

Universidad del Valle de Guatemala
Departamento de Matemática
Licenciatura en Matemática Aplicada

Estudiante: Rudik Roberto Rompich
Correo: rom19857@uvg.edu.gt
Carné: 19857

IA3028 - Data Mining - Catedrático: Luis Pedro Flores
7 de julio de 2021

Laboratorio 1

Preguntas a responder:

1. ¿Qué influía más en que un pasajero sobreviviera?

Solución. Los siguientes factores fueron determinantes:

- a) Pclass (la clase que el pasajero ocupaba). Según la Figura 1, la clase fue uno de los factores más elementales para determinar la supervivencia. Si se pertenecía a la primera clase las probabilidades eran muy altas, en la segunda eran indistintas y en la tercera era casi una condena de muerte.
- b) Sex (el sexo del pasajero). El sexo fue uno de los factores que determinar la supervivencia. Cabe resaltar que la mayoría de los pasajeros a bordo eran hombres; los cuales murieron en su mayoría.
- c) Fare (la tarifa del pasaje). La tarifa es algo interesante para resaltar, ya que la mayoría de pasajes vendidos fueron los más baratos y en donde más se produjeron muertes. Por lo tanto, un pasaje más caro implicaba una muerte casi segura.
- d) Embarked (lugar de embarque al Titanic). Aunque intuitivamente el lugar de abordaje no debería implicar nada; cabe resaltar que las muertes que más se produjeron fueron de las personas que embarcaron en Southampton (la ciudad de origen).

Factores que no tuvieron ninguna implicación:

- a) Age (edad). Pareciera indicar que el factor de la edad no fue determinante para sobrevivir o no.

□

2. ¿Qué % de pasajeros sobrevivieron?

Solución. Según la Figura 1 el 38.38 % de los pasajeros sobrevivieron.

□

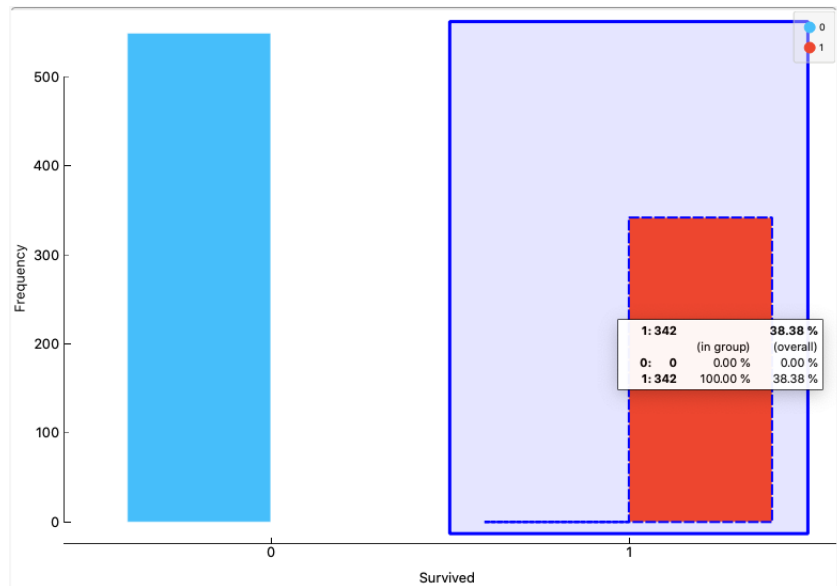


Figura 1: 0 es muerto; 1 es sobreviviente.

3. Haga un modelo que pueda predecir la probabilidad de que los pasajeros sobrevivan.

Solución. Se evaluaron 3 modelos: kNN, Logistic Refression y Tree. □

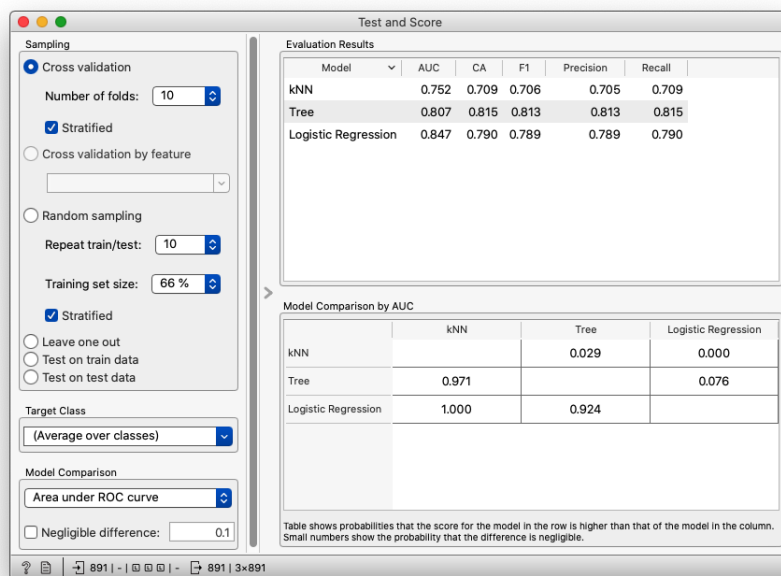


Figura 2: Modelo kNN, Logistic Regression, Tree.

4. Hay datos de pasajeros que no se tiene el “tag” si sobrevivieron o no. Corra el modelo para entender si los pasajeros sobrevivieron.

Solución. A continuación se observa un pequeño resumen de los datos de acuerdo a la predicción del modelo: □

Predictions

Show probabilities for

0
1

Restore Original Order

	Logistic Regression	Name	PassengerId	Pclass	Sex	Age	Fare	Embarked
1	0.11 → 0	Kelly, Mr. Ja...	892	3	male	34.50	7.8292	Q
2	0.37 → 0	Wilkes, Mrs. ...	893	3	female	47.00	7.0000	S
3	0.15 → 0	Myles, Mr. T...	894	2	male	62.00	9.6875	Q
4	0.08 → 0	Wirz, Mr. Alb...	895	3	male	27.00	8.6625	S
5	0.57 → 1	Hirvonen, Mr...	896	3	female	22.00	12.2875	S
6	0.12 → 0	Svensson, M...	897	3	male	14.00	9.2250	S
7	0.63 → 1	Connolly, Mi...	898	3	female	30.00	7.6292	Q
8	0.25 → 0	Caldwell, Mr...	899	2	male	26.00	29.0000	S
9	0.72 → 1	Abraham, Mrs...	900	3	female	18.00	7.2292	C
10	0.10 → 0	Davies, Mr. J...	901	3	male	21.00	24.1500	S
11	0.08 → 0	Ilieff, Mr. Ylio	902	3	male	?	7.8958	S
12	0.30 → 0	Jones, Mr. C...	903	1	male	46.00	26.0000	S
13	0.92 → 1	Snyder, Mrs. ...	904	1	female	23.00	82.2667	S
14	0.09 → 0	Howard, Mr. ...	905	2	male	63.00	26.0000	S
15	0.84 → 1	Chaffee, Mrs...	906	1	female	47.00	61.1750	S
16	0.89 → 1	del Carlo, Mr...	907	2	female	24.00	27.7208	C
17	0.30 → 0	Keane, Mr. D...	908	2	male	35.00	12.3500	Q
18	0.16 → 0	Assaf, Mr. G...	909	3	male	21.00	7.2250	C
19	0.53 → 1	Ilmakangas, ...	910	3	female	27.00	7.9250	S
20	0.52 → 1	Assaf Khalil, ...	911	3	female	45.00	7.2250	C
21	0.36 → 0	Rothschild, ...	912	1	male	55.00	59.4000	C
22	0.14 → 0	Olsen, Mast...	913	3	male	9.00	3.1708	S
23	0.90 → 1	Flegenheim, ...	914	1	female	?	31.6833	S
24	0.63 → 1	Williams, Mr...	915	1	male	21.00	61.3792	C
25	0.91 → 1	Ryerson, Mr...	916	1	female	48.00	262.3750	C
26	0.04 → 0	Robins, Mr. A...	917	3	male	50.00	14.5000	S
27	0.95 → 1	Ostby, Miss. ...	918	1	female	22.00	61.9792	C
28	0.15 → 0	Daher, Mr. S...	919	3	male	22.50	7.2250	C
29	0.34 → 0	Brady, Mr. Jo...	920	1	male	41.00	30.5000	S
30	0.13 → 0	Samaan, Mr. ...	921	3	male	?	21.6792	C
31	0.14 → 0	Louch, Mr. C...	922	2	male	50.00	26.0000	S
32	0.27 → 0	Jefferys, Mr. ...	923	2	male	24.00	31.5000	S
33	0.48 → 0	Dean, Mrs. B...	924	3	female	33.00	20.5750	S
34	0.51 → 1	Johnston, Mr...	925	3	female	?	23.4500	S
35	0.56 → 1	Mock, Mr. Ph...	926	1	male	30.00	57.7500	C
36	0.17 → 0	Katavelas, M...	927	3	male	18.50	7.2292	C

Figura 3: Predicción

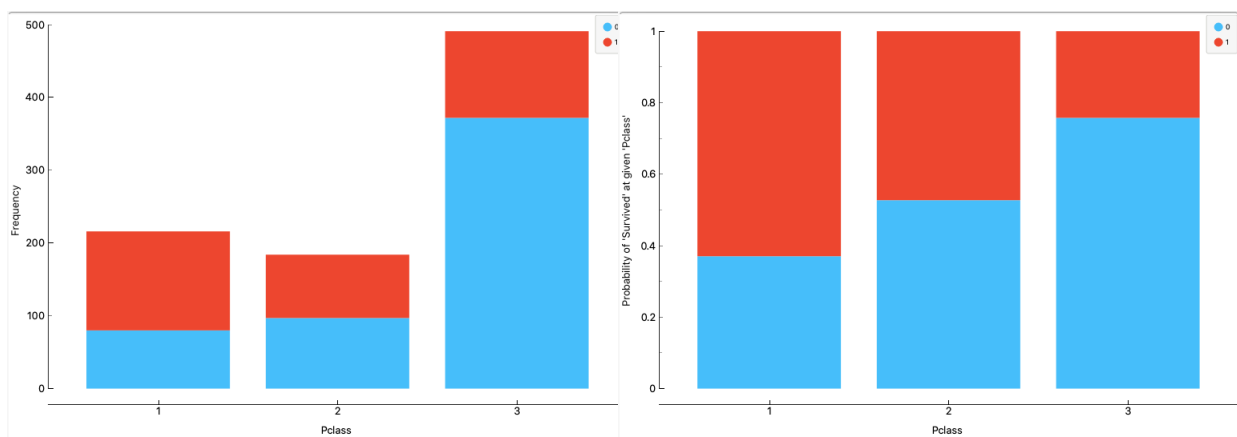
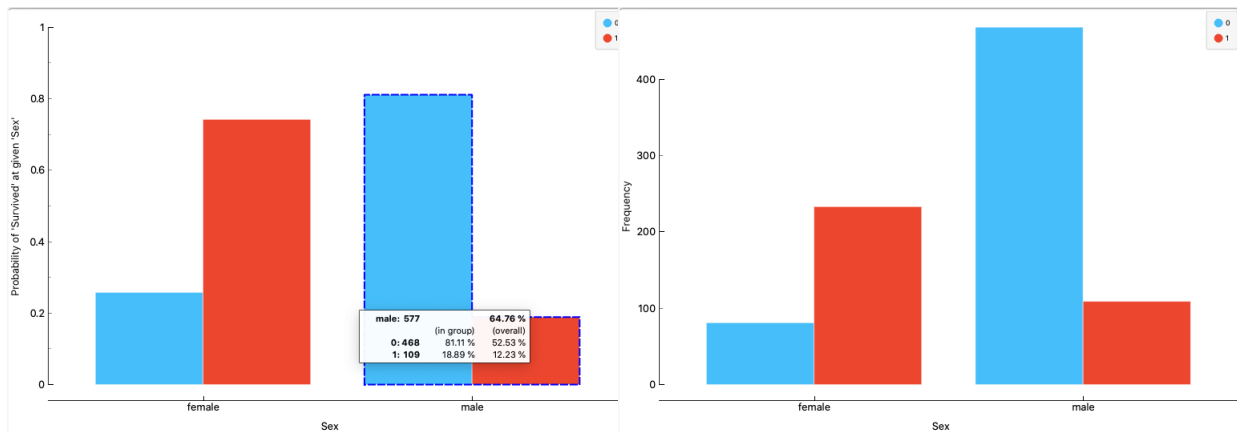


Figura 4: Clase



(a) Probabilidad de sobrevivir.

(b) Sobrevivientes.

Figura 5: Sexo

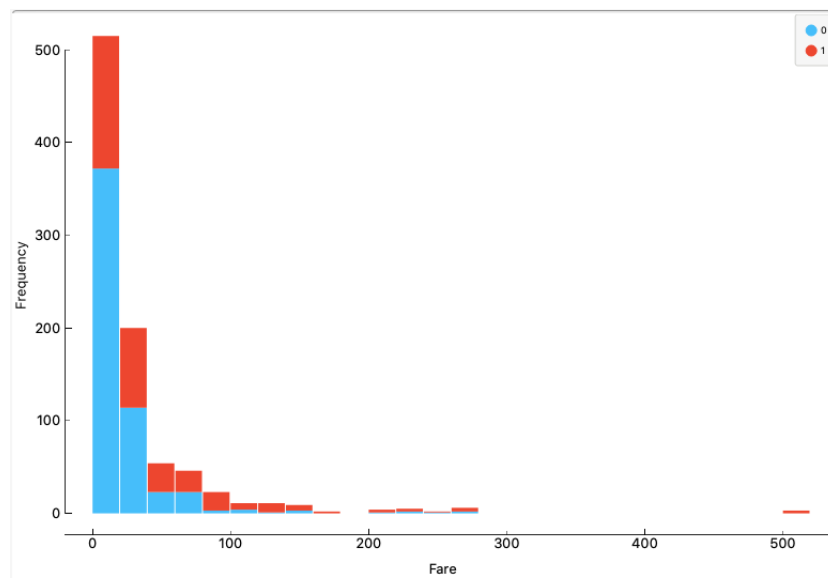


Figura 6: Fare (tarifa).

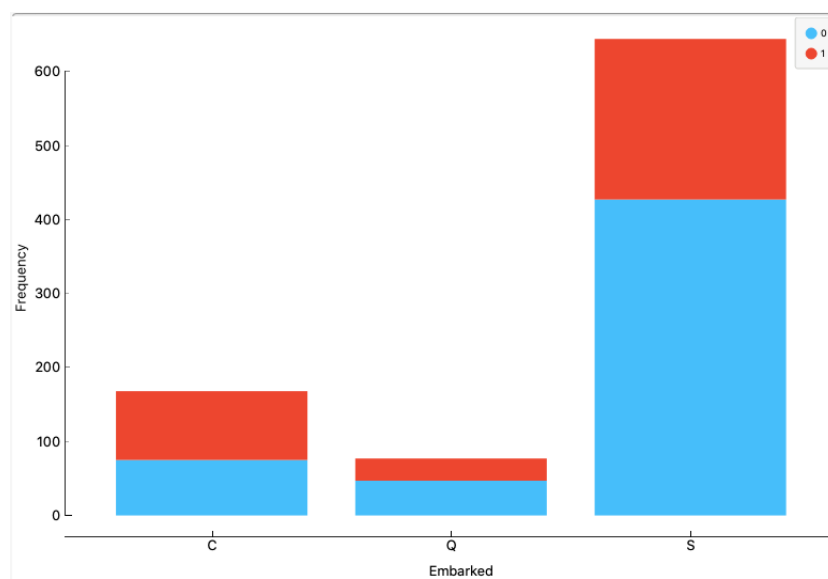


Figura 7: Embarked (embarque).