

GABRIEL BARDAJÍ BIESCAS 1568297

HEALTH INSURANCE CROSS SELL PREDICTION

APRENTATGE COMPUTACIONAL



ÍNDEX

- 01 INTRODUCCIÓ AL DATASET**
- 02 ANÀLISI DELS ATRIBUTS**
- 03 FEATURE SELECTION**
- 04 DIVISIÓ TRAIN-TEST I SELECCIÓ HIPERPARÀMETRES**
- 05 MODEL SELECTION**
- 06 CONCLUSIONS**



INTRODUCCIÓ AL DATASET

- Recopilació de dades sobre els clients d'una asseguradora de vehicles.
- **381.109** files i **12** columnas

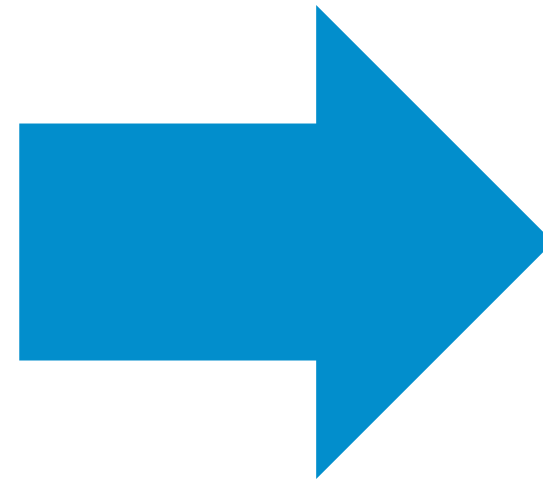


Gender	Age	Driving_License	Region_Code	Previously_Insured	Vehicle_Age	Vehicle_Damage	Annual_Premium	Policy_Sales_Channel	Vintage	Response
Male	44	1	28.0	0	> 2 Years	Yes	40454.0	26.0	217	1
Male	76	1	3.0	0	1-2 Year	No	33536.0	26.0	183	0
Male	47	1	28.0	0	> 2 Years	Yes	38294.0	26.0	27	1
Male	21	1	11.0	1	< 1 Year	No	28619.0	152.0	203	0
Female	29	1	41.0	1	< 1 Year	No	27496.0	152.0	39	0



ANÀLISI DELS ATRIBUTS

id	int64
gender	object
age	int64
driving_license	int64
region_code	float64
previously_insured	int64
vehicle_age	object
vehicle_damage	object
annual_premium	float64
policy_sales_channel	float64
vintage	int64
response	int64



id	0
Gender	0
Age	0
Driving_License	0
Region_Code	0
Previously_Insured	0
Vehicle_Age	0
Vehicle_Damage	0
Annual_Premium	0
Policy_Sales_Channel	0
Vintage	0
Response	0
dtype: int64	
(381109, 12)	



ANÀlisi DELS ATRIBUTS - TRACTAMENT DE VARIABLES CATEGÒRIQUES

1

VEHICLE_AGE

- '> 2 Years' = 0
- '1-2 Year' = 1
- '< 1 Year' = 2

2

VEHICLE_DAMAGE

- Yes = 1
- No = 0

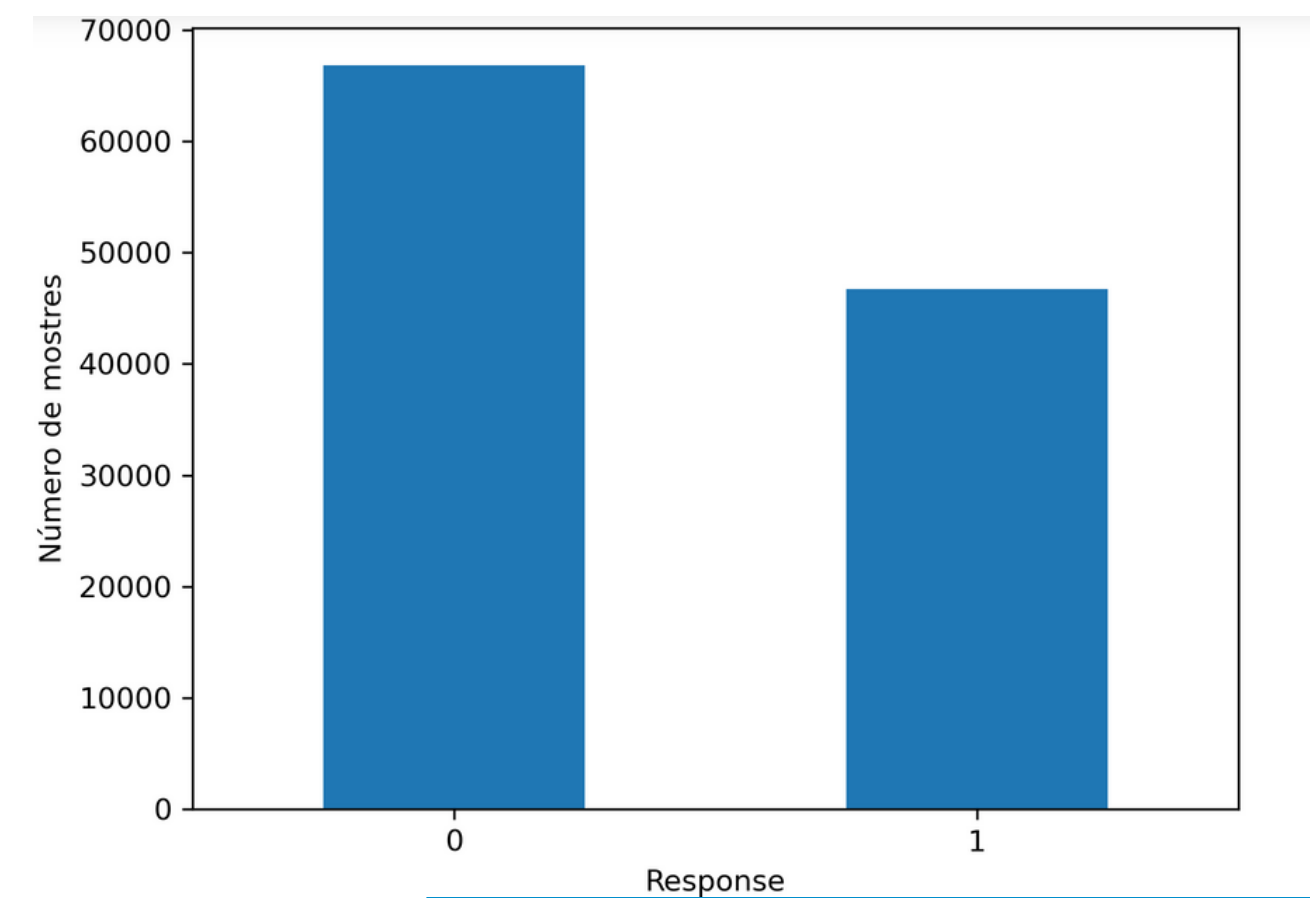
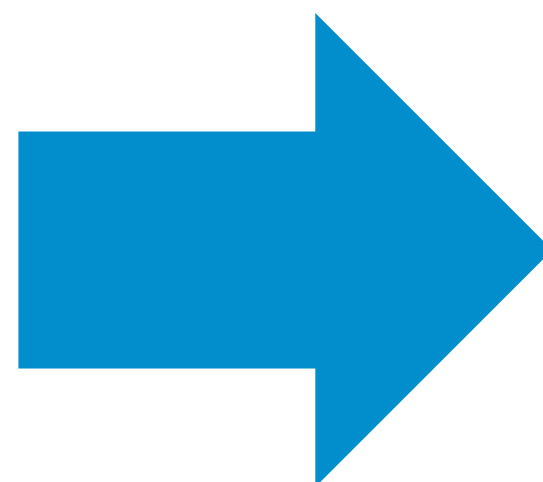
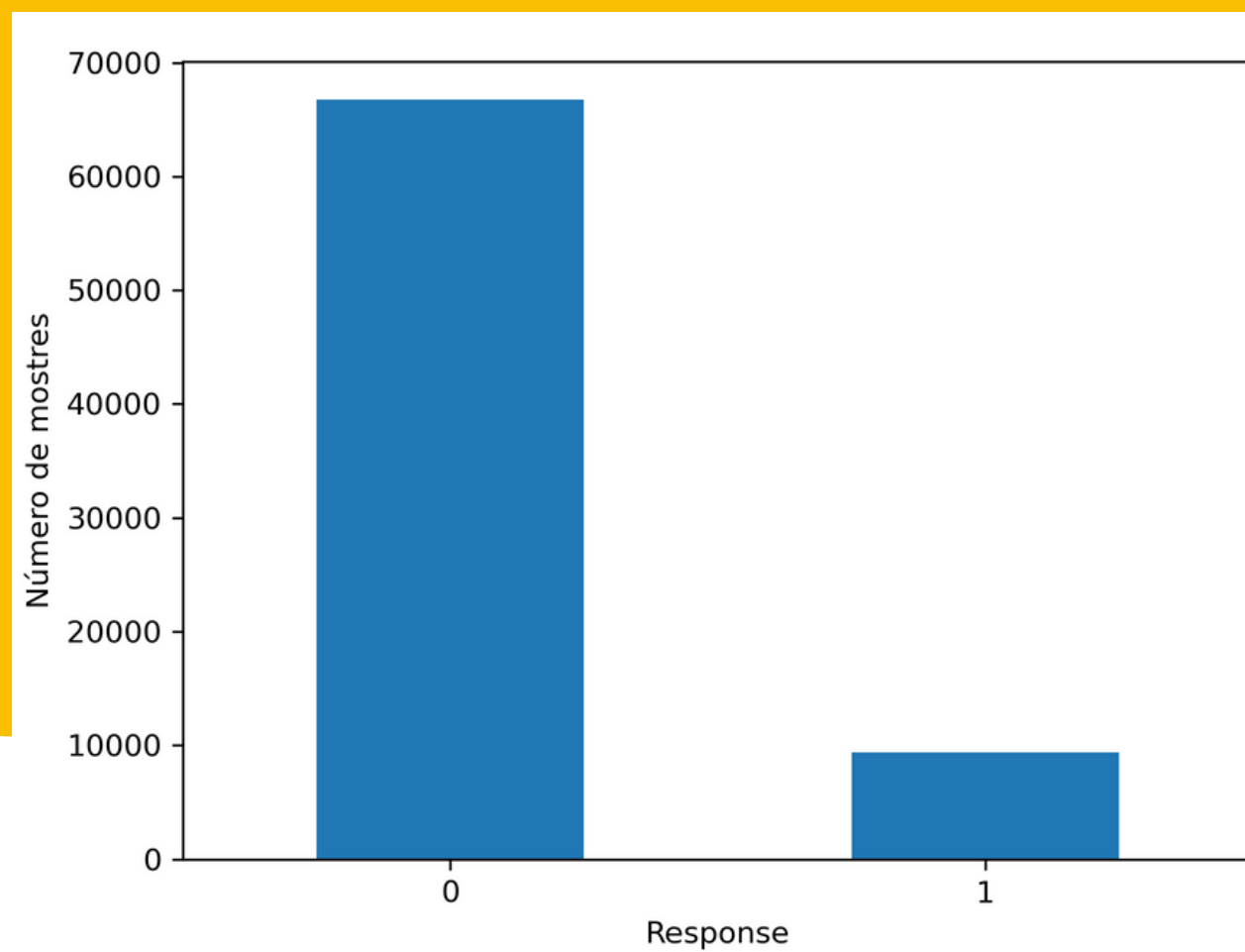
3

GENDER

- Male = 1
- Female = 0



ANÀLISI DELS ATRIBUTS





FEATURE SELECTION

Gender	1	0.15	0.02	0	0.08	0.16	0.09	0	0.11	0	0.05
Age	0.15	1	0.08	0.04	0.25	0.77	0.27	0.07	0.58	0	0.11
Previously_Insured	0.02	0.08	1	0	0.01	0.04	0.02	0.01	0.04	0	0.01
Region_Code	0	0.04	0	1	0.02	0.04	0.03	0.01	0.04	0	0.01
Previously_Insured	0.08	0.25	0.01	0.02	1	0.38	0.82	0	0.22	0	0.34
Vehicle_Age	0.16	0.77	0.04	0.04	0.38	1	0.4	0.04	0.55	0	0.22
Vehicle_Damage	0.09	0.27	0.02	0.03	0.82	0.4	1	0.01	0.22	0	0.35
Annual_Premium	0	0.07	0.01	0.01	0	0.04	0.01	1	0.11	0	0.02
Sales_Channel	0.11	0.58	0.04	0.04	0.22	0.55	0.22	0.11	1	0	0.14
Vintage	0	0	0	0	0	0	0	0	0	1	0
Response	0.05	0.11	0.01	0.01	0.34	0.22	0.35	0.02	0.14	0	1
	Gender	Age	Previously_Insured	Region_Code	Previously_Insured	Vehicle_Age	Vehicle_Damage	Annual_Premium	Sales_Channel	Vintage	Response

Age - 0.11

Previously_Insured - 0.34

Vehicle_Age - 0.22

Vehicle_Damage - 0.35

Policy_Sales_Channel 0.14



FEATURE SELECTION

1 DESBALANCEJAT AMB TOTS ELS ATRIBUTS

2 DESBALANCEJAT AMB TOTS ELS ATRIBUTS

3 BALANCEJAT AMB TOTS ELS ATRIBUTS

4 BALANCEJAT AMB TOTS ELS ATRIBUTS



FEATURE SELECTION

SMALL-DATASET

'Number of rows (clients) for small-dataset: 7622'

'Number of columns (characteristics): 12'



DIVISIÓ TRAIN-TEST I SELECCIÓ HIPERPARAMETRES

ESTANDARITZACIÓ



TRAIN-TEST



RANDOMIZEDSEARCH



VISUALITZACIÓ DELS MÈTODESD'APRENENTATGE

	DESBALANCEJADA		BALANCEJADA	
	TOTS ATR	MÉS CORR	TOTS ATR	MÉS CORR
KNN	0.8752	0.8760	0.7685	0.7721
LOGISTIC REGRESSION	0.8770	0.8768	0.7449	0.7449
GRADIENT DESCENT	0.8774	0.8775	0.7599	0.7584
RANDOM FOREST	0.8655	0.7253	0.7614	0.7692
NAIVE BAYES	0.63833	0.6461	0.7490	0.7498



CONCLUSIONS



ENLLAÇ AL GITHUB



https://github.com/biel1703/Kaggle_individual_ApC

**GRÀCIES PER LA
VOSTRA ATENCIÓ**