

Tarea 12

Nombres y código:

Ruiz Ortiz Cesar Adrián 208020817

Padilla Martin Juan Pablo 217294261

Álvarez Gutiérrez David Alejandro 217294016

Sección: D01

Materia: MINERIA DE DATOS

Profesor: ISRAEL ROMAN GODINEZ

CARRERA: INGENIERIA EN INFORMATICA

CICLO ESCOLAR: 2020 A



Instrucciones: Para el conjunto presentado en la Tabla 1 que corresponde a un problema de clasificación, realizar un análisis comparativo del desempeño de los modelos One-R, Naïve Bayes, y K-NN. Para ello, evalúe cada modelo utilizando la técnica de K-Fold Cross Validation con un a K=2. Para cada iteración de K, se deberá presentar la matriz de confusión llenada con la información evaluada del bloque de prueba. Como resultado final de ambas iteraciones, se presentará el promedio de la medida de exactitud. Al final indicar cuál es el modelo que mejor resultado presenta de acuerdo con el porcentaje de “Exactitud”

prome

Sail	Temperatura	Outlook	Company	Sailboat
Yes	16	rainy	big	big
Yes	15	rainy	big	small
No	14	rainy	med	big
No	19	rainy	med	small
Yes	25	sunny	big	big
Yes	31	sunny	big	small
Yes	25	sunny	med	big
Yes	32	sunny	med	big
Yes	31	sunny	med	small
Yes	30	sunny	no	small
No	35	sunny	no	big
No	15	rainy	med	big
No	20	rainy	no	big
No	15	rainy	no	big
No	18	rainy	no	small
No	20	rainy	no	small
Yes	29	sunny	big	big
No	25	sunny	big	small
No	32	sunny	med	big
No	36	sunny	med	big

Tranformamos la columna de temperatura a categorico utilizando un rango

Sail	Temperatura	Outlook	Company	Sailboat
Yes	1	rainy	big	big
Yes	1	rainy	big	small
No	1	rainy	med	big
No	2	rainy	med	small
Yes	2	sunny	big	big
Yes	3	sunny	big	small
Yes	2	sunny	med	big
Yes	3	sunny	med	big
Yes	3	sunny	med	small
Yes	3	sunny	no	small
No	4	sunny	no	big
No	1	rainy	med	big
No	2	rainy	no	big
No	1	rainy	no	big
No	1	rainy	no	small
No	2	rainy	no	small
Yes	3	sunny	big	big
No	2	sunny	big	small

One-r

$K=1$

Tabelle der Gewinne

Tabelle	Gewinn		Outlook	Knoten Summe	Gewinn	
	1	2			1	2
1	0	2				
2	0	2				
3	4	0				
4	0	0				

Gewinn	Gewinn		Outlook	Knoten Summe	Gewinn	
	1	2			1	2
1	0	0				
2	0	0				
3	4	0				
4	0	0				

Tabelle Verluste

Verluste	Verluste		Gewinn	Gewinn Tabelle
	1	2		
1	2	→ 1/2	1/2	2/10
2	2	→ 1/2	1/2	
3	3	→ 1/2	0/4	
Outlook	Summe → 10		0/4	2/10
	Summe → 10		0/6	
	1/2 → 1/2		0/1	2/10
	1/2 → 1/2		0/4	
	1/2 → 1/2		2/5	
	Summe → 10		1/5	2/10
	1/2 → 1/2		1/5	

Arbeitsblatt

Arbeitsblatt	Arbeitsblatt		Arbeitsblatt	Arbeitsblatt
	1	2		
1	1	→ 1/2	1/2	2/10
2	1	→ 1/2	1/2	
3	1	→ 1/2	0/4	
Outlook	Summe → 10		0/4	2/10
	Summe → 10		0/6	
	1/2 → 1/2		0/1	2/10
	1/2 → 1/2		0/4	
	1/2 → 1/2		2/5	
	Summe → 10		1/5	2/10
	1/2 → 1/2		1/5	

Arbeitsblatt

Arbeitsblatt	Arbeitsblatt		Arbeitsblatt	Arbeitsblatt
	1	2		
1	1	→ 1/2	1/2	2/10
2	1	→ 1/2	1/2	
3	1	→ 1/2	0/4	
Outlook	Summe → 10		0/4	2/10
	Summe → 10		0/6	
	1/2 → 1/2		0/1	2/10
	1/2 → 1/2		0/4	
	1/2 → 1/2		2/5	
	Summe → 10		1/5	2/10
	1/2 → 1/2		1/5	

KTC

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

table de données

	1	2	3	4
Landmark	1	0	2	1
	1	0	2	1
	1	0	2	1
	1	0	2	1

Calculs

	1	2	3	4
Landmark	1			

Naive bayes

NAIVE BAYES - A

TABLAS DE FRECUENCIA Y VEROSIMILITUD

	Temperature		Outlook		Windy		Rain		Sum
	Yes	No	Yes	No	Yes	No	Yes	No	
[14-16]	2/1	3/1	rain	2/1	2/1	big	4/1	0/1	8
[17-19]	2/1	1/1	Sunny	0/1	0/1	med	3/1	2/1	5
[20-22]	4/1	0/1				med	3/1	2/1	6
[23-25]	0/1	0/1				no	1/1	0/1	1

	Temperature		Outlook		Windy		Rain		Sum
	Yes	No	Yes	No	Yes	No	Yes	No	
[14-16]	3/10	2/10	rain	3/10	3/10	big	5/10	1/10	8/10
[17-19]	2/10	1/10	Sunny	1/10	1/10	med	4/10	2/10	6/10
[20-22]	5/10	1/10				med	4/10	2/10	6/10
[23-25]	0/10	1/10				no	1/10	0/10	1/10

11 $P_{Yes} = \frac{1}{10} \times \frac{2}{10} \times \frac{2}{10} \times \frac{5}{10} \times \frac{8}{10} = 0.0042 \rightarrow 0.2556$
 $P_{No} = \frac{1}{10} \times \frac{1}{10} \times \frac{1}{5} \times \frac{2}{10} \times \frac{2}{10} = 0.0008 \rightarrow 0.1641$

12 $P_{Yes} = \frac{3}{10} \times \frac{3}{10} \times \frac{1}{10} \times \frac{5}{10} \times \frac{8}{10} = 0.0090 \rightarrow 0.4210$
 $P_{No} = \frac{2}{10} \times \frac{3}{10} \times \frac{3}{5} \times \frac{2}{10} \times \frac{2}{10} = 0.015 \rightarrow 0.5989$

13 $P_{Yes} = \frac{3}{10} \times \frac{3}{10} \times \frac{2}{5} \times \frac{5}{10} \times \frac{8}{10} = 0.0054 \rightarrow 0.5217$
 $P_{No} = \frac{2}{10} \times \frac{3}{10} \times \frac{1}{5} \times \frac{2}{10} \times \frac{2}{10} = 0.0050 \rightarrow 0.4782$

14 $P_{Yes} = \frac{3}{10} \times \frac{3}{10} \times \frac{2}{10} \times \frac{5}{10} \times \frac{8}{10} = 0.0054 \rightarrow 0.5217$
 $P_{No} = \frac{2}{10} \times \frac{3}{10} \times \frac{1}{5} \times \frac{2}{10} \times \frac{2}{10} = 0.0050 \rightarrow 0.4782$

15 $P_{Yes} = \frac{3}{10} \times \frac{3}{10} \times \frac{2}{10} \times \frac{5}{10} \times \frac{8}{10} = 0.0054 \rightarrow 0.5217$
 $P_{No} = \frac{2}{10} \times \frac{3}{10} \times \frac{1}{5} \times \frac{2}{10} \times \frac{2}{10} = 0.0050 \rightarrow 0.4782$

16 $P_{Yes} = \frac{3}{10} \times \frac{3}{10} \times \frac{2}{10} \times \frac{5}{10} \times \frac{8}{10} = 0.0054 \rightarrow 0.5217$
 $P_{No} = \frac{2}{10} \times \frac{3}{10} \times \frac{1}{5} \times \frac{2}{10} \times \frac{2}{10} = 0.0050 \rightarrow 0.4782$

17 $P_{Yes} = \frac{3}{10} \times \frac{3}{10} \times \frac{2}{10} \times \frac{5}{10} \times \frac{8}{10} = 0.0054 \rightarrow 0.5217$
 $P_{No} = \frac{2}{10} \times \frac{3}{10} \times \frac{1}{5} \times \frac{2}{10} \times \frac{2}{10} = 0.0050 \rightarrow 0.4782$

18 $P_{Yes} = \frac{3}{10} \times \frac{3}{10} \times \frac{2}{10} \times \frac{5}{10} \times \frac{8}{10} = 0.0054 \rightarrow 0.5217$
 $P_{No} = \frac{2}{10} \times \frac{3}{10} \times \frac{1}{5} \times \frac{2}{10} \times \frac{2}{10} = 0.0050 \rightarrow 0.4782$

19 $P_{Yes} = \frac{3}{10} \times \frac{3}{10} \times \frac{2}{10} \times \frac{5}{10} \times \frac{8}{10} = 0.0054 \rightarrow 0.5217$
 $P_{No} = \frac{2}{10} \times \frac{3}{10} \times \frac{1}{5} \times \frac{2}{10} \times \frac{2}{10} = 0.0050 \rightarrow 0.4782$

20 $P_{Yes} = \frac{3}{10} \times \frac{3}{10} \times \frac{2}{10} \times \frac{5}{10} \times \frac{8}{10} = 0.0054 \rightarrow 0.5217$
 $P_{No} = \frac{2}{10} \times \frac{3}{10} \times \frac{1}{5} \times \frac{2}{10} \times \frac{2}{10} = 0.0050 \rightarrow 0.4782$

MATRIZ DE CONFUSION - A

MODELO	OBJETIVO	
	Yes	No
Yes	7	0
No	0	1

VALOR POSITIVO PREDICTIVO / PRECISION = $\frac{1}{7}$
 VALOR NEGATIVO PREDICTIVO = 1
 RECALL = 1
 ESPECIFICIDAD = $\frac{1}{1}$
 EXACTITUD = $\frac{1}{8} = 0.125$

MATRIZ DE CONFUSION - B

MODELO	OBJETIVO	
	Yes	No
Yes	7	0
No	7	2

VALOR POSITIVO PREDICTIVO / PRECISION = 1
 VALOR NEGATIVO PREDICTIVO = $\frac{3}{9}$
 RECALL = $\frac{1}{8}$
 ESPECIFICIDAD = 1
 EXACTITUD = $\frac{3}{10} = 0.3$

ACC = 0.25

knn

$K=1$

KNN -

$Q_1 = 0 + 0 + 0 + 1 = 1 = 1/4 = 0.25$
 $Q_2 = 1 + 0 + 0 + 0 = 1 = 1/4 = 0.25$
 $Q_3 = 1 + 0 + 0 + 1 = 2 = 2/4 = 0.5$
 $Q_4 = 1 + 0 + 0 + 0 = 1 = 1/4 = 0.25$
 $Q_5 = 0 + 0 + 1 + 0 = 1 = 1/4 = 0.25$

Actual class
for class

	Yes	No
Q1	1	0
Q2	0	1
Q3	1	0
Q4	0	1
Q5	0	1

Value of class = 1/4
Value of class = 1/4
Recall = 1
Precision = 1/2
F1 score = 1/10

$K=2$

KNN

$Q_1 = 1 + 1 + 1 + 0 = 3 = 3/4 = 0.75$
 $Q_2 = 1 + 0 + 1 + 0 = 2 = 2/4 = 0.5$
 $Q_3 = 1 + 0 + 1 + 0 = 2 = 2/4 = 0.5$
 $Q_4 = 1 + 0 + 1 + 1 = 3 = 3/4 = 0.75$
 $Q_5 = 1 + 0 + 1 + 0 = 2 = 2/4 = 0.5$

Actual class
for class

	Yes	No
Q1	1	0
Q2	0	1
Q3	0	1
Q4	1	0
Q5	0	1

Value of class = 3/4
Value of class = 1/4
Recall = 3/4
Precision = 1/2
F1 score = 3/4

$Q_1 = 0.75$