



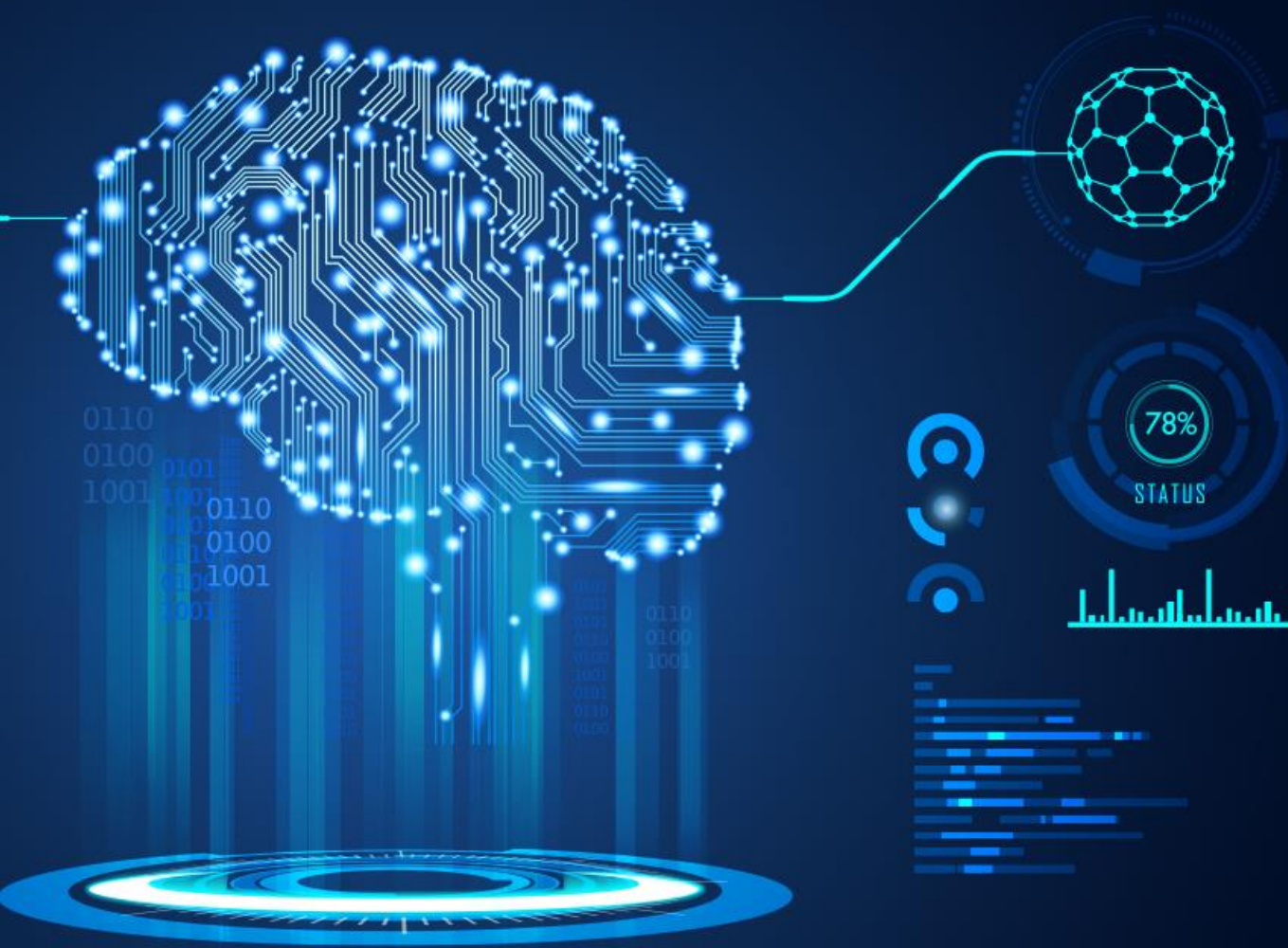
MACHINE LEARNING INMERSION

GRUPO 8: CAMPAÑA DE ACEPTACIÓN DEL PRODUCTO DE AHORRO

Integrantes:

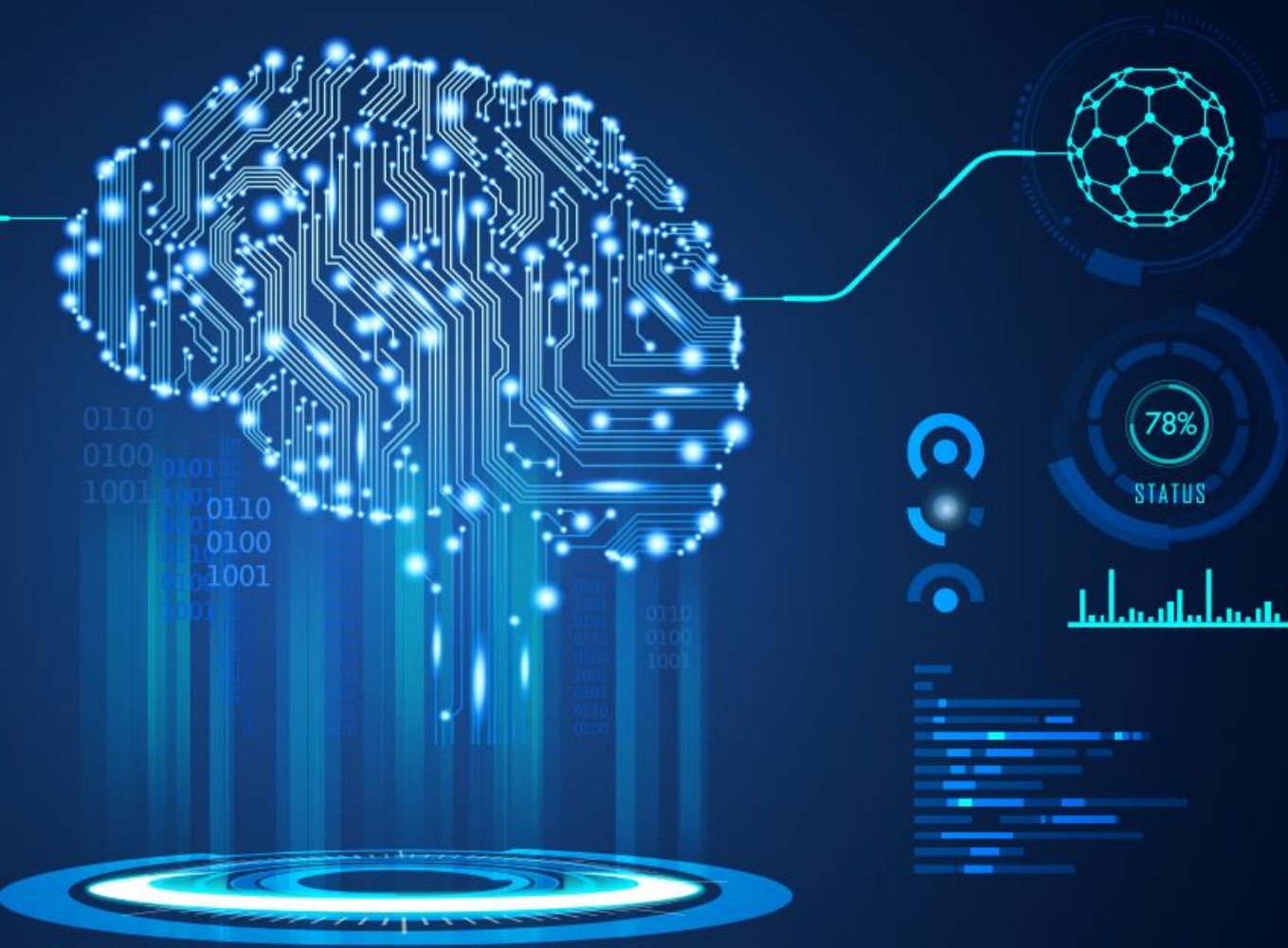
Pamela Fiestas Pittman

Víctor Ortega Mory



ÍNDICE

1. Comprensión del negocio
2. Preparación del Datsaset
3. Análisis descriptivo
4. Limpieza de los datos
5. Elaboración de Machine Learning
6. Anexos



ÍNDICE

1. **Comprensión del negocio**
2. **Preparación del Datsaset**
3. **Análisis descriptivo**
4. **Limpieza de los datos**
5. **Elaboración de Machine Learning**
6. **Anexos**

COMPRENSIÓN DEL NEGOCIO

Contexto

Negocio : Empresa del sector financiero

Objetivo:

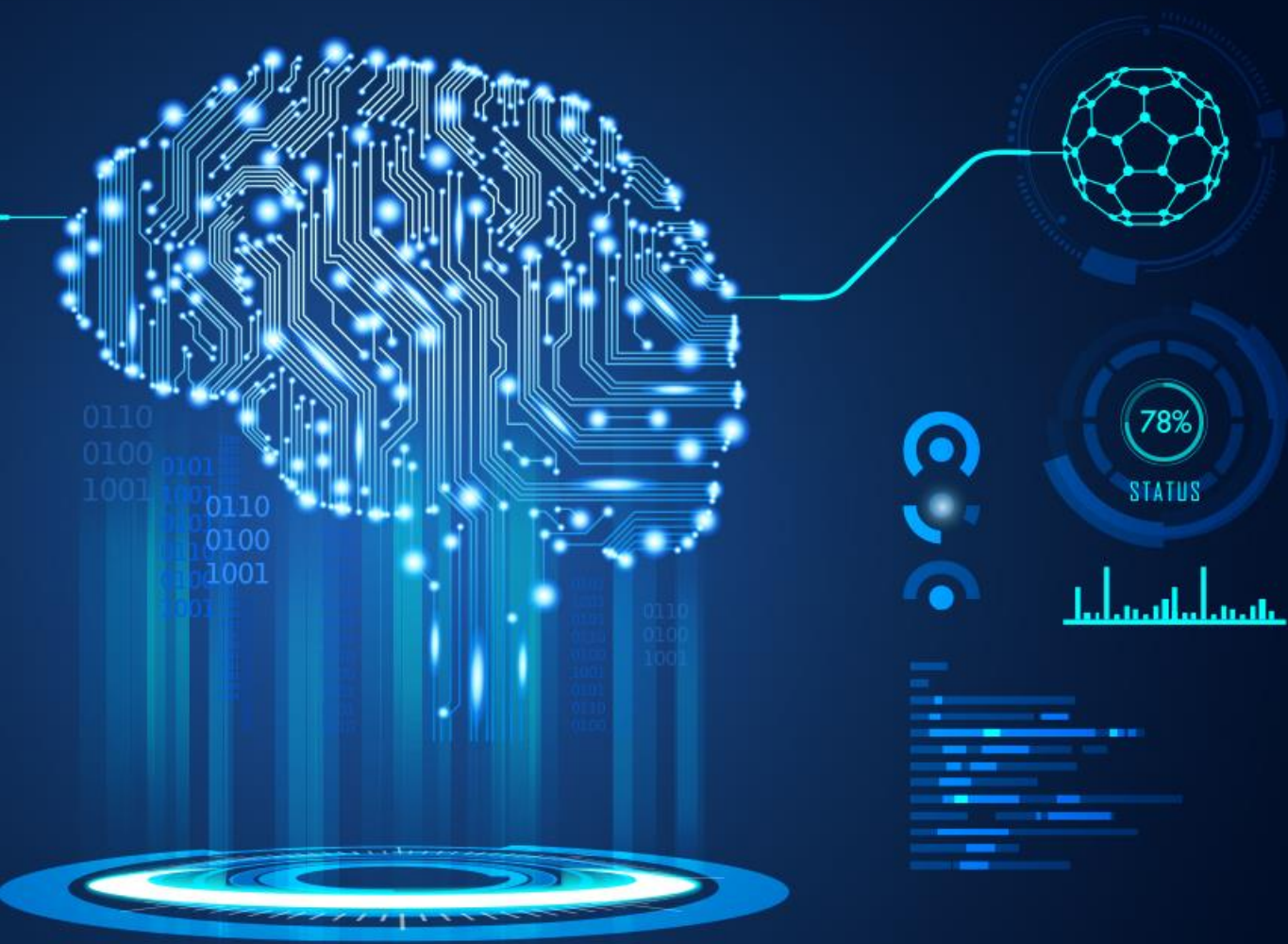
- Incrementar la efectividad de las futuras campañas en la adquisición de productos pasivos

Necesidad

- Identificar los clientes con mayor probabilidad de aceptar productos pasivos

Propuesta de solución

- Desarrollar un Modelo supervisado que permita determinar la probabilidad de aceptación de la campaña de las personas posibles a contactar.



ÍNDICE

1. **Comprensión del negocio**
2. **Preparación del Datsaset**
3. **Análisis descriptivo**
4. **Limpieza de los datos**
5. **Elaboración de Machine Learning**
6. **Anexos**

DATASET: VARIABLES CON LAS QUE SE CUENTA

Características del Dataset

- Data recolectada de un sistema de Visión Única del Cliente.
- La data esta anonimizada.
- Contiene 2,240 registros con data de clientes y 13 características o variables de clientes.
- Una variable de suscripción a producto a un depósito a plazos.
- Desarrollamos procesos de imputación para solucionar la completitud de información.

variable	Descripción de la Variable	Completitud de base
coddoc	Código único de la persona o cliente	100.00%
edad	Edad de la persona o cliente	99.96%
estciv	Estado civil de la persona o cliente	99.51%
educación	Nivel de educación de la persona o cliente	99.24%
mora	El cliente o persona ha tenido comportamiento moroso en los últimos 12 meses?	99.60%
balance	Saldo o deuda total en la entidad. Podría el cliente o persona tener saldo a favor	99.69%
vivienda	El cliente o persona tiene vivienda propia?	99.29%
préstamo	El cliente o persona tiene préstamo personal?	99.60%
dia	Ultimo día del mes contactado	99.78%
duración	Duración del último contacto, en segundos	99.46%
campana	Número de contactos realizados durante esta campaña y para este cliente	99.51%
pdias	Número de días que pasaron después de que el cliente fue contactado la última vez de una campaña (-1 significa que el cliente no fue previamente contactado)	99.96%
previo	Número de contactos hechos antes de esta campaña y para este cliente	99.96%
adq_ahorro	El cliente se ha suscrito el cliente un depósito a plazo?	100.00%

DATASET: TIPOS DE VARIABLES

variable	Descripción de la Variable	Tipo de variable
coddoc	Código único de la persona o cliente	OBJECT
edad	Edad de la persona o cliente	FLOAT64
estciv	Estado civil de la persona o cliente	OBJECT
educación	Nivel de educación de la persona o cliente	OBJECT
mora	El cliente o persona ha tenido comportamiento moroso en los últimos 12 meses?	OBJECT
balance	Saldo o deuda total en la entidad. Podría el cliente o persona tener saldo a favor	FLOAT64
vivienda	El cliente o persona tiene vivienda propia?	OBJECT
préstamo	El cliente o persona tiene préstamo personal?	OBJECT
dia	Ultimo día del mes contactado	FLOAT64
duración	Duración del último contacto, en segundos	FLOAT64
campana	Número de contactos realizados durante esta campaña y para este cliente	FLOAT64
pdias	Número de días que pasaron después de que el cliente fue contactado la última vez de una campaña (-1 significa que el cliente no fue previamente contactado)	FLOAT64
previo	Número de contactos hechos antes de esta campaña y para este cliente	FLOAT64
adq_ahorro	El cliente se ha suscrito el cliente un depósito a plazo?	INT64

DATASET: PRESENCIA DE MISSING VALUES

Cuantitativas	Cantidad de MV
edad	1
balance	7
día	5
duración	12
campana	11
pdías	1
previo	1

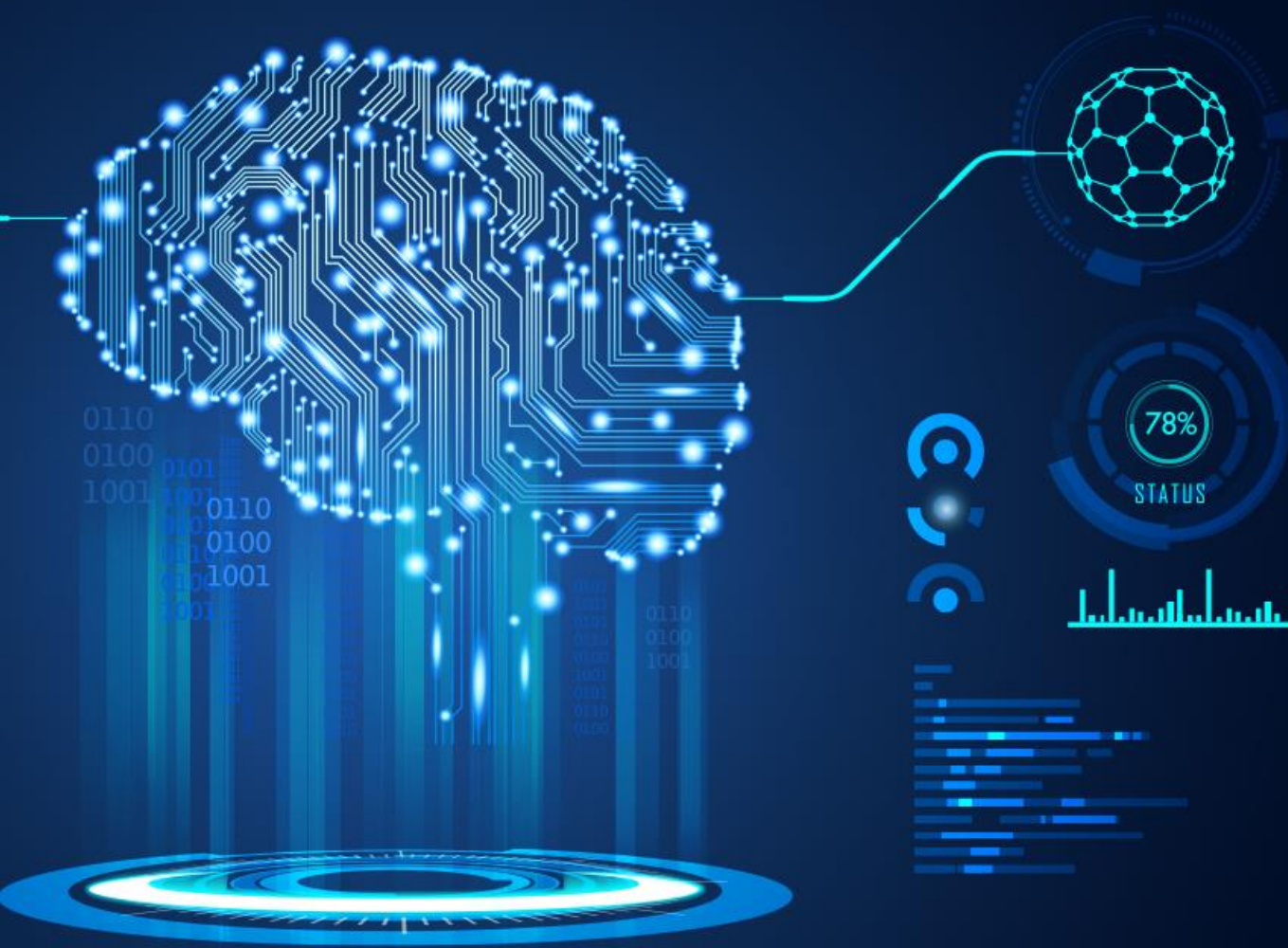
Categóricas	Cantidad de MV
estciv 11	11
educación	17
mora	9
vivienda	16
préstamo	9

DATASET: VARIABLES CON LENGUAJE HUMANO

estciv	education	mora	vivienda	préstamo
married	primaria	no	no	no
married	secundaria	no	si	si
single	terciario	no	si	no
married	terciario	no	si	si
single	terciario	no	no	no

DATASET: VARIABLES CON LENGUAJE HUMANO

Dataset	Cantidad
Adquirió productos pasivos	258
No adquirió productos pasivos	1982

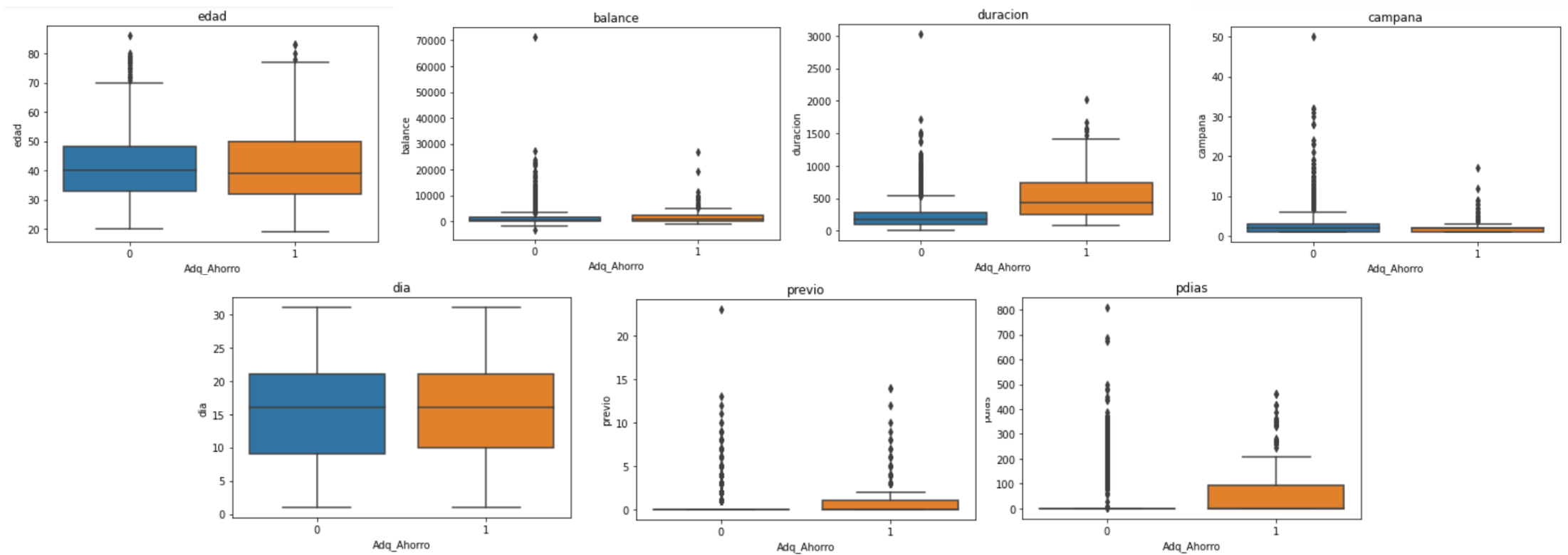


ÍNDICE

1. Comprensión del negocio
2. Preparación del Datsaset
3. **Análisis descriptivo**
4. Limpieza de los datos
5. Elaboración de Machine Learning
6. Anexos

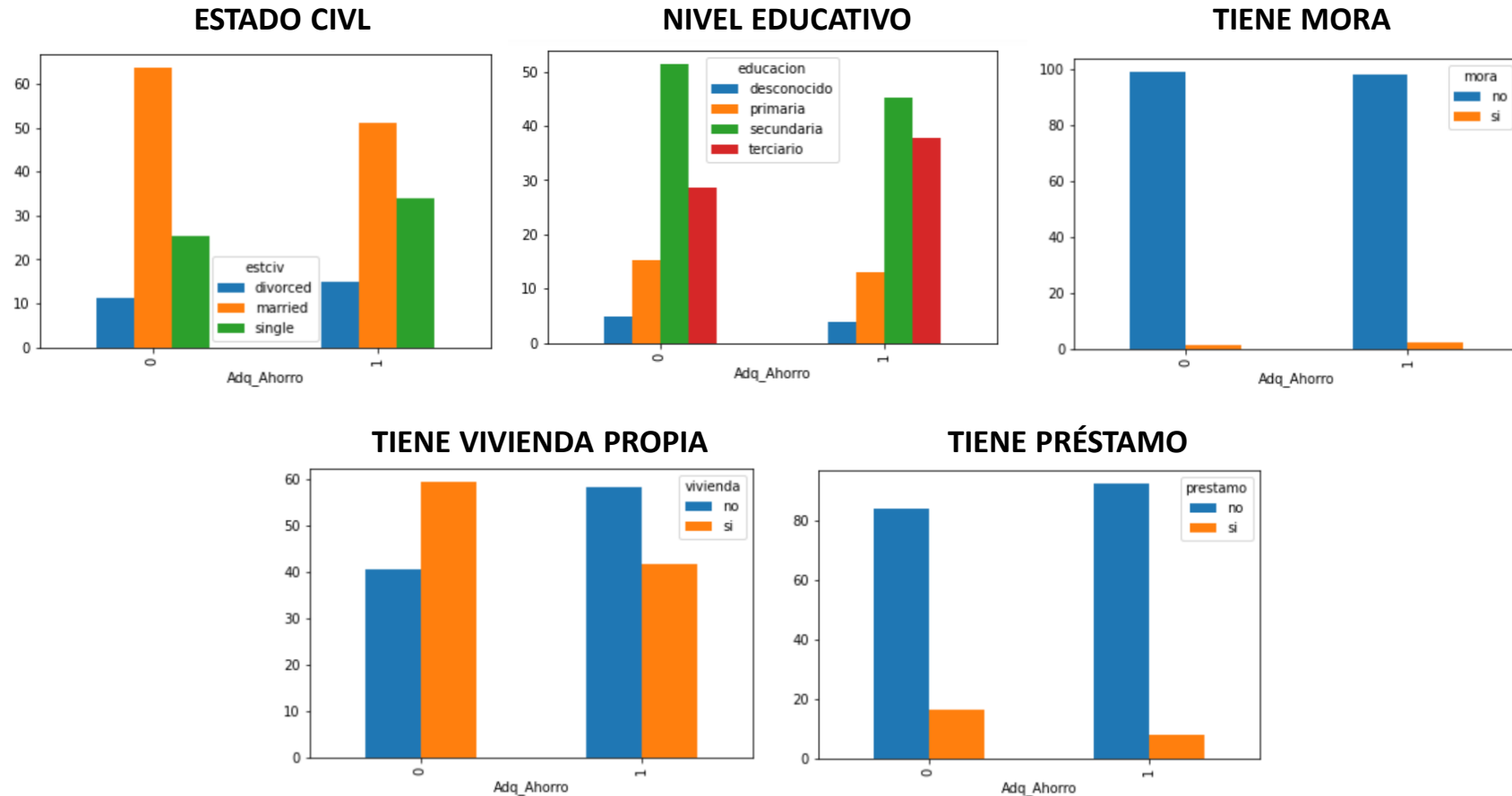
DATASET: ESTADÍSTICAS DESCRIPTIVAS

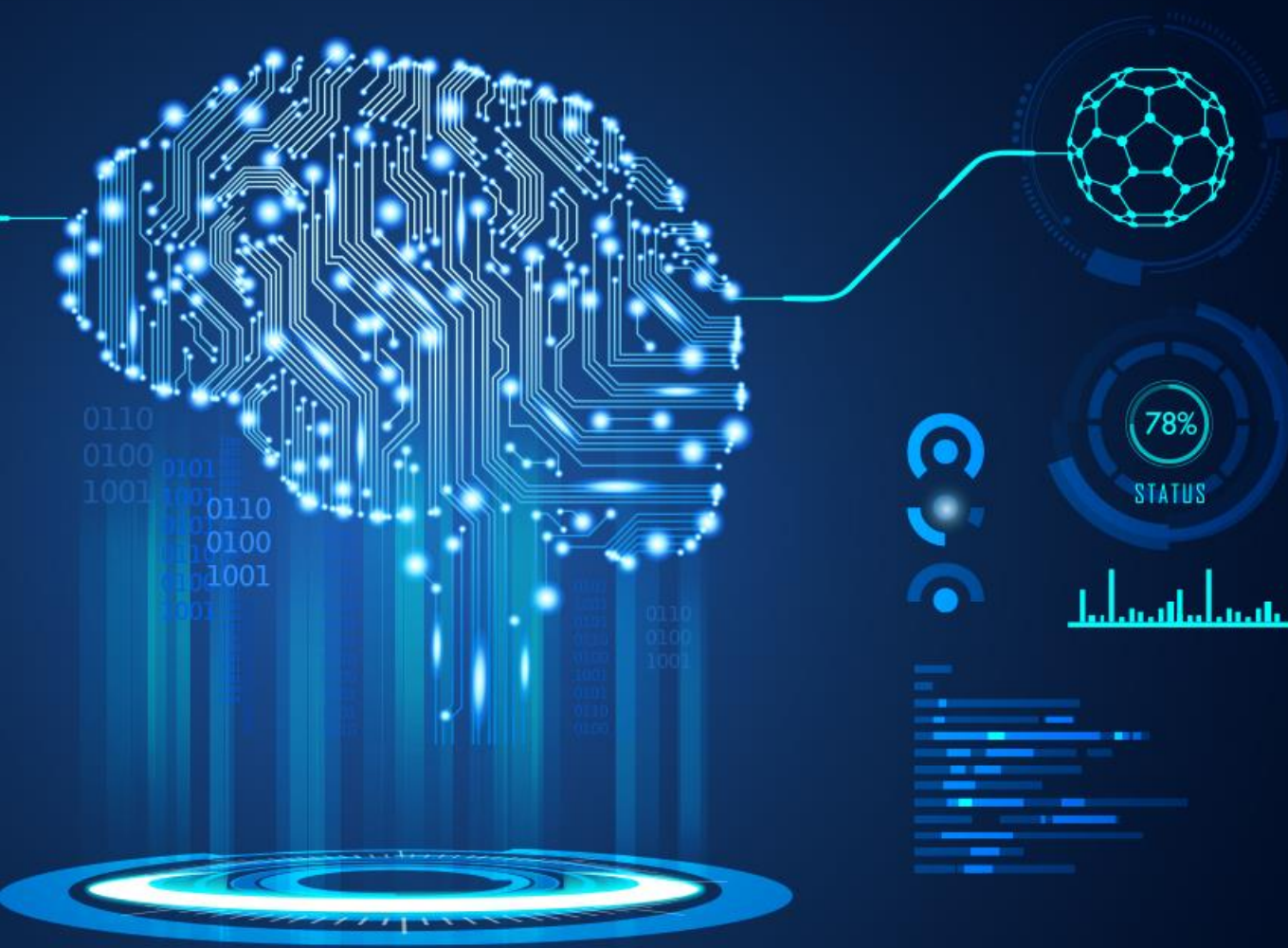
La duración del último contacto parece ser la variable más diferenciadora entre las personas que adquieren un producto de ahorro.



DATASET: ESTADÍSTICAS DESCRIPTIVAS

La tenencia de vivienda propia y el nivel educativo parecen ser las variables más diferenciadora entre las personas que adquieren un producto de ahorro.





ÍNDICE

1. **Comprensión del negocio**
2. **Preparación del Datsaset**
3. **Análisis descriptivo**
4. **Limpieza de los datos**
5. **Elaboración de Machine Learning**
6. **Anexos**

PASO N°1: IMPUTACIÓN DE VARIABLES MISSING

Se considera la mediana para imputar variables numéricas y la moda para variables categóricas.

PASO N°2: TRANSFORMACIÓN DE VARIABLES

Las variables edad, día, campana, pdías, previo y duración deben ser interger y no float64.

PASO N°3: LABEL ENCODER

Se traslada a lenguaje de programación todas las variables categóricas.

PASO N°4: LABEL ENCODER

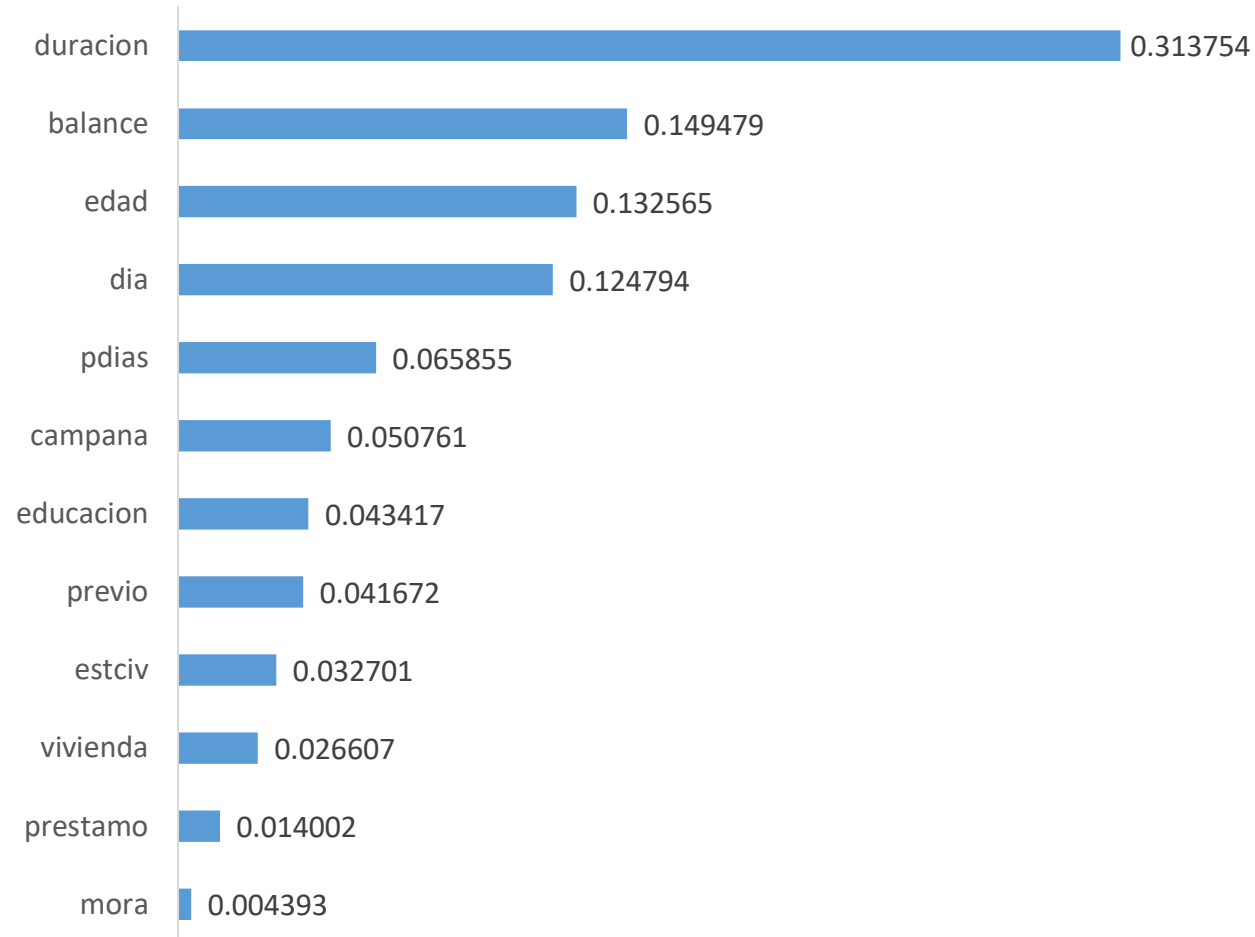
Se realiza análisis gráficos para observar outliers. Sin embargo, el presente proyecto considera la totalidad de las variables.

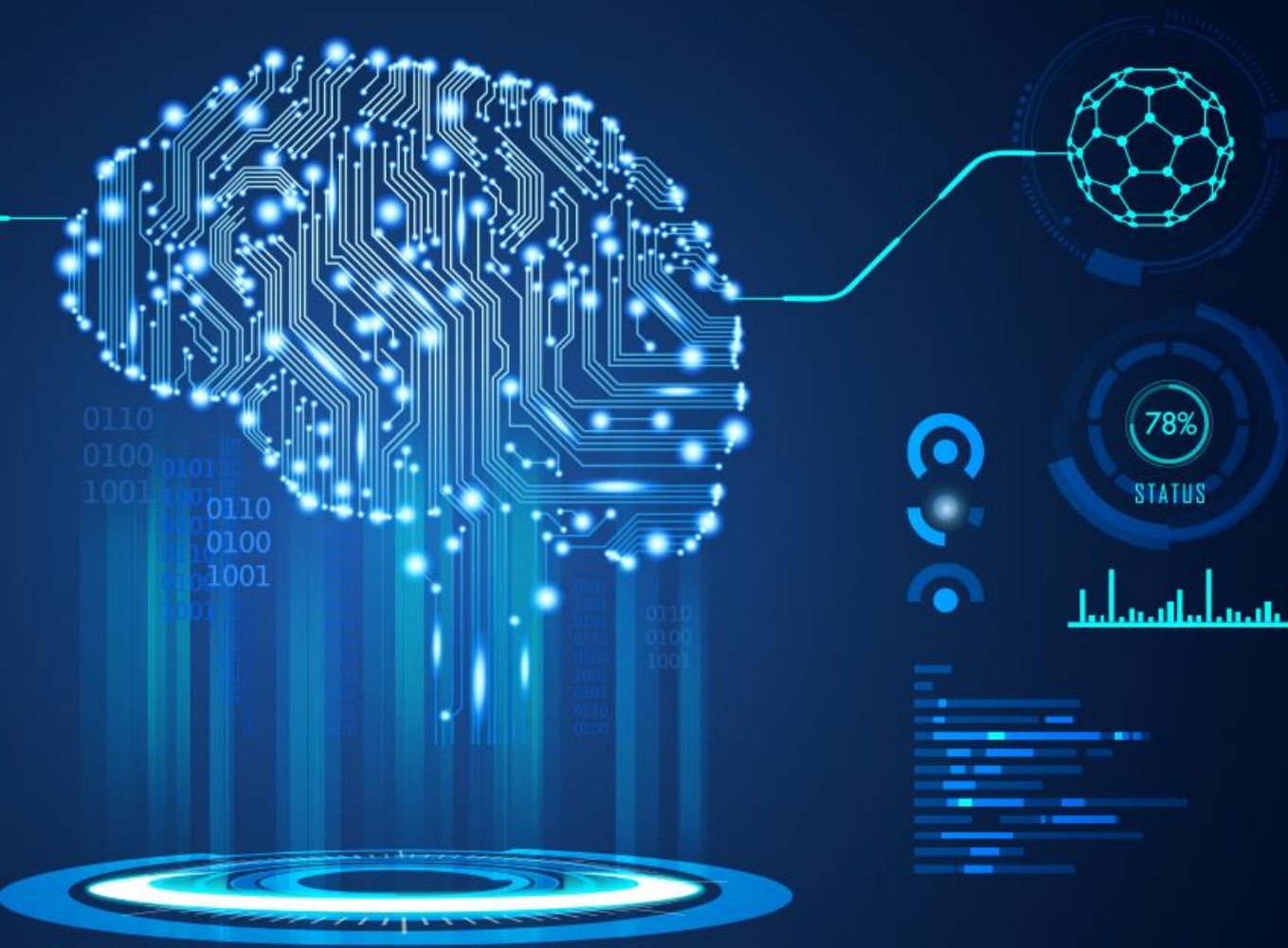
PASO N°5: BALANCEO DEL DATASET

Se balancea al número menor , teniendo como total (516,13)

PASO N°6: SELECCIÓN DE VARIABLES

Identificamos primero las variables que discriminan más, las cuales son duración, balance, edad y día





ÍNDICE

1. **Comprensión del negocio**
2. **Preparación del Datsaset**
3. **Análisis descriptivo**
4. **Limpieza de los datos**
5. **Elaboración de Machine Learning**
6. **Anexos**

RESULTADOS DEL MODELO

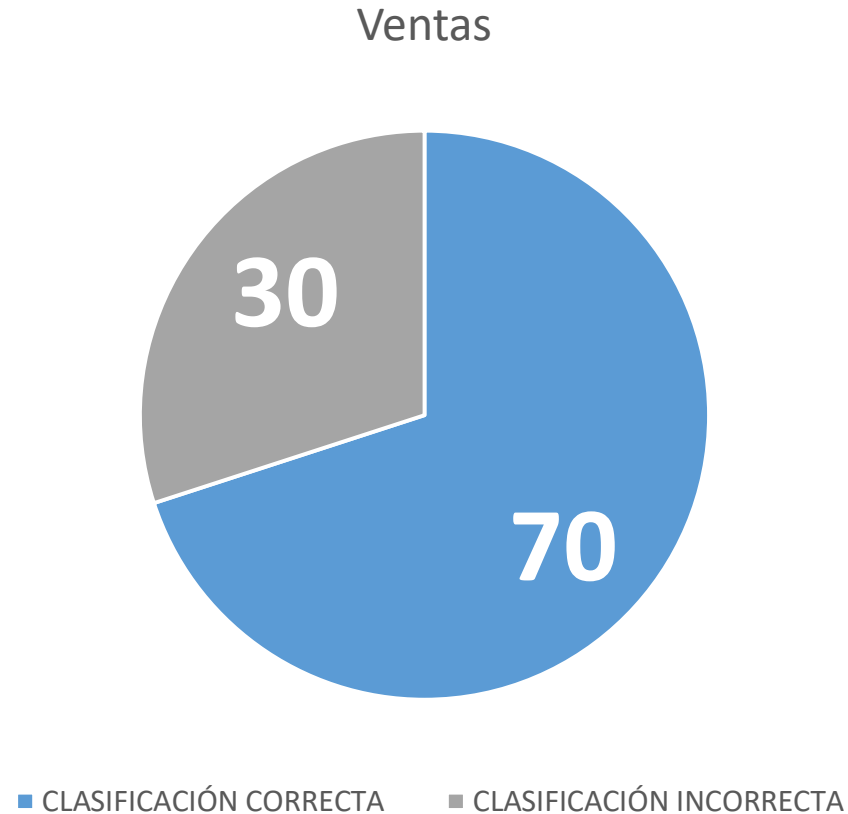
Después de varios ensayos desarrollamos el siguiente modelo seleccionando variables entre la importancia y la observación descriptiva. Con buenos indicadores y pocas diferencias entre los resultados del Train y el Test

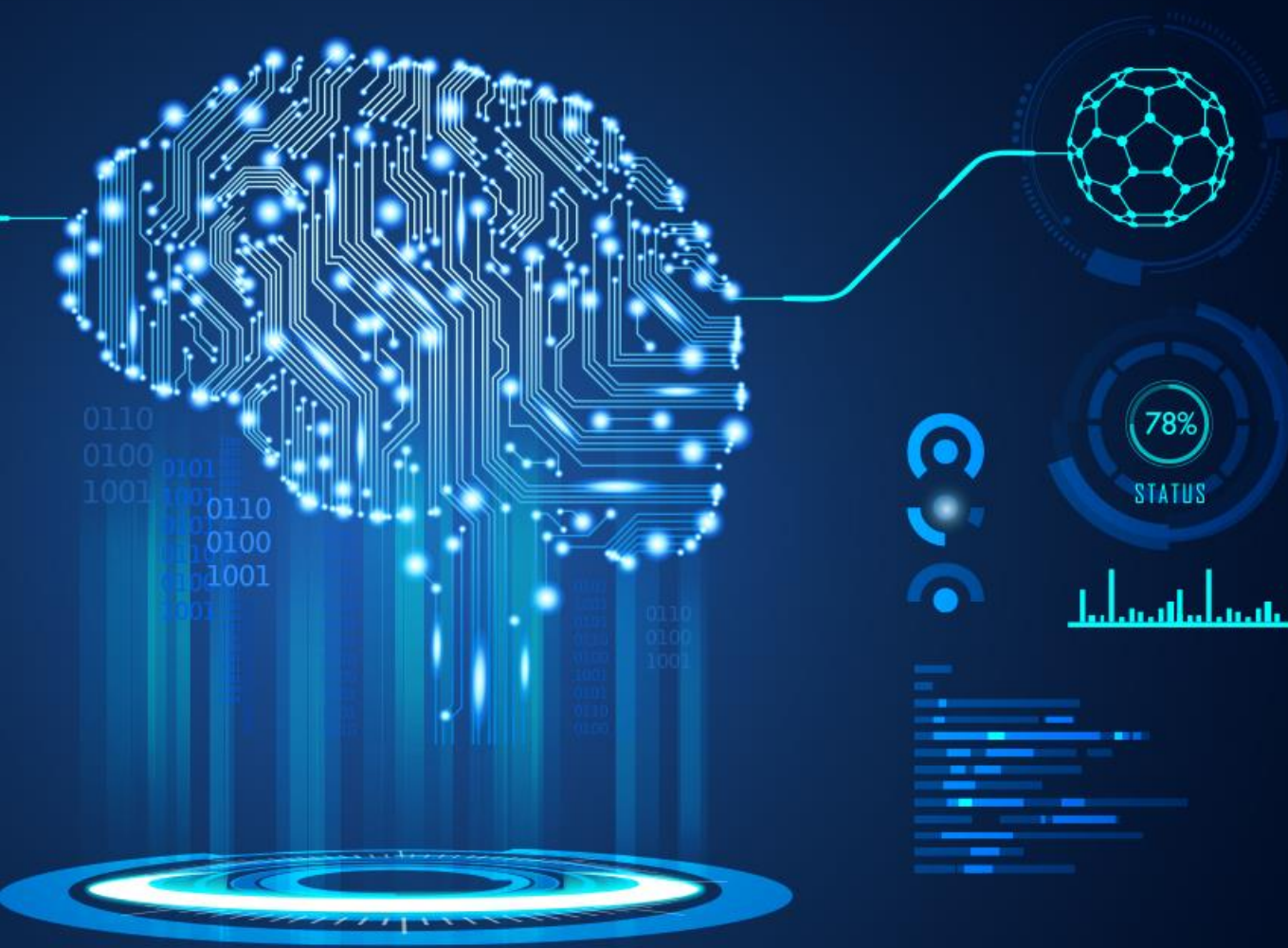
Partición	Indicadores		
	Accuracy	Precisión	Recall
Train	73%	80%	66%
Test	74%	76%	69%
Diferencias	1%	4%	3%

$$P(x) = \frac{1}{1 + e^{-(-1.4805 - 0.0041 \times campana + 0.0001 \times balance + 0.0147 \times edad - 0.9163 \times vivienda - 0.0855 \times education)}}$$

RESULTADOS DEL MODELO

Al aplicar el modelo se espera que se logre clasificar de forma correcta a un 70% de clientes entre los que aceptarían la campaña y los que la rechazarían para hacer que los resultados más eficientes en ventas.



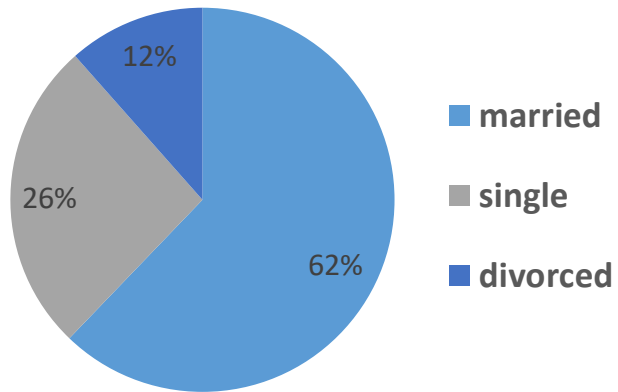


ÍNDICE

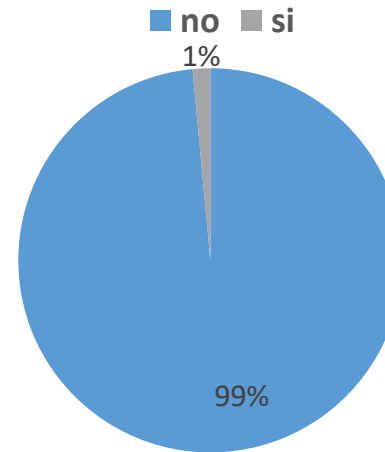
1. Comprensión del negocio
2. Preparación del Datsaset
3. Análisis descriptivo
4. Limpieza de los datos
5. Elaboración de Machine Learning
6. Anexos

ANEXO1: ESTADÍSTICAS DESCRIPTIVAS DE VARIABLES CATEGÓRICAS

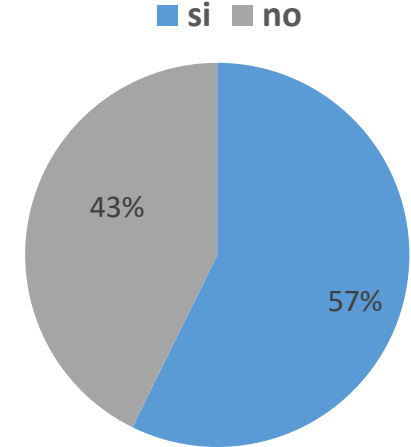
ESTADO CIVIL



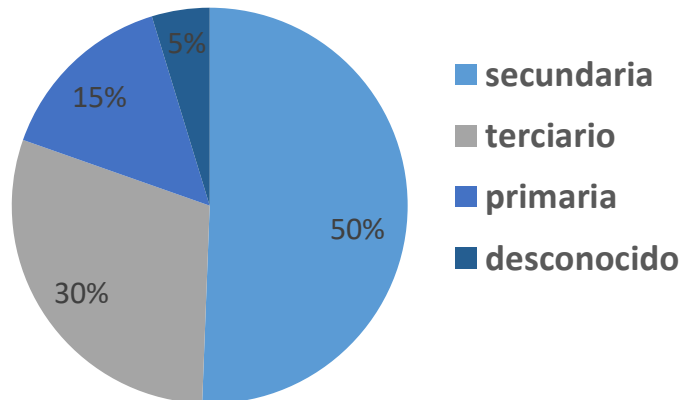
TIENE MORA



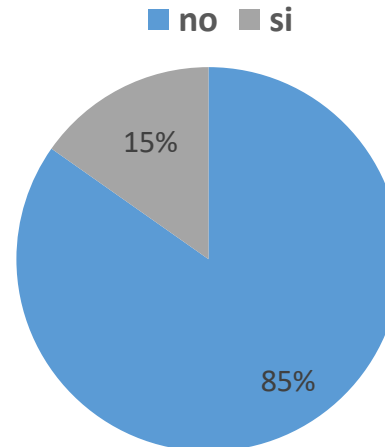
TIENE VIVIENDA PROPIA



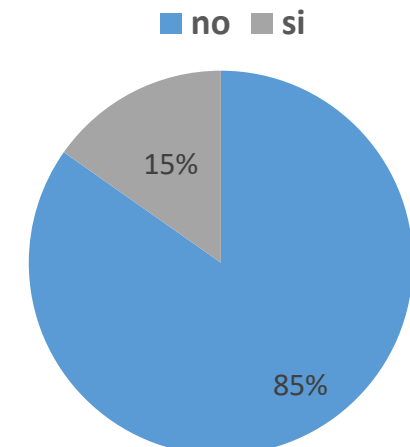
NIVEL EDUCATIVO



ADQUIRIÓ UN PRODUCTO DE AHORRO



TIENE PRÉSTAMO



ANEXO2: ESTADÍSTICAS DESCRIPTIVAS DE VARIABLES CONTÍNUAS

