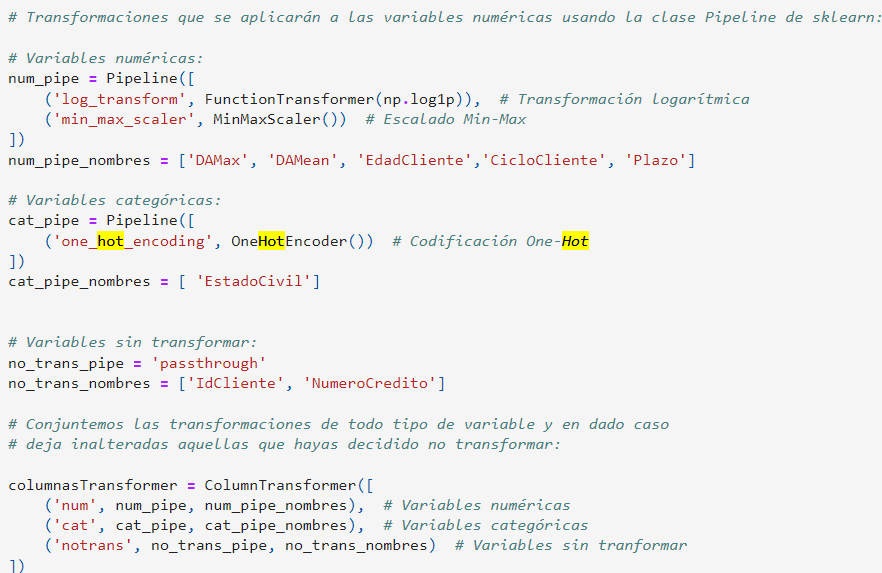


Realizamos la partición del conjunto de datos en entrenamiento del 80% y pruebas del 20%



**Transformaciones y uso del one-hot encoding**

Se aplica una transformación logarítmica y un escalado Min-Max a las columnas númericas, se aplica codifiación One-Hot a las columnas categóricas, se dejan algunas columnas sin transformación y por último se combina todas las transformaciones en un solo transformador.

****