

# Proyecto Entrega 1

## 1. Describir el problema predictivo a resolver

Actualmente un banco quiere determinar cuales de sus clientes aceptaran adquirir un CDT a un plazo determinado y cuales no, debido a que desean invertir en campañas de marketing para lograr captar la mayor cantidad de recursos. Partiendo de lo anterior se requiere un modelo de clasificación que logre diferenciar entre los clientes con alta probabilidad de aceptar y aquellos que no.

## 2. Dataset que se va a utilizar.

Los datos a utilizar se encuentran en el siguiente enlace:

<https://www.kaggle.com/competitions/prueba-oficina-de-datos/overview>

En el anterior podemos encontrar los datos de entrenamiento, los cuales se encuentran organizados de la siguiente manera:

Train(1).csv

ID	Edad	Tipo_Trabajo	Estado_Civil	Educacion	Incumplimiento	Vivienda	Consumo	Contacto	Mes	Dias	Campana	Dias_Ultima_Camp	No_Contactos	Resultado_Anterior	emp_var_rate	cons_price_idx	cons_conf_idx	...		
14110	44	admin.	married	university.degree	no	yes	yes	cellular	jul	mon	4	999	0	nonexistent	1.4	93918.0	-42.7	4962.0	5228.1	0
13250	38	blue-collar	married	high.school	no	yes	no	cellular	jul	wed	5	999	0	nonexistent	1.4	93918.0	-42.7	4963.0	5228.1	0
13728	52	blue-collar	single	unknown	unknown	no	yes	cellular	jul	fri	11	999	0	nonexistent	1.4	93918.0	-42.7	4962.0	5228.1	0
27503	40	blue-collar	married	basic.9y	no	yes	no	cellular	may	thu	1	999	1	failure	-1.8	92893.0	-46.2	1266.0	5099.1	0
17265	31	admin.	single	university.degree	no	no	yes	cellular	aug	tue	2	999	0	nonexistent	1.4	93444.0	-36.1	4963.0	5228.1	0
25436	33	admin.	single	university.degree	no	yes	no	cellular	may	thu	4	999	1	failure	-1.8	92893.0	-46.2	1327.0	5099.1	0
29516	29	admin.	single	high.school	no	yes	yes	cellular	jun	mon	2	999	0	nonexistent	-2.9	92963.0	-40.8	1.26	5076.2	1
9633	36	admin.	married	university.degree	no	yes	no	telephone	jun	wed	1	999	0	nonexistent	1.4	94465.0	-41.8	4958.0	5228.1	0
2185	36	admin.	married	university.degree	unknown	yes	no	telephone	may	wed	2	999	0	nonexistent	1.1	93994.0	-36.4	4859.0	5191.0	0
9136	25	blue-collar	single	basic.9y	no	yes	yes	telephone	jun	thu	4	999	0	nonexistent	1.4	94465.0	-41.8	4961.0	5228.1	0
12542	24	services	married	high.school	no	yes	no	cellular	jul	mon	4	999	0	nonexistent	1.4	93918.0	-42.7	4.96	5228.1	0
20972	31	admin.	married	university.degree	no	no	no	cellular	nov	wed	2	999	0	nonexistent	-0.1	93.2	-42.0	4.12	5195.8	0
31055	26	student	single	university.degree	no	no	no	cellular	nov	fri	1	7	1	success	-3.4	92649.0	-30.1	0.7140000000000001	5017.5	1
2947	52	blue-collar	married	basic.4y	no	no	no	telephone	may	fri	2	999	0	nonexistent	1.1	93994.0	-36.4	4859.0	5191.0	0
27021	43	blue-collar	married	basic.4y	no	no	yes	cellular	may	wed	1	999	1	failure	-1.8	92893.0	-46.2	1281.0	5099.1	0
30582	80	retired	divorced	basic.4y	no	no	no	cellular	oct	tue	1	999	0	nonexistent	-3.4	92431.0	-26.9	0.7440000000000001	5017.5	0
11527	52	blue-collar	married	unknown	no	no	yes	cellular	jul	mon	11	999	0	nonexistent	1.4	93918.0	-42.7	4962.0	5228.1	0
25852	40	blue-collar	married	basic.4y	no	no	no	cellular	may	fri	1	999	0	nonexistent	-1.8	92893.0	-46.2	1313.0	5099.1	0
18099	56	blue-collar	married	basic.9y	unknown	no	no	cellular	aug	fri	4	999	0	nonexistent	1.4	93444.0	-36.1	4964.0	5228.1	0
4577	46	admin.	single	high.school	no	no	no	telephone	may	mon	2	999	0	nonexistent	1.1	93994.0	-36.4	4857.0	5191.0	0
29412	28	student	single	high.school	no	yes	no	cellular	jun	fri	3	999	0	nonexistent	-2.9	92963.0	-40.8	1268.0	5076.2	1
31835	56	unemployed	married	basic.9y	no	no	no	cellular	may	mon	1	999	1	failure	-1.8	93876.0	-40.0	0.701	5008.7	1

Además también encontramos los datos de prueba, a los que se les realizará el diagnóstico y aplicaran las predicciones.

## Test(1).csv

ID	Edad	Tipo_Trabajo	Estado_Civil	Educacion	Incumplimiento	Vivienda	Consumo	Contacto	Mes	Dias	Campana	Dias_Ultima_Camp	No_Contactos	Resultado_Anterior	emp_var_rate	cons_price_idx	cons_conf_idx	eu
30785	98	retired	married	basic.4y	unknown	yes	no	cellular	oct	fri	2,999	0	nonexistent	-3.4,92431.0	-26.9	0.73	5017.5	
14243	45	self-employed	divorced	university.degree	no	no	no	cellular	jul	tue	4,999	0	nonexistent	1.4,93918.0	-42.7	4.961	0.5228.1	
20996	41	blue-collar	married	basic.4y	no	yes	no	telephone	nov	thu	3,999	0	nonexistent	-0.1,93.2	-42.0	4.076	0.5195.8	
31958	68	retired	married	basic.4y	no	yes	no	cellular	jun	thu	2,999	1	failure	-1.7,94055.0	-39.8	0.742	4.991.6	
10457	28	services	married	high.school	no	no	yes	cellular	jul	tue	1,999	0	nonexistent	1.4,93918.0	-42.7	4.962	0.5228.1	
19093	31	technician	divorced	university.degree	no	yes	no	cellular	aug	fri	3,999	0	nonexistent	1.4,93444.0	-36.1	4.963	0.5228.1	
14405	37	admin.	single	high.school	no	yes	no	cellular	jul	tue	2,999	0	nonexistent	1.4,93918.0	-42.7	4.961	0.5228.1	
16875	47	admin.	married	university.degree	no	unknown	unknown	cellular	aug	thu	2,999	0	nonexistent	1.4,93444.0	-36.1	4.964	0.5228.1	
30787	26	admin.	single	university.degree	no	no	no	cellular	oct	fri	2,999	1	failure	-3.4,92431.0	-26.9	0.73	5017.5	
6197	31	admin.	single	high.school	no	yes	no	telephone	may	fri	25,999	0	nonexistent	1.1,93994.0	-36.4	4.864	0.5191.0	
5807	38	blue-collar	married	basic.4y	unknown	no	no	telephone	may	thu	6,999	0	nonexistent	1.1,93994.0	-36.4	4.86	5.191.0	
25203	39	blue-collar	single	basic.9y	unknown	yes	no	cellular	may	wed	2,999	1	failure	-1.8,92893.0	-46.2	1.334	0.5099.1	
5492	40	technician	married	professional.course	no	no	yes	telephone	may	wed	2,999	0	nonexistent	1.1,93994.0	-36.4	4.857	0.5191.0	
4992	35	services	married	high.school	unknown	yes	no	telephone	may	tue	8,999	0	nonexistent	1.1,93994.0	-36.4	4.857	0.5191.0	
19430	55	admin.	divorced	university.degree	no	yes	no	cellular	nov	mon	2,999	0	nonexistent	-0.1,93.2	-42.0	4.191	0.5195.8	
8291	32	blue-collar	married	high.school	no	no	yes	telephone	jun	mon	1,999	0	nonexistent	1.4,94465.0	-41.8	4.96	5.228.1	
15930	29	technician	married	professional.course	no	yes	no	cellular	aug	fri	1,999	0	nonexistent	1.4,93444.0	-36.1	4.966	0.5228.1	
30385	39	blue-collar	married	basic.9y	no	yes	no	cellular	sep	fri	1,3	1	success	-3.4,92379.0	-29.8	0.8029999999999999	5.017.5	
24411	52	admin.	single	high.school	no	no	no	cellular	may	mon	2,999	0	nonexistent	-1.8,92893.0	-46.2	1.354	0.5099.1	
17934	50	admin.	married	unknown	no	yes	no	cellular	aug	thu	1,999	0	nonexistent	1.4,93444.0	-36.1	4.963	0.5228.1	
148	38	blue-collar	single	basic.9y	no	yes	no	telephone	may	mon	2,999	0	nonexistent	1.1,93994.0	-36.4	4.857	0.5191.0	
4042	35	admin.	single	university.degree	no	yes	yes	telephone	may	fri	1,999	0	nonexistent	1.1,93994.0	-36.4	4.857	0.5191.0	
6960	59	retired	married	basic.6y	unknown	no	no	telephone	jun	wed	5,999	0	nonexistent	1.4,94465.0	-41.8	4.864	0.5228.1	
32828	31	technician	single	professional.course	no	unknown	unknown	cellular	oct	thu	2,999	0	nonexistent	-1.1,94601.0	-49.5	10.25	0.4963.6	

### 3. Cuales son las métricas de desempeño requeridas (de machine learning y de negocio)

Las propuestas por la competición ROC – AUC

### 4. Un primer criterio sobre cuál sería el desempeño deseable en producción.

Diferenciar entre los clientes con alta probabilidad de aceptar un CDT y aquellos que no.