



MARKETING BANCARIO

Depósitos a Plazos. Tener o no tener, esa es la cuestión.

QUIEN SOY ?



LEONARDO GONZALEZ

Ingeniero en Negocio -
economía

2012 Universidad Católica Silva Henríquez.

Magister Políticas Públicas

2014 Universidad Diego Portales (congelado)

Desarrollo Front End

2020 Coding Dojo - Yellow Belt

Data Science

2021 Acámica - Digital House.

Ingeniero Comercial licenciado
en economía con más de ocho
años de experiencia en
proyectos de analítica financiera
para diferentes rubros.

Stack tecnológico:

python, Spotfire BI, Excel, SAP BO,
BI open source, SQL, JDE y otros.



TABLA DE CONTENIDOS

**PROBLEMA Y
DATOS
DISPONIBLES**

01

02

**ANÁLISIS
DESCRIPTIVO
DE DATOS**

**DATA
WORKFLOW**

03

04

CONCLUSIONES



01

PROBLEMA Y DATOS DISPONIBLES

PROBLEMA

Problema. Predecir si el cliente suscribe un depósito a plazo (y).

Datos Disponibles

- Los datos están relacionados con campañas de marketing directo de una institución bancaria.
- Las campañas de marketing se basaron en llamadas telefónicas.
- A menudo, se requería más de un contacto con el mismo cliente, para poder acceder si el producto (depósito a plazo bancario) estaría ('sí') o no ('no') suscrito.

DATOS DISPONIBLES

CLIENTE

- Edad
- Trabajo
- Estado civil
- Educación

BANCARIO

- Incumplimiento
- Hipoteca
- Prestamo

CONTACTO

- Forma de contacto
- Mes
- Día
- Duración del contacto

CAMPAÑA

- Campaña mkt
- Días último contacto
- Resultado de campaña

40.000 filas
17 columnas

RESULTADO

- Depósito a plazo suscrito o no (y)

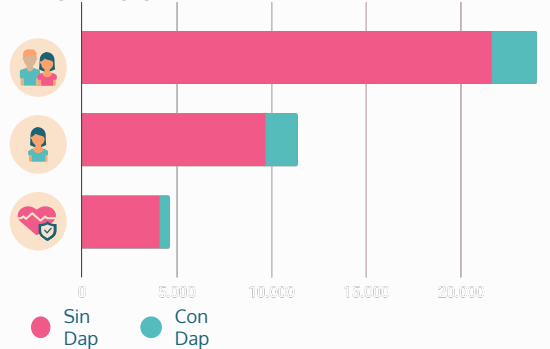
02

ANÁLISIS DESCRIPTIVO DE DATOS

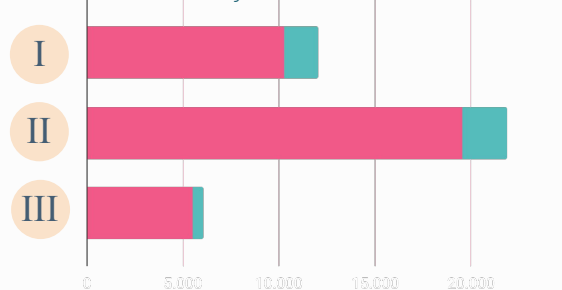


CLIENTES SEGÚN LOS DATOS

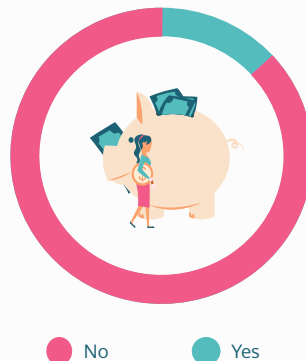
ESTADO CIVIL (Mayoritariamente casados)



EDUCACION (Mayoritariamente educación media)



PRODUCTO



11%

De los clientes contactados Si contrata el producto

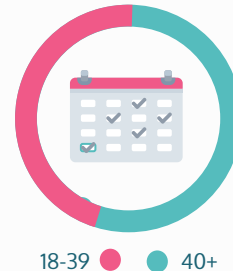
HIPOTECA



55%

Cuenta con un crédito hipotecario.

EDAD



64%

De los clientes es mayor 40 años

EMPLEO Y ESTADO BANCARIO

¿Cómo son los clientes más propensos a tomar depósitos a plazo?

● Sin Dap
● Con Dap

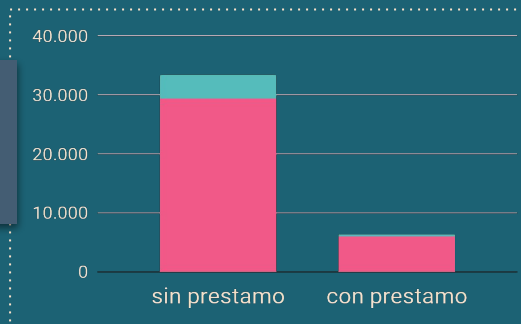
TIPO DE EMPLEO VS DAP.



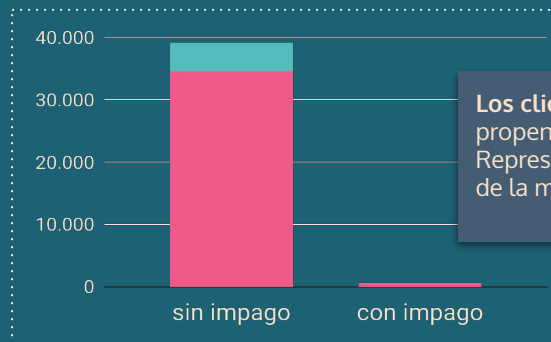
Clientes sin préstamos personales son quienes mayormente toman un depósito a plazo.

Empleados con cargos de mayor preparación son quienes mayormente contratan DAP.

DEUDA CLIENTE VS DAP.



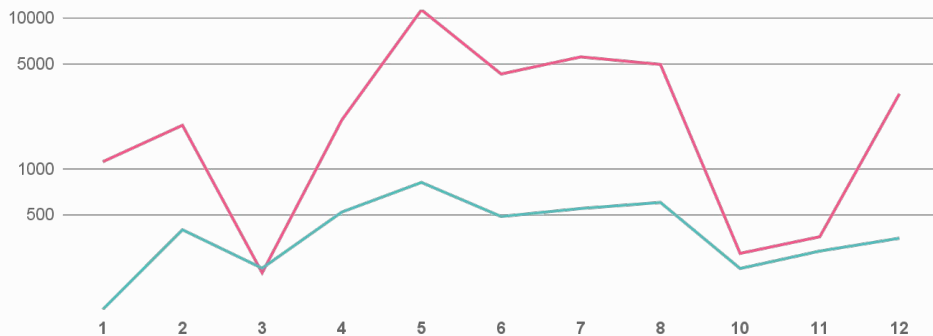
DEFAULT VS DAP.



Los clientes sin impago son más propensos a contratar DAP. Representan además mayor parte de la muestra.

CLIENTES EN EL TIEMPO

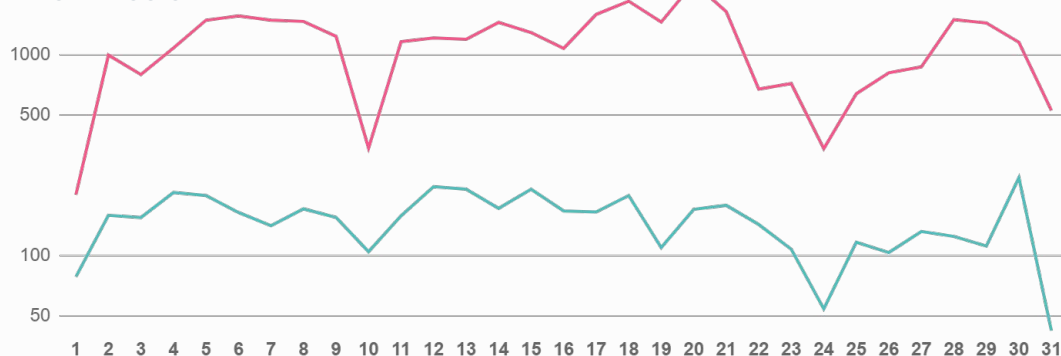
DISTRIBUCIÓN MENSUAL



El contrato de productos de ahorro durante el año tiene **mejor resultado entre abril y octubre**. En tanto el inicio y final de año no parecen ser fechas atractivas.

En cuanto a la distribución por día del mes podemos destacar que **no existe un día preferido para tomar el producto**.

DISTRIBUCIÓN DIARIA



ESTRUCTURA DE DATOS

Los datos están compuestos por variables categóricas ordinales, binarias y numéricas.



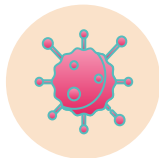
DESBALANCE

Existe un problema de balance de datos, donde predomina un 89% de clientes que no contratan DAP.



FALTANTES

Existen categorías que contiene el valor "unknown", lo que indica que son ausentes y deben ser abordados.



MULTIVARIABLES

Es necesario transformar los datos a un formato estándar.



	age	job	marital	education	default	balance	housing
0	46	management	single	tertiary	no	593	yes
1	42	admin.	married	tertiary	no	1536	no
2	33	blue-collar	married	secondary	no	370	yes
3	29	blue-collar	single	secondary	no	1472	no
4	29	technician	married	secondary	no	767	yes



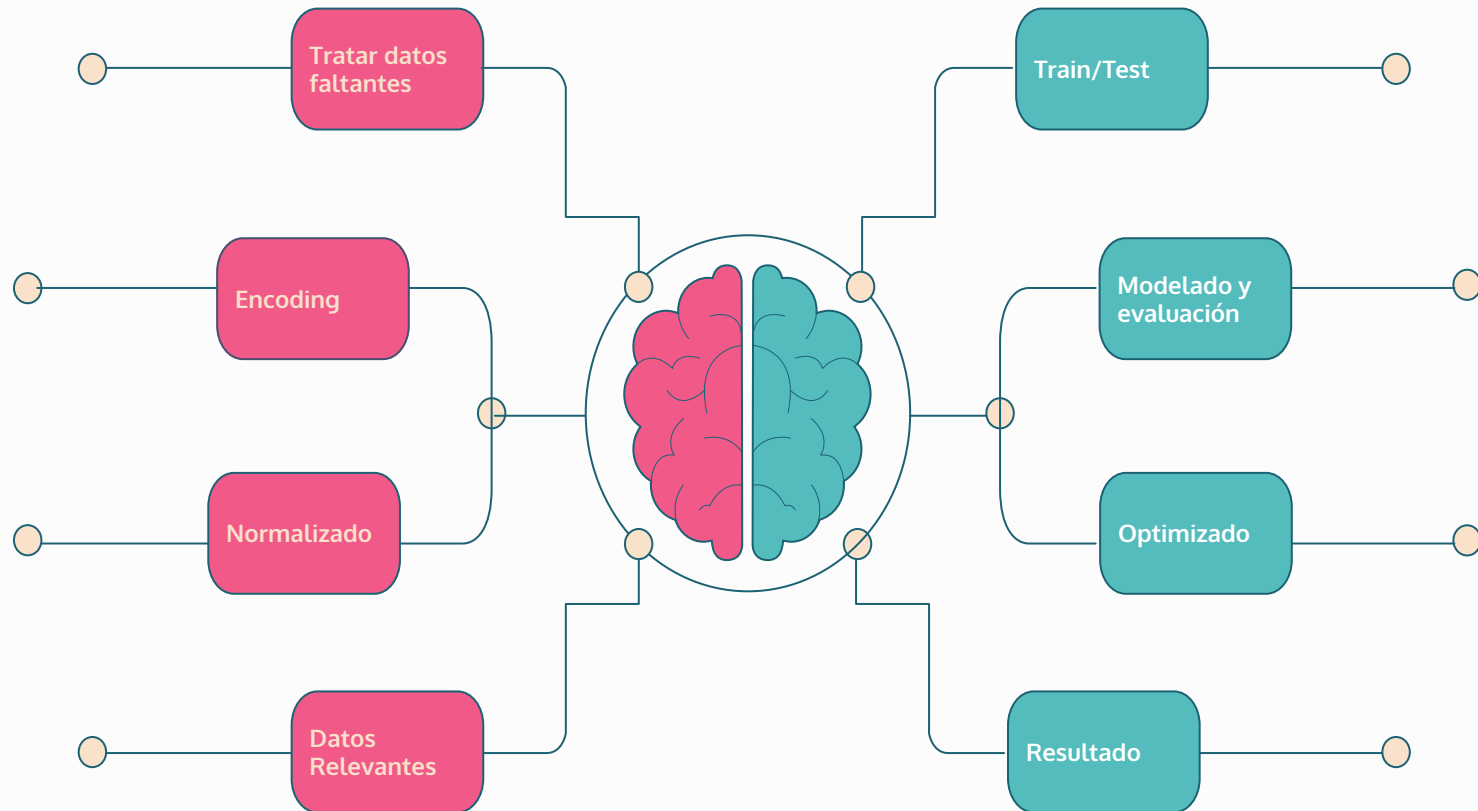
03

DATA WORKFLOW



FLUJO DE TRABAJO

PRE PROCESO



MODELO

Hitos relevantes.

Faltantes



Poutcome 81% de faltantes

Contact 29% faltantes

Education 4% de faltantes

Empleo 0,4% de faltantes

Encoding



Educación: I, II y III + faltantes imputados por su trabajo.

Empleo: se imputo por la moda.

Contacto: fue imputado por la moda.

Binarias: Yes = 1 ; No = 0

Mes: Se categorizó según el mes 1 a 12.

Trabajo y Estado Civil: fueron tratados como dummies.

Normalizado



[Edad , Balance , Campaña ,
Pdays, Previous.]



Balanceo

Sampling: realizamos una estrategia de Oversampling para ampliar la muestra de valor minoritarios en una proporción 60/30, respecto de un 90/10 original.



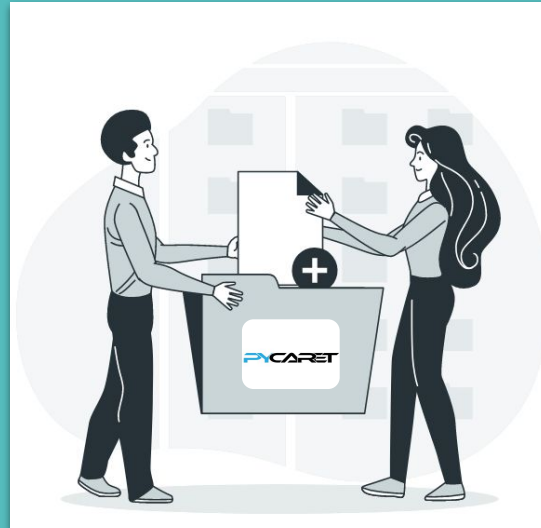
Modelado

Pycaret: Librería de código abierto que permite evaluar varios modelos a la vez, utilizando el proceso. Se abordará como un problema de clasificación.

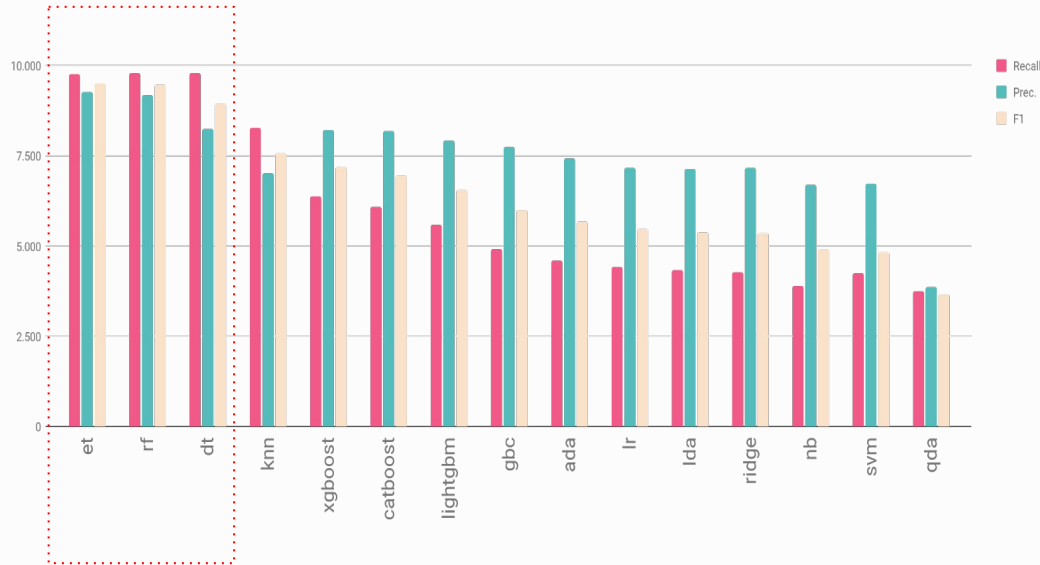


Optimizado

Mejoras: Buscaremos el modelo con mejor F1 e intentaremos mejorar.



PRIMEROS RESULTADO



Parámetros y rendimiento.

- Los modelos de Árboles y bosques son los mejores evaluados para este problema.
- El mejor F1 lo obtiene el modelo Extra Tree Classifier y Random Forest seguidos por un Árbol de decisiones.
- Utilizaremos et y rf para evaluar y crear nuestros modelos.

Model	Recall	Prec.	F1
et	9.761	9.279	9.514
rf	9.786	9.178	9.472
dt	9.779	8.249	8.949

CREANDO MODELO

EXTRA TREE

	Recall	Prec.	F1
0	0,9805	0,9167	0,9475
1	0,9746	0,9392	0,9566
2	0,9752	0,9387	0,9566
3	0,9799	0,9347	0,9568
4	0,9776	0,9288	0,9525
5	0,9752	0,925	0,9494
6	0,9799	0,9248	0,9516
7	0,9711	0,9175	0,9435
8	0,9717	0,9242	0,9474
9	0,9752	0,9292	0,9516
Mean	0,9761	0,9279	0,9514
SD	0,0032	0,0075	0,0043

Observamos que la media de F1 del moldeo "et" es 9,5 muy poco por sobre "rf" de 9,4.



La diferencia entre los modelos creados en términos de su F1 no es altamente significativo utilizando un despliegue en 10. Por tanto es de esperar que ambos modelos obtienen resultados similares.

RANDOM FOREST

	Recall	Prec.	F1
0	0,9841	0,92	0,951
1	0,9782	0,9242	0,9504
2	0,9776	0,9267	0,9515
3	0,9829	0,9297	0,9555
4	0,9799	0,9146	0,9461
5	0,9764	0,9204	0,9476
6	0,9799	0,9131	0,9453
7	0,9746	0,8968	0,9341
8	0,9758	0,9169	0,9454
9	0,9764	0,9159	0,9452
Mean	0,9786	0,9178	0,9472
SD	0,0029	0,0087	0,0054

OPTIMIZANDO

EXTRA TREE

	Recall	Prec.	F1
0	0,6171	0,6057	0,6113
1	0,6018	0,5913	0,5965
2	0,6065	0,6044	0,6054
3	0,6242	0,5967	0,6101
4	0,6346	0,5724	0,6019
5	0,6216	0,5946	0,6078
6	0,6068	0,5898	0,5982
7	0,6057	0,5816	0,5934
8	0,6083	0,6019	0,605
9	0,6029	0,5877	0,5952
Mean	0,6129	0,5926	0,6025
SD	0,0103	0,0099	0,0061

Al intentar optimizar nuestros modelos no obtenemos mejores resultados que en la primera versión.



Este proceso de búsqueda de hiperparámetros se realizó bajo parámetros por defecto lo que no quiere decir que podemos realizarlo de manera manual.

RANDOM FOREST

	Recall	Prec.	F1
0	0,5794	0,649	0,6122
1	0,5853	0,6446	0,6135
2	0,5711	0,6594	0,6121
3	0,5776	0,657	0,6148
4	0,5738	0,6506	0,6098
5	0,5891	0,6587	0,622
6	0,5685	0,6373	0,6009
7	0,562	0,6292	0,5937
8	0,5634	0,651	0,604
9	0,5699	0,6355	0,6009
Mean	0,574	0,6472	0,6084
SD	0,0084	0,0099	0,0079

PREDICCIONES



EXTRA TREE

F1

0.96

Utilizamos los datos reservados para la prueba que mediante el **método predict** nos permite conocer la probabilidad de la clase predicha usando el modelo entrenado.

Ambos modelos obtiene un buen resultado, con un mejor desempeño del modelo "et". Sin embargo probaremos ambos en los datos de testeo que no conocemos.

```
ExtraTreesClassifier(bootstrap=False, ccp_alpha=0.0, class_weight=None,
criterion='gini', max_depth=None, max_features='auto',
max_leaf_nodes=None, max_samples=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, n_estimators=100, n_jobs=-1,
oob_score=False, random_state=123, verbose=0,
warm_start=False)
```



RANDOM FOREST

F1

0.95

```
RandomForestClassifier(bootstrap=True, ccp_alpha=0.0, class_weight=None,
criterions='gini', max_depth=None, max_features='auto',
max_leaf_nodes=None, max_samples=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, n_estimators=100,
n_jobs=-1, oob_score=False, random_state=123, verbose=0,
warm_start=False)
```



04

CONCLUSIONES



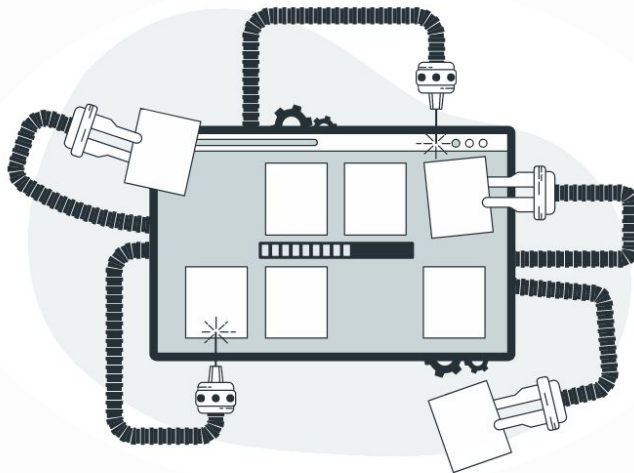
CONCLUSIONES

DATOS

Si bien los datos tienen características e información de clientes sería interesante contar con otros datos que nos permitan dar mas informacion a nuestros modelos com: Genero, Fechas, Montos DAP, Ubicación y Tasas de interés.

PROCESO

Para este problema se abordaron diferentes preprocesamiento de datos como también diferentes parametrizaciones de los modelos, sin embargo estos no tuvieron mejores resultado que los expuestos. Sin embargo esto no está exento de mejoras.



MODELOS

Los modelos utilizados fueron guiados por los resultado de Pycaret y no se abordaron otros modelos manuales que pudieran mejorar el performace de nuestros resultados.

NEGOCIO

Los clientes evaluados en su mayoría no contratan DAP lo que representa una realidad para el negocio y esto puede estar explicado por otras razones que este set de datos no logra explicar como pueden ser facilidades de contratación, entendimiento del producto, impuestos, tasas de interés entre otros.

THANKS!

Leonardo Gonzalez.

