

Pruebas desarrolladas mediante aplicación de entorno Google Colab

Para el proceso de pruebas del algoritmo de predicción CART aplicado al modelo de árboles de decisión para la minería de datos se contó con 1028 registros, a su vez se crearon siete subconjuntos de datos referentes a las variables establecidas en la Tabla 1. Variables seleccionadas para el estudio, a fin de desarrollar las pruebas y obtener la generación de datos más óptimos que a continuación se detallan.

Tabla 1. Variables seleccionadas para el estudio

Nro.	Variables 2018	Variables 2019	Variables 2020
1	dia	dia	dia
2	hora	hora	hora
3	tipologia	tipologia	tipologia
4	parroquia_urbana	parroquia_urbana	parroquia_urbana
5	causas	causas	causas
6	nro_heridos	nro_heridos	nro_heridos
7	nro_fallecidos	nro_fallecidos	nro_fallecidos

- **Variable “día”**

Prueba 1 – Caso 1

Como se presenta en la Figura 1, se representaron todas las estadísticas de los 1028 registros contenidos dentro de la variable día, que fueron utilizados para el algoritmo del modelo de árboles de decisión, para mayor calidad de la imagen verificar el anexo 1.

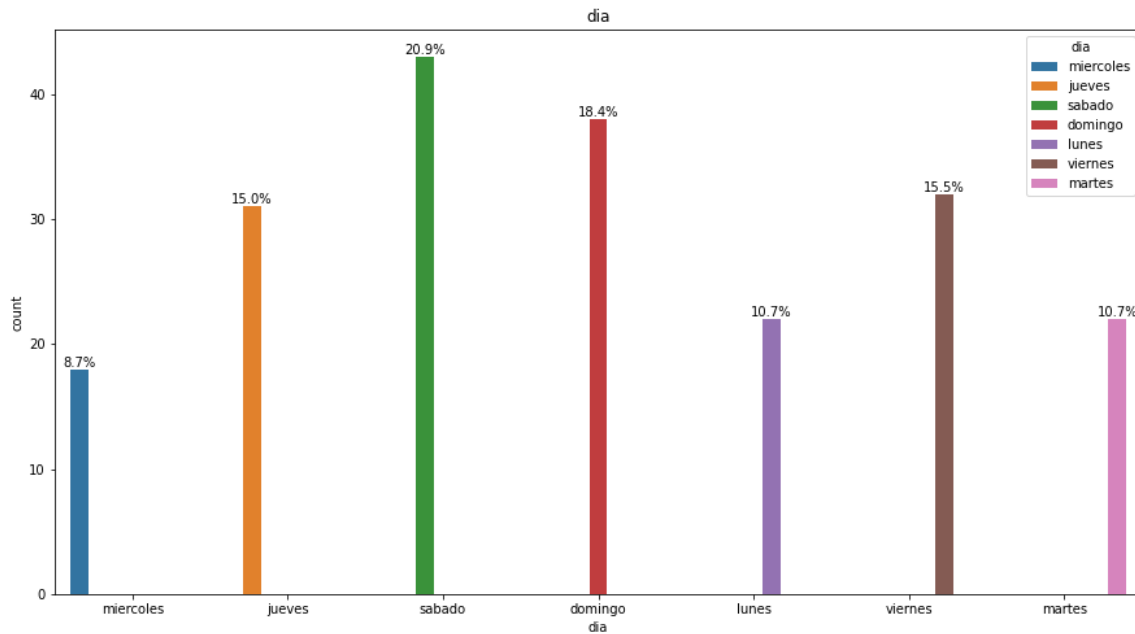


Figura 1. Representación de datos originales de variable “dia”

En la Figura 2, se establece las variables “X” y “Y”, por lo que la variable “X” tomó la división para los registros de datos de entrada, mientras que la variable “Y” almacena los registros de destino, lo que permitieron obtener los resultados de datos predichos mediante el algoritmo de clasificación de árboles de decisión.

```
[68] X = df.iloc[:, [1,2,3,4,5,6]] # atributos de entrada seran las primeras columnas
      Y = df.iloc[:,[0]] # atributos de destino

      #presentacion de los atributos de entrada
      X.head()
```

	hora	tipologia	parroquia_urbana	causas	nro_heridos	nro_fallecidos
0	4	13	4	1	0	0
1	6	13	5	2	0	0
2	6	17	4	2	2	0
3	2	8	6	2	0	0
4	3	13	5	1	0	0

```
[69] #presentacion de los atributos de destino
      Y.head()
```

	dia
0	2
1	2
2	2
3	2
4	2

Figura 2. Asignación de variable "X" para datos entrada y "Y" para datos destino.

Como se presenta en la Figura 3, dividimos los datos en las variables de entrenamiento denominada “X_train” y “y_train”, que sirvieron para crear el modelo; mientras que con las variables de predicción “X_test” y “y_test” se utilizaron para verificar la efectividad de la predicción del modelo, por lo que se utilizó el 80% de datos para entrenar el modelo y para la predicción se estableció un 20% de datos, de igual manera se estableció utilizar la semilla para el generador aleatorio con el valor de 7, para que los resultados obtenidos se puedan reproducir y se mantengan fijos.

Dividimos nuestros datos en las tablas de testing y training "X_train" y "y_train", seran para crear el modelo.

"X_test" y "y_test" serviran para saber que tan bien predice el modelo.

```
[308] X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.2, random_state=7)# stratify=y
      print('Son {} datos para entrenamiento y {} datos para prueba'.format(X_train.shape[0], X_test.shape[0]))
```

Son 822 datos para entrenamiento y 206 datos para prueba

Me muestra la información de la tabla de entrenamiento

```
[309] X_train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 822 entries, 213 to 28
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   hora                   822 non-null   int64
1   tipologia              822 non-null   int64
2   parroquia_urbana       822 non-null   int64
3   causas                 822 non-null   int64
4   nro_heridos            822 non-null   int64
5   nro_fallecidos         822 non-null   int64
dtypes: int64(6)
memory usage: 45.0 KB
```

Figura 3. Verificación de código para datos de entrenamiento y predicción del modelo.

Dentro de la Figura 4, se utilizó la librería de “sklearn” para crear el modelo de árbol de decisión, entrenar el modelo de predicción y generar las matrices de confusión, a su vez para la implementación del modelo se consideró que la profundidad de árbol inicial será de máximo 3 nodos internos para reducir el consumo de memoria, la complejidad y el tamaño de los árboles, en donde se verificó el ajuste del árbol a los datos; del mismo modo, se estableció la utilización del criterio de “entropía” para minimizar la pérdida de registros en el modelado; luego se generó la gráfica del árbol generado por el modelo, tal como se observa en la Figura 5, para mayor calidad de imagen revisar el anexo 2.

```
[ ] #Cargamos la libreria DecisionTreeClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import confusion_matrix, classification_report, plot_confusion_matrix

#llamamos al constructor del arbol de decision
classifier = DecisionTreeClassifier(max_depth=3,criterion = 'entropy')

#Entrenamos el modelo
arbol_modelo = classifier.fit(X_train,y_train)

y_pred = classifier.predict(X_test)
```

Figura 4. Implementación de código el modelo de clasificación de árboles de decisión.

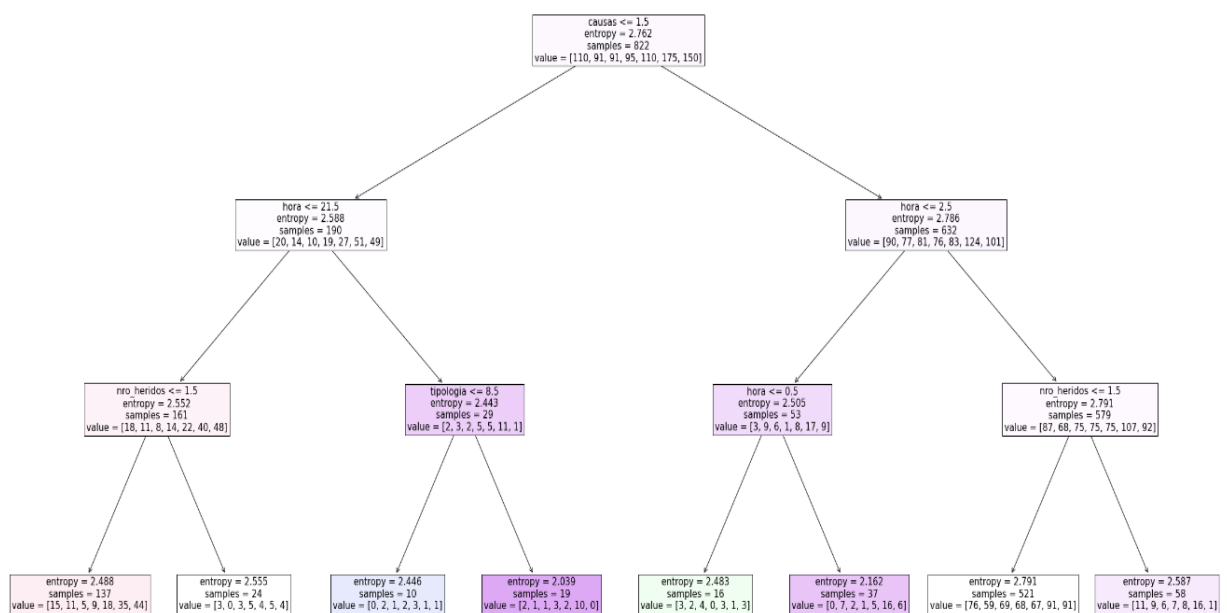


Figura 5. Árbol de decisión de la variable "dia" generado por ejecución de la prueba 1.

En la Tabla 2. Exactitud y precisión del modelo de predicción de variable “dia” prueba 1, se presenta el informe de datos generados que permitieron verificar la muestra de las principales métricas de clasificación en el modelo aplicado; además dentro de la matriz de confusión presentada en la Figura 6, se evidencia la cantidad de los valores predichos por el modelo, en donde los valores obtenidos en la diagonal principal respecto a los valores predichos incorrectamente están acorde al porcentaje generado en la exactitud del modelo.

Tabla 2. Exactitud y precisión del modelo de predicción de variable “dia” prueba 1

	precision	recall	f1-score	support
1	0	0	0	22
2	0	0	0	22
3	0,142857	0,055556	0,08	18
4	0	0	0	31
5	0	0	0	32
6	0,194969	0,72093	0,306931	43
7	0,257143	0,236842	0,246575	38
accuracy	0,199029	0,199029	0,199029	0,199029
macro avg	0,084996	0,144761	0,090501	206
weighted avg	0,100614	0,199029	0,116543	206

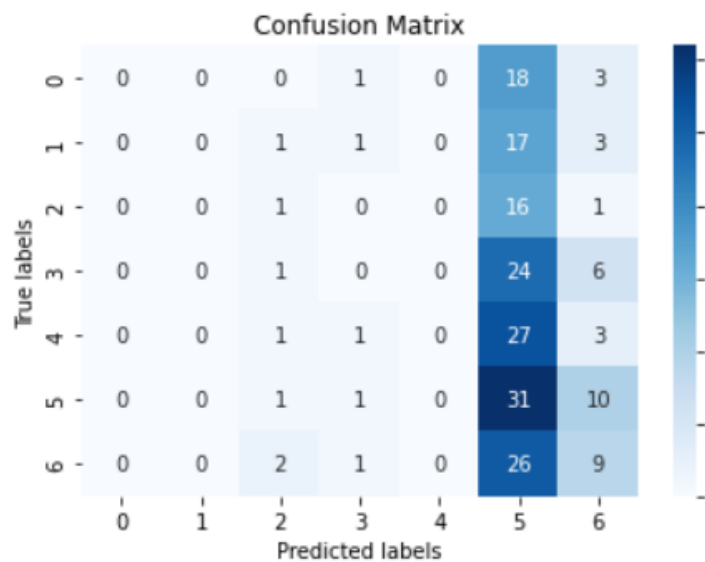


Figura 6. Matriz de confusión de variable "dia" prueba 1.

El código generado para la minería de datos de variable “dia” prueba 1 - caso 1, se encuentra presentado en el anexo 3.

Prueba 2 - Caso 2

En esta prueba se presentó las variables generadas por rangos en donde únicamente se consideró el rango de la variable “dia”, mientras los datos contenidos en las variables “hora” y “tipologia” permanecieron con los datos originales, tal como se presenta en el anexo 28.

En la Tabla 3. Exactitud y precisión del modelo de predicción de variable “dia” prueba 2.,se presentó el informe de datos generados que permitieron establecer el valor de las principales métricas de clasificación en el modelo aplicado; además dentro de la matriz de confusión presentada en la

Figura 7 , se evidencia la cantidad de los valores predichos por el modelo, en donde los valores obtenidos en la diagonal principal respecto a los valores predichos incorrectamente están acorde al porcentaje generado en la exactitud del modelo.

Tabla 3. Exactitud y precisión del modelo de predicción de variable “dia” prueba 2.

	precision	recall	f1-score	support
1	0,328947368	0,403225806	0,362318841	62
2	0	0	0	63
3	0,415384615	0,666666667	0,511848341	81
accuracy	0,383495146	0,383495146	0,383495146	0,383495146
macro avg	0,248110661	0,356630824	0,291389061	206
weighted avg	0,262334421	0,383495146	0,310308174	206

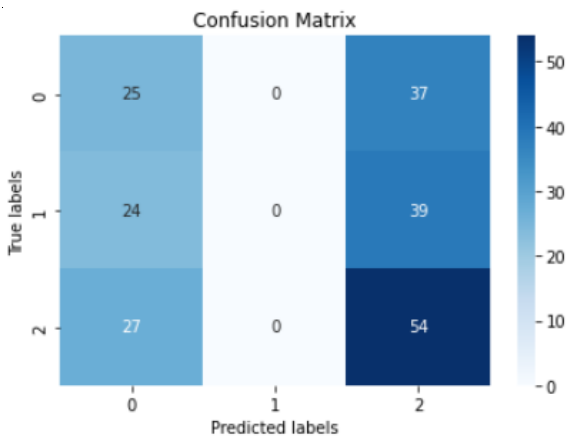


Figura 7. Matriz de confusión de variable "dia" prueba 2.

El presente código generado para la minería de datos de variable “dia” prueba 2 - caso 2, se encuentra presentado en el anexo 4.

Prueba 3 - Caso 2

En esta prueba se presentó las variables generadas por rangos en donde se consideró el rango de la variable “dia”, “hora” y “tipologia”, los cuales fueron seleccionados para verificar la variación del porcentaje de exactitud del modelo, tal como se presenta en el anexo 28.

En la Tabla 4. Exactitud y precisión del modelo de predicción de variable “dia” prueba 3, se presentó el informe del valor de las nuevas métricas de clasificación del modelo; además dentro de la matriz de confusión presentada en la

Figura 8, se evidencia la cantidad de los valores predichos por el modelo, siendo acordes al porcentaje generado en la exactitud del modelo.

Tabla 4. Exactitud y precisión del modelo de predicción de variable “dia” prueba 3

	precision	recall	f1-score	support
1	0,296296	0,258065	0,275862	62
2	0	0	0	63
3	0,388158	0,728395	0,506438	81
accuracy	0,364078	0,364078	0,364078	0,364078
macro avg	0,228151	0,32882	0,260767	206
weighted avg	0,241802	0,364078	0,28216	206

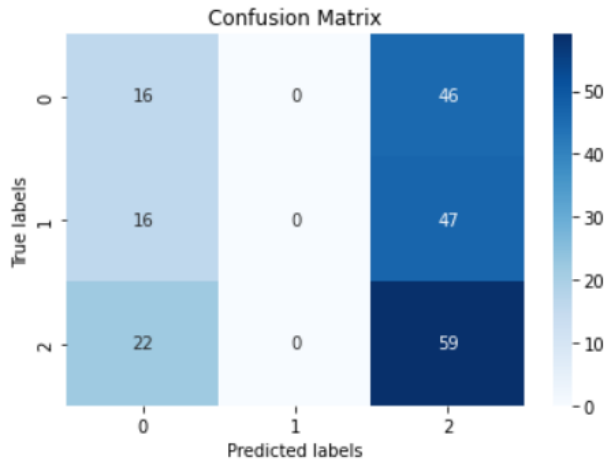


Figura 8. Matriz de confusión de variable "dia" prueba 3.

El presente código generado para la minería de datos de variable “dia” prueba 3 - caso 2, se encuentra presentado en el anexo 5.

Variable “hora”

Prueba 1 – Caso 1

Como se presenta en la **¡Error! No se encuentra el origen de la referencia.**, se representaron todos los porcentajes de los 1028 registros contenidos dentro de la variable hora, para mayor calidad de la imagen verificar el anexo 6.

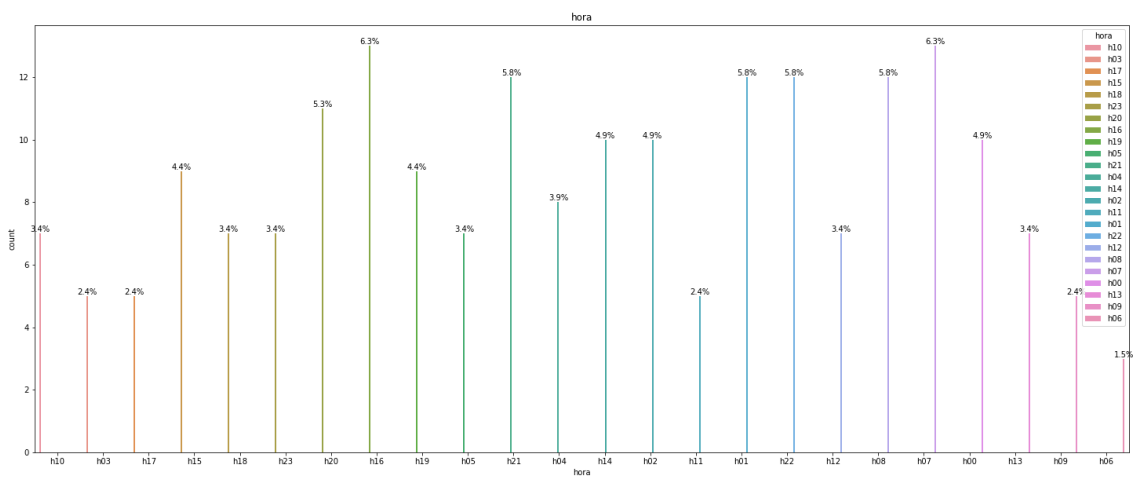


Figura 9. Representación de valores originales de variable "hora"

En esta prueba se utilizó la variable “hora” como atributo de referencia para la predicción de datos, en donde las variables restantes permanecieron con los datos originales identificados en el caso 1.

En la Tabla 5. Exactitud y precisión del modelo de predicción de variable "hora" prueba 1., se presentó el informe del valor de las nuevas métricas de clasificación del modelo; además dentro de la matriz de confusión presentada en la Figura 10, se evidencia la cantidad de los valores predichos por el modelo, siendo acordes al porcentaje generado en la exactitud del modelo.

Tabla 5. Exactitud y precisión del modelo de predicción de variable "hora" prueba 1.

	precision	recall	f1-score	support
0	0	0	0	10
1	0	0	0	12
2	0	0	0	10
3	0,053571	0,6	0,098361	5
4	0	0	0	8
5	0	0	0	7
6	0	0	0	3
7	0	0	0	13
8	0	0	0	12
9	0	0	0	5
10	0	0	0	7
11	0	0	0	5
12	0,048193	0,571429	0,088889	7
13	0	0	0	7
14	0	0	0	10
15	0	0	0	9
16	0	0	0	13
17	0	0	0	5
18	0	0	0	7
19	0	0	0	9
20	0	0	0	11
21	0	0	0	12
22	0,037037	0,083333	0,051282	12
23	0	0	0	7
accuracy	0,038835	0,038835	0,038835	0,038835
macro avg	0,005783	0,052282	0,009939	206
weighted avg	0,005095	0,038835	0,008395	206

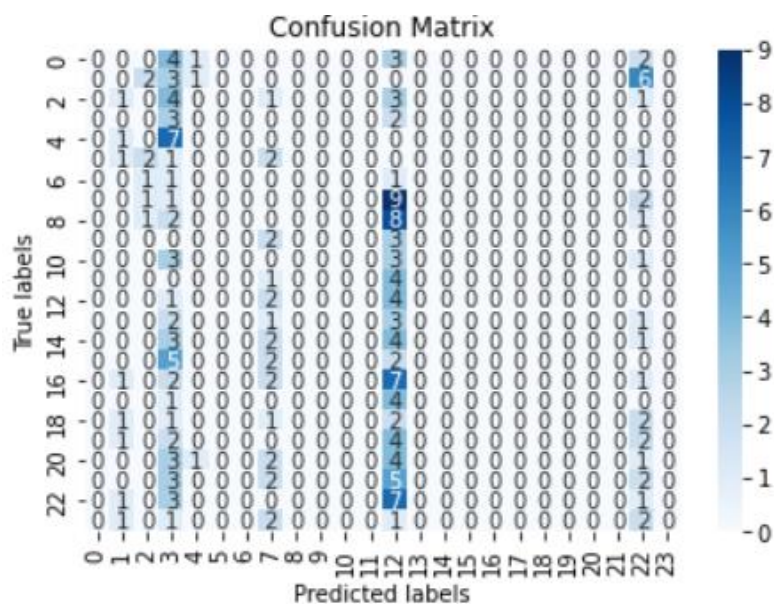


Figura 10. Matriz de confusión de variable "hora" prueba 1.

El presente código generado para la minería de datos de variable “hora” prueba 1 - caso 1, se encuentra presentado en el anexo 7.

Prueba 2 – Caso 2

En esta prueba se presentó las variables generadas por rangos en donde únicamente se consideró el rango de la variable “hora”, mientras los datos contenidos en las variables “dia” y “tipologia” permanecen con los datos originales, tal como se presenta en el anexo 28.

En la Tabla 6. Exactitud y precisión del modelo de predicción de variable "hora" prueba 2., se presentó el informe de datos generados que permitieron establecer el valor de la exactitud del modelo aplicado; además dentro de la matriz de confusión presentada en la Figura 11, se evidencia la cantidad de los valores predichos, siendo acordes al porcentaje generado en la exactitud del modelo.

Tabla 6. Exactitud y precisión del modelo de predicción de variable "hora" prueba 2.

	precision	recall	f1-score	support
0	0,283784	0,567568	0,378378	37
1	0,375	0,096774	0,153846	31
2	0,131579	0,172414	0,149254	29
3	0,171875	0,333333	0,226804	33
4	0	0	0	34
5	0,227273	0,119048	0,15625	42
accuracy	0,218447	0,218447	0,218447	0,218447
macro avg	0,198252	0,214856	0,177422	206
weighted avg	0,199797	0,218447	0,180314	206

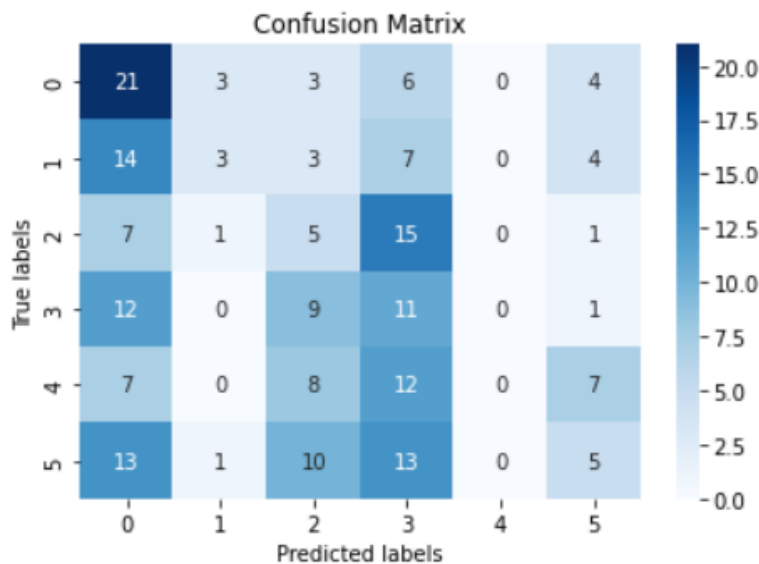


Figura 11. Matriz de confusión de variable "hora" prueba 2.

El presente código generado para la minería de datos de variable “hora” prueba 2 - caso 2, se encuentra presentado en el anexo 8.

Prueba 3 – Caso 2

En esta prueba se presentó las variables generadas por rangos en donde se consideró el rango de las variables “hora”, “dia” y “tipologia” con los rangos de valores del caso 2, tal como se presenta en el anexo 28.

En la Tabla 7. Exactitud y precisión del modelo de predicción de variable "hora" prueba 3., se presentó el informe de datos generados que permitieron establecer el valor de la

exactitud del modelo aplicado; además dentro de la matriz de confusión presentada en la Figura 12, se evidencia la cantidad de los valores predichos, siendo acordes al porcentaje generado en la exactitud del modelo.

Tabla 7. Exactitud y precisión del modelo de predicción de variable "hora" prueba 3.

	precision	recall	f1-score	support
0	0,282051	0,594595	0,382609	37
1	0,666667	0,064516	0,117647	31
2	0,131579	0,172414	0,149254	29
3	0,171875	0,333333	0,226804	33
4	0	0	0	34
5	0,173913	0,095238	0,123077	42
accuracy	0,213592	0,213592	0,213592	0,213592
macro avg	0,237681	0,210016	0,166565	206
weighted avg	0,232498	0,213592	0,168863	206

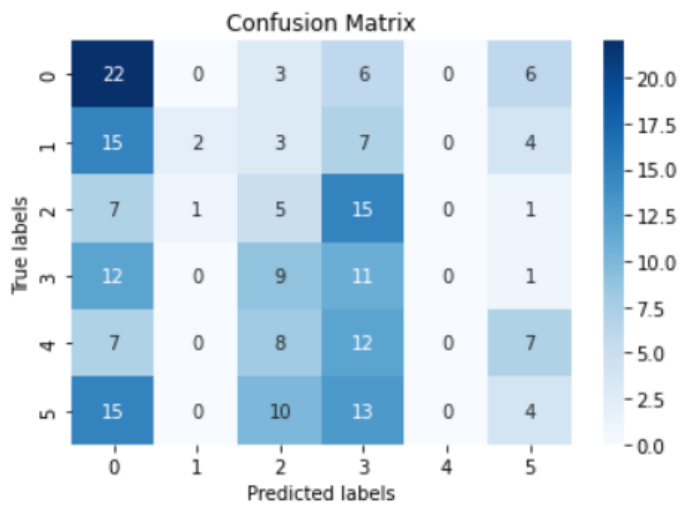


Figura 12. Matriz de confusión de variable "hora" prueba 3.

El presente código generado para la minería de datos de variable “hora” prueba 3 - caso 3, se encuentra presentado en el anexo 9.

Prueba 4 – Caso 3

En esta prueba se presentó las variables generadas por rangos en donde únicamente se consideró el rango de la variable “hora”, mientras los datos contenidos en las variables “dia” y “tipologia” permanecen con los datos originales del caso 3, tal como se presenta en el anexo 28.

En la Tabla 8. Exactitud y precisión del modelo de predicción de variable "hora" prueba 4., se presentó el informe de datos generados que permitieron establecer el valor de la exactitud del modelo aplicado; además dentro de la matriz de confusión presentada en la Figura 13, se evidencia la cantidad de los valores predichos, siendo acordes al porcentaje generado en la exactitud del modelo.

Tabla 8. Exactitud y precisión del modelo de predicción de variable "hora" prueba 4.

	precision	recall	f1-score	support
0	0,452055	0,634615	0,528	52
1	0	0	0	45
2	0,309278	0,588235	0,405405	51
3	0,388889	0,241379	0,297872	58
accuracy	0,373786	0,373786	0,373786	0,373786
macro avg	0,287556	0,366057	0,307819	206
weighted avg	0,300173	0,373786	0,317516	206

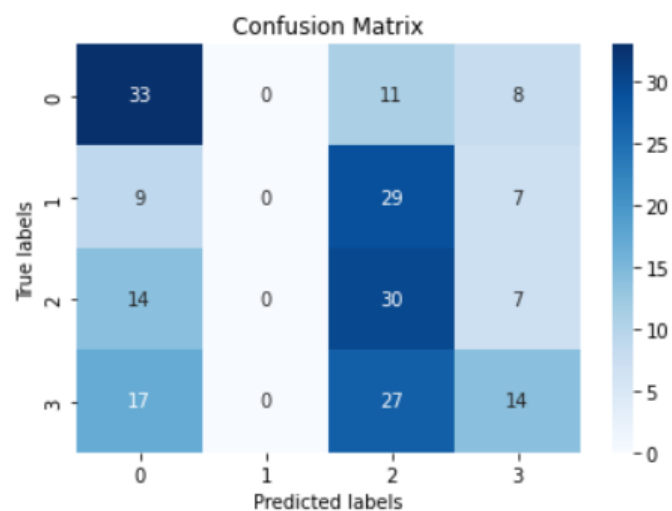


Figura 13. Matriz de confusión de variable "hora" prueba 4.

El presente código generado para la minería de datos de variable “hora” prueba 4 - caso 3, se encuentra presentado en el anexo 10.

Prueba 5 – Caso 3

En esta prueba se presentó las variables generadas por rangos en donde se consideró el rango de las variables “hora”, “dia” y “tipologia” con los rangos de valores del caso 3, tal como se presenta en el anexo 28.

En la Tabla 9. Exactitud y precisión del modelo de predicción de variable "hora" prueba 5., se presentó el informe de datos generados que permitieron establecer el valor de la exactitud del modelo aplicado; además dentro de la matriz de confusión presentada en la Figura 14, se evidencia la cantidad de los valores predichos, siendo acordes al porcentaje generado en la exactitud del modelo.

Tabla 9. Exactitud y precisión del modelo de predicción de variable "hora" prueba 5.

	precision	recall	f1-score	support
0	0,428571	0,634615	0,511628	52
1	0	0	0	45
2	0,309278	0,588235	0,405405	51
3	0,40625	0,224138	0,288889	58
accuracy	0,368932	0,368932	0,368932	0,368932
macro avg	0,286025	0,361747	0,301481	206
weighted avg	0,299133	0,368932	0,310854	206

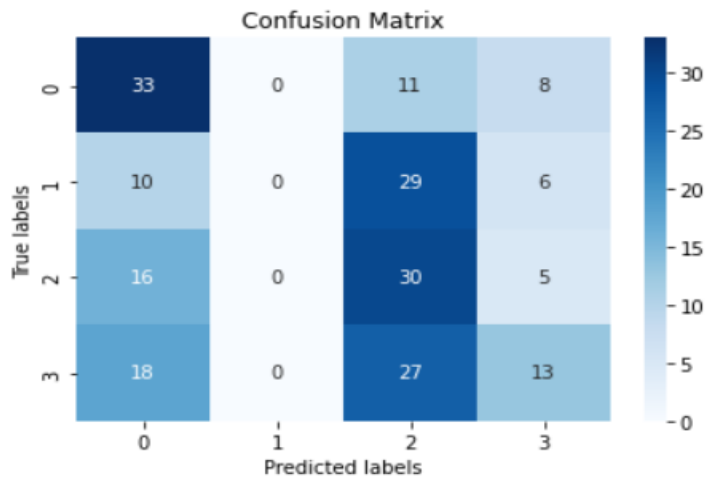


Figura 14. Matriz de confusión de variable "hora" prueba 5.

El presente código generado para la minería de datos de variable “hora” prueba 5 - caso 3, se encuentra presentado en el anexo 11.

Variable “tipologia”

Prueba 1 – Caso 1

Como se presenta en la Figura 15, se representaron todos los porcenjates de los 1028 registros contenidos dentro de la variable “tipologia”, para mayor calidad de la imagen verificar el anexo 12.

En esta prueba se utilizó la variable “tipologia” como atributo de referencia para la predicción de datos, en donde las variables restantes permanecieron con los datos originales identificados en el caso 1.

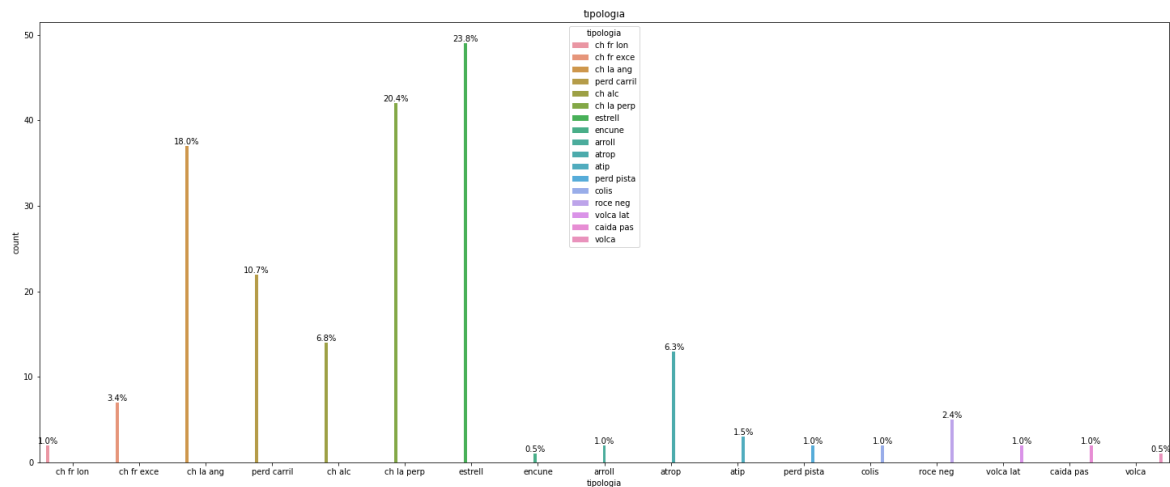


Figura 15. Representación de valores originales de variable "tipologia"

En la Tabla 10. Exactitud y precisión del modelo de predicción de variable "tipología" prueba 1., se presentó el informe del valor de las nuevas métricas de clasificación del modelo, en donde se puede observar que la exactitud es de 37,37%, lo que representa que el valor obtenido es relativamente bajo; además dentro de la matriz de confusión presentada en la Figura 16, se evidencia la cantidad de los valores predichos por el modelo, siendo acordes al porcentaje generado en la exactitud del modelo.

Tabla 10. Exactitud y precisión del modelo de predicción de variable "tipología" prueba 1.

	precision	recall	f1-score	support
1	0	0	0	2
2	0	0	0	3
3	0,833333	0,384615	0,526316	13
4	0	0	0	2
6	0	0	0	7
7	0	0	0	2
8	0	0	0	37
9	0,44	0,52381	0,478261	42
10	1	0,428571	0,6	14
11	0	0	0	2
12	0	0	0	1
13	0,305556	0,897959	0,455959	49
14	0	0	0	22
15	0	0	0	2
16	0	0	0	5
19	0	0	0	1
20	0	0	0	2
accuracy	0,373786	0,373786	0,373786	0,373786
macro avg	0,151699	0,131468	0,121208	206
weighted avg	0,28294	0,373786	0,279956	206

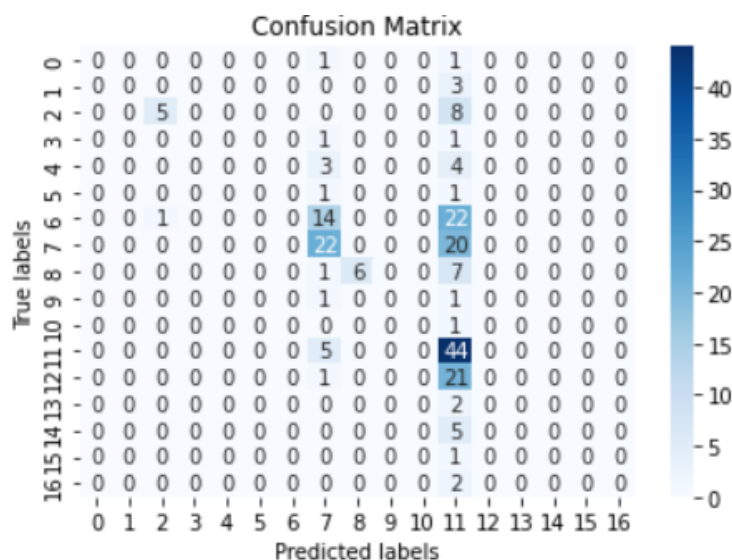


Figura 16. Matriz de confusión de variable "tipologia" prueba 1.

El presente código generado para la minería de datos de variable “tipologia” prueba 1 - caso 1, se encuentra presentado en el anexo 13.

Prueba 2 – Caso 2

En esta prueba se presentó la variable “tipologia” en donde se realizó la comparación de tipologías viales desarrollando un registro general, mientras que las demás variables del Dataset permanecen con los datos originales del caso 2.

En la Tabla 11. Exactitud y precisión del modelo de predicción de variable "tipologia" prueba 2., se presentó el informe de datos generados que permitieron establecer el valor de la exactitud del modelo aplicado, en donde se puede ver que representa un 56,31% de exactitud, el cual muestra que el nivel de porcentaje es mayor al obtenido anteriormente en la prueba 1, aquí se considera que el valor obtenido es considerable; además dentro de la matriz de confusión presentada en la Figura 17, se evidencia la cantidad de los valores predichos, siendo acordes al porcentaje generado en la exactitud del modelo.

Tabla 11. Exactitud y precisión del modelo de predicción de variable "tipologia" prueba 2.

	precision	recall	f1-score	support
1	0	0	0	2
2	0	0	0	3
3	0,833333	0,384615	0,526316	13
4	0	0	0	2

5	0,595092	0,95098	0,732075	102
6	0	0	0	2
7	0	0	0	1
8	0,378378	0,285714	0,325581	49
9	0	0	0	24
10	0	0	0	5
11	0	0	0	3
accuracy	0,563107	0,563107	0,563107	0,563107
macro avg	0,164255	0,147392	0,143998	206
weighted avg	0,437249	0,563107	0,473142	206

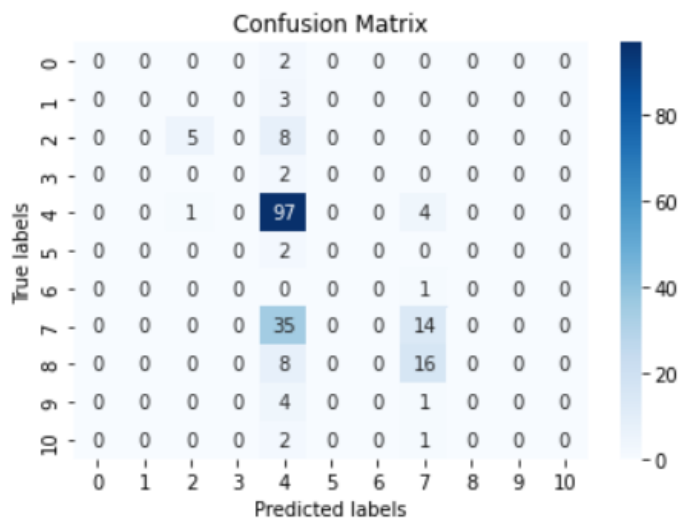


Figura 17. Matriz de confusión de variable "tipologia" prueba 2.

El presente código generado para la minería de datos de variable “tipologia” prueba 2 - caso 2, se encuentra presentado en el anexo 14.

Prueba 3 – Caso 2 y Prueba 4 – Caso 3

En estas pruebas se presentó las variables generadas por rangos en donde se consideró el rango de las variables “hora”, “dia” y “tipologia” con todos los rangos de datos del caso 2, a su vez se realizó las pruebas considerando los datos del caso 3, tal como se presenta en el anexo 28.

En la Tabla 12. Exactitud y precisión del modelo de predicción de variable "tipologia" prueba 3 y prueba 4., se presentó el informe de datos generados que permitieron establecer

el valor de la exactitud del modelo aplicado; además dentro de la matriz de confusión presentada en la Figura 18, se evidencia la cantidad de los valores predichos, siendo acordes al porcentaje generado en la exactitud del modelo; estos resultados son equivalentes a los resultados generados por la prueba 4, ya que contienen las mismas características y valores generados por la predicción.

Tabla 12. Exactitud y precisión del modelo de predicción de variable "tipologia" prueba 3 y prueba 4.

	precision	recall	f1-score	support
1	0	0	0	2
2	0	0	0	3
3	0,833333	0,384615	0,526316	13
4	0	0	0	2
5	0,595092	0,95098	0,732075	102
6	0	0	0	2
7	0	0	0	1
8	0,378378	0,285714	0,325581	49
9	0	0	0	24
10	0	0	0	5
11	0	0	0	3
accuracy	0,563107	0,563107	0,563107	0,563107
macro avg	0,164255	0,147392	0,143998	206
weighted avg	0,437249	0,563107	0,473142	206

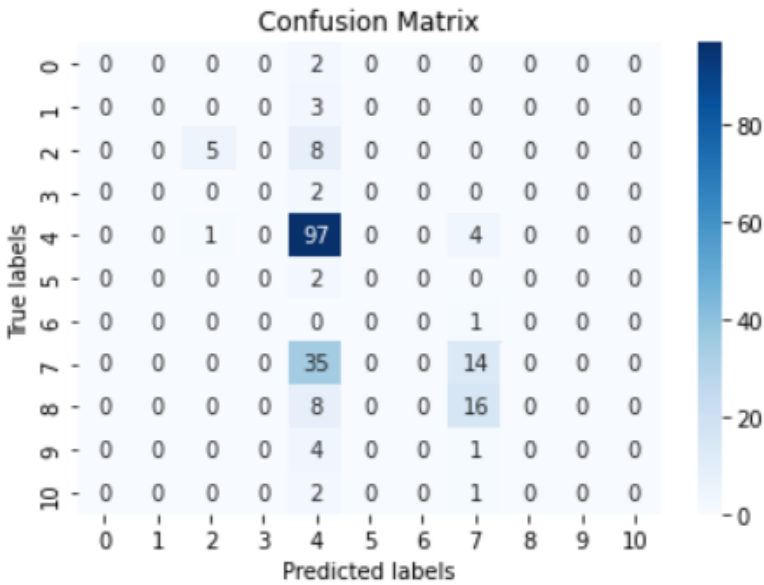


Figura 18. Matriz de confusión de variable "tipologia" prueba 3 y prueba 4.

El presente código generados para la minería de datos de variable “tipologia” prueba 3 - caso 2, se encuentra presentado en el anexo 15 y los datos de la variable “tipologia” prueba 4 - caso 3, se encuentran presentados en el anexo 16.

Variable “parroquia_urbana”

Prueba 1 – Caso 1

Como se presenta en la Figura 19, se representaron todas las estadísticas de los 1028 registros contenidos dentro de la variable “parroquia_urbana”, para mayor calidad de la imagen verificar el anexo 17.

En esta prueba se utilizó la variable “parroquia_urbana” como atributo de referencia para la predicción de datos, en donde las variables restantes permanecieron con los datos originales identificados en el caso 1, tal como se presenta en el anexo 28.

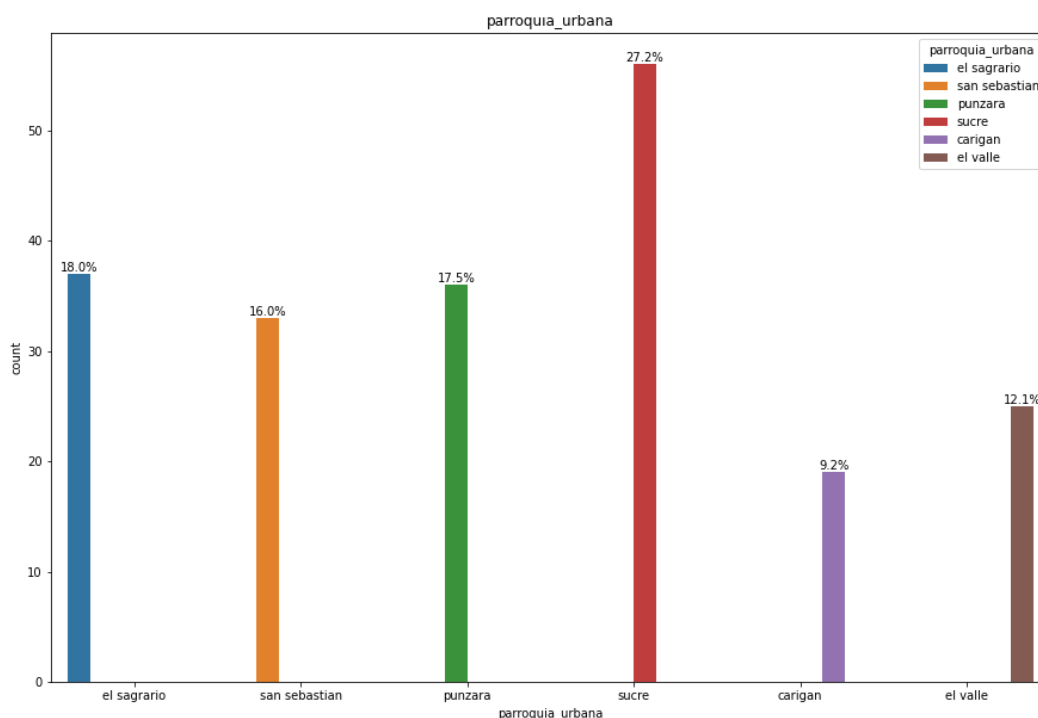


Figura 19. Representación de valores originales de variable "parroquia_urbana".

En la Tabla 13. Exactitud y precisión del modelo de predicción de variable "parroquia_urbana" prueba 1.,se presentó el informe del valor de las nuevas métricas de clasificación del modelo; además dentro de la matriz de confusión presentada en la Figura 20, se evidencia la cantidad de los valores predichos por el modelo, siendo acordes al porcentaje generado en la exactitud del modelo.

Tabla 13. Exactitud y precisión del modelo de predicción de variable "parroquia_urbana" prueba 1.

	precision	recall	f1-score	support
1	0,357143	0,405405	0,379747	37
2	0	0	0	33
3	0	0	0	25
4	0,291925	0,839286	0,43318	56
5	0,666667	0,055556	0,102564	36
6	0	0	0	19
accuracy	0,31068	0,31068	0,31068	0,31068
macro avg	0,219289	0,216708	0,152582	206
weighted avg	0,26001	0,31068	0,203888	206

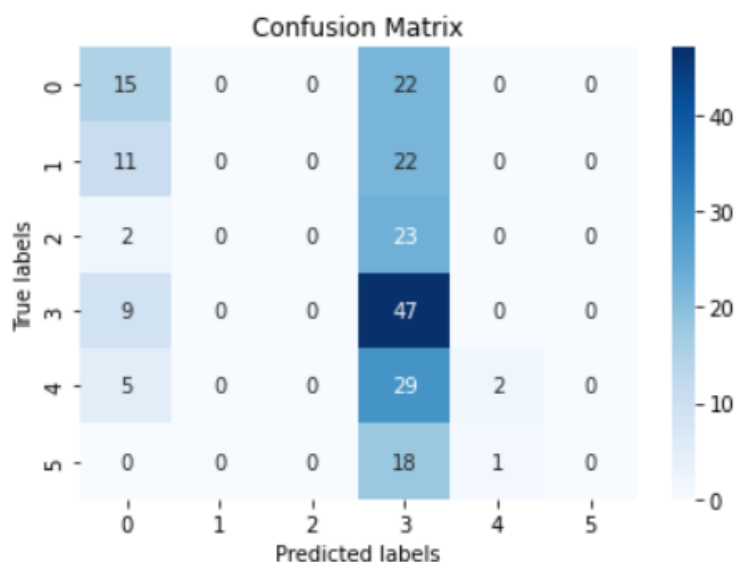


Figura 20. Matriz de confusión de variable "parroquia_urbana" prueba 1.

El presente código generado para la minería de datos de variable “parroquia_urbana” prueba 1 - caso 1, se encuentra presentado en el anexo 18.

Prueba 2 – Caso 2

En esta prueba se presentó las variables generadas por rangos en donde se consideró el rango de las variables “hora”, “dia” y “tipologia” con los rangos de valores del caso 2, tal como se presenta en el anexo 28.

En la Tabla 14. Exactitud y precisión del modelo de predicción de variable "parroquia_urbana" prueba 2., se presentó el informe de datos generados que permitieron establecer el valor de la exactitud del modelo aplicado; además dentro de la matriz de confusión presentada en la Figura 21, se evidencia la cantidad de los valores predichos, siendo acordes al porcentaje generado en la exactitud del modelo.

Tabla 14. Exactitud y precisión del modelo de predicción de variable "parroquia_urbana" prueba 2.

	precision	recall	f1-score	support
1	0,272727	0,162162	0,20339	37
2	0	0	0	33
3	0	0	0	25
4	0,277174	0,910714	0,425	56
5	0	0	0	36
6	0	0	0	19
accuracy	0,276699	0,276699	0,276699	0,276699
macro avg	0,09165	0,178813	0,104732	206
weighted avg	0,124333	0,276699	0,152065	206

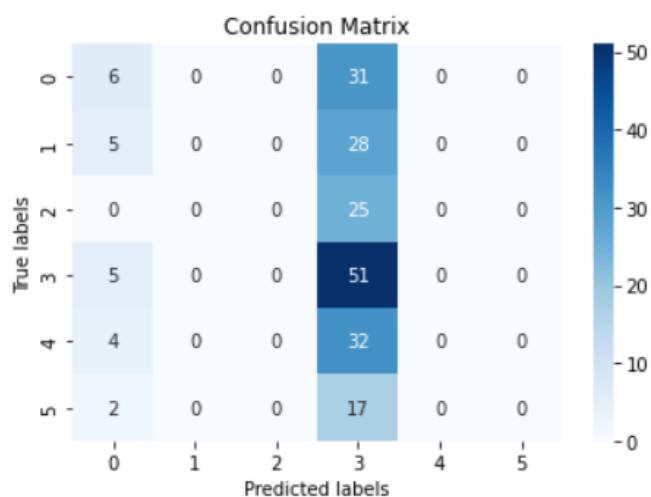


Figura 21. Matriz de confusión de variable "parroquia_urbana" prueba 2.

El presente código generado para la minería de datos de variable “parroquia_urbana” prueba 2 - caso 2, se encuentra presentado en el anexo 19.

Prueba 3 – Caso 3

En esta prueba se presentó las variables generadas por rangos en donde se consideró el rango de las variables “hora”, “dia” y “tipologia” con los rangos de valores del caso 3, tal como se presenta en el anexo 28. En la descripción de la

Tabla 15. Exactitud y precisión del modelo de predicción de variable "parroquia_urbana" prueba 3., se presentó el informe de datos generados que permitieron establecer el valor de la exactitud del modelo aplicado; además dentro de la matriz de confusión presentada en la Figura 22, se evidencia la cantidad de los valores predichos, siendo acordes al porcentaje generado en la exactitud del modelo.

Tabla 15. Exactitud y precisión del modelo de predicción de variable "parroquia_urbana" prueba 3.

	precision	recall	f1-score	support
1	0,285714	0,108108	0,156863	37
2	0	0	0	33
3	0	0	0	25
4	0,276042	0,946429	0,427419	56
5	0	0	0	36
6	0	0	0	19
accuracy	0,276699	0,276699	0,276699	0,276699
macro avg	0,093626	0,175756	0,09738	206
weighted avg	0,126358	0,276699	0,144366	206

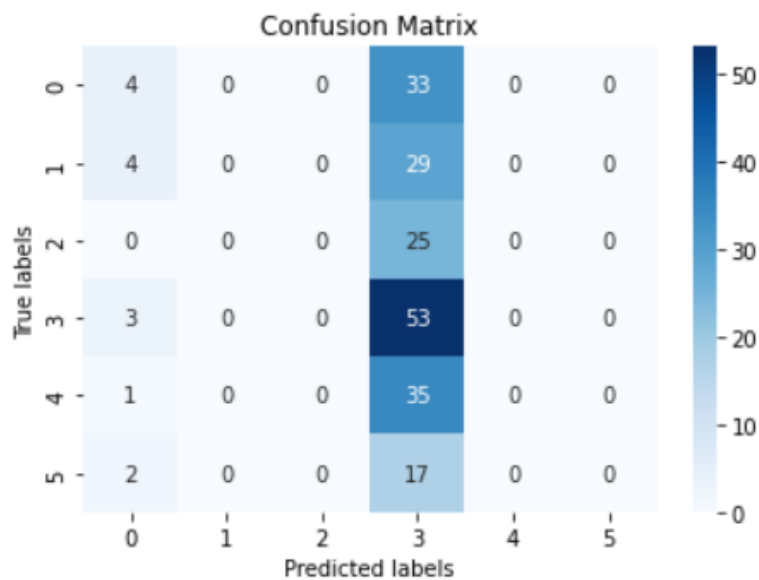


Figura 22. Matriz de confusión de variable "parroquia_urbana" prueba 3.

El presente código generado para la minería de datos de variable “parroquia_urbana” prueba 2 - caso 2, se encuentra presentado en el anexo 20.

Variable “causas”

Prueba 1 – Caso 1

Como se presenta en la Figura 23, se representaron todos los porcenjates de los 1028 registros contenidos dentro de la variable “causas”.

En esta prueba se utilizó la variable “causas” como atributo de referencia para la predicción de datos, en donde las variables restantes permanecieron con los datos originales identificados en el caso 1, tal como se presenta en el anexo 28.

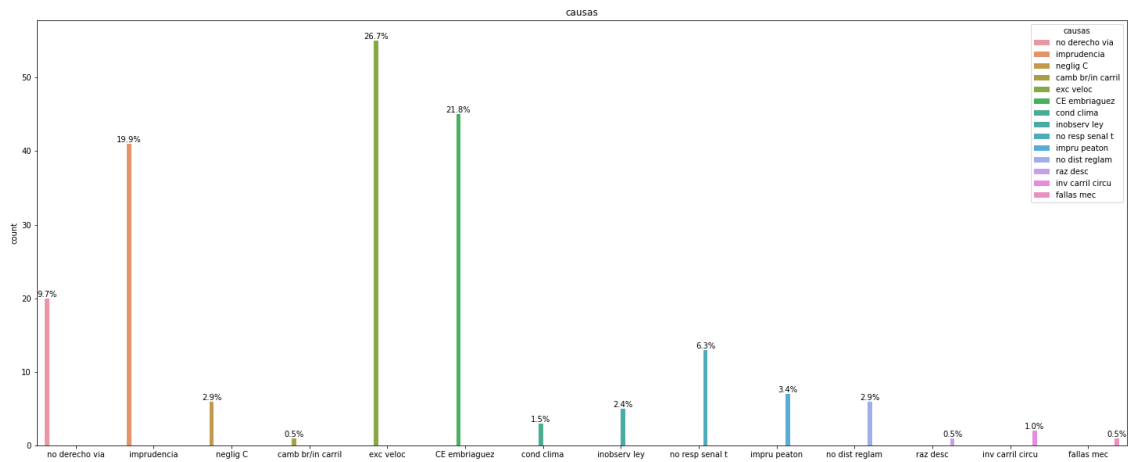


Figura 23. Representación de valores originales de variable "causas".

En la Tabla 16. Exactitud y precisión del modelo de predicción de variable "causas" prueba 1., se presentó el informe del valor de las nuevas métricas de clasificación del modelo; además dentro de la matriz de confusión presentada en la Figura 24, se evidencia la cantidad de los valores predichos por el modelo, siendo acordes al porcentaje generado en la exactitud del modelo.

Tabla 16. Exactitud y precisión del modelo de predicción de variable "causas" prueba 1.

	precision	recall	f1-score	support
1	0,222222	0,044444	0,074074	45
2	0,529412	0,219512	0,310345	41
3	0,240506	0,95	0,383838	20
4	0,534247	0,709091	0,609375	55
5	0	0	0	1
6	0	0	0	1
7	0	0	0	13
8	0	0	0	2
9	0,357143	0,714286	0,47619	7
10	0,428571	1	0,6	6
11	0	0	0	1
12	0	0	0	3
13	0	0	0	6

	precision	recall	f1-score	support
16	0	0	0	5
accuracy	0,38835	0,38835	0,38835	0,38835
macro avg	0,16515	0,25981	0,175273	206
weighted avg	0,344519	0,38835	0,311569	206

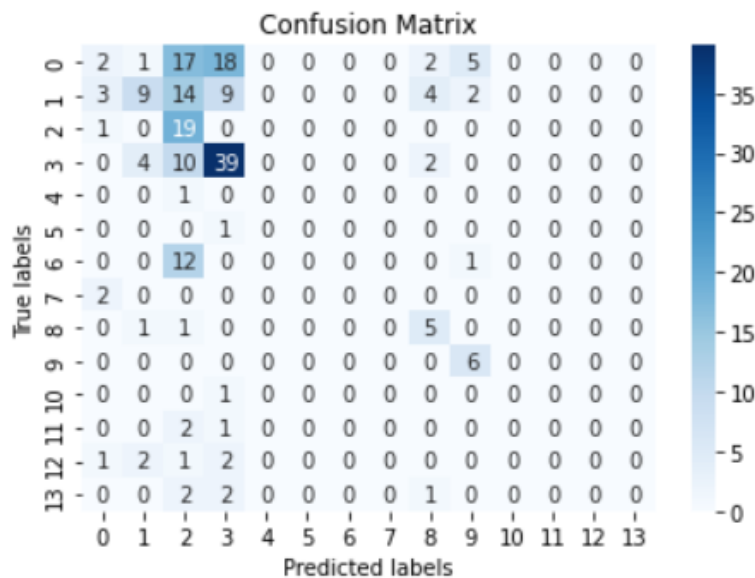


Figura 24. Matriz de confusión de variable "causas" prueba 1.

El presente código generado para la minería de datos de variable “causas” prueba 1 - caso 1, se encuentra presentado en el anexo 21.

Prueba 2 – Caso 2

En esta prueba se presentó las variables generadas por rangos en donde se consideró el rango de las variables “hora”, “día” y “tipología” con los rangos de valores del caso 2, tal como se presenta en el anexo 28.

En la Tabla 17. Exactitud y precisión del modelo de predicción de variable "causas" prueba 2., se presentó el informe de datos generados que permitieron establecer el valor de la exactitud del modelo aplicado; además dentro de la matriz de confusión presentada en la Figura 25, se evidencia la cantidad de los valores predichos, siendo acordes al porcentaje generado en la exactitud del modelo.

Tabla 17. Exactitud y precisión del modelo de predicción de variable "causas" prueba 2.

	precision	recall	f1-score	support
1	0,24	0,266667	0,252632	45
2	0,239437	0,414634	0,303571	41
3	0	0	0	20
4	0,534247	0,709091	0,609375	55
5	0	0	0	1
6	0	0	0	1
7	0	0	0	13
8	0	0	0	2
9	0,416667	0,714286	0,526316	7
10	0	0	0	6
11	0	0	0	1
12	0	0	0	3
13	0	0	0	6
16	0	0	0	5
accuracy	0,354369	0,354369	0,354369	0,354369
macro avg	0,102168	0,150334	0,12085	206
weighted avg	0,256879	0,354369	0,296188	206

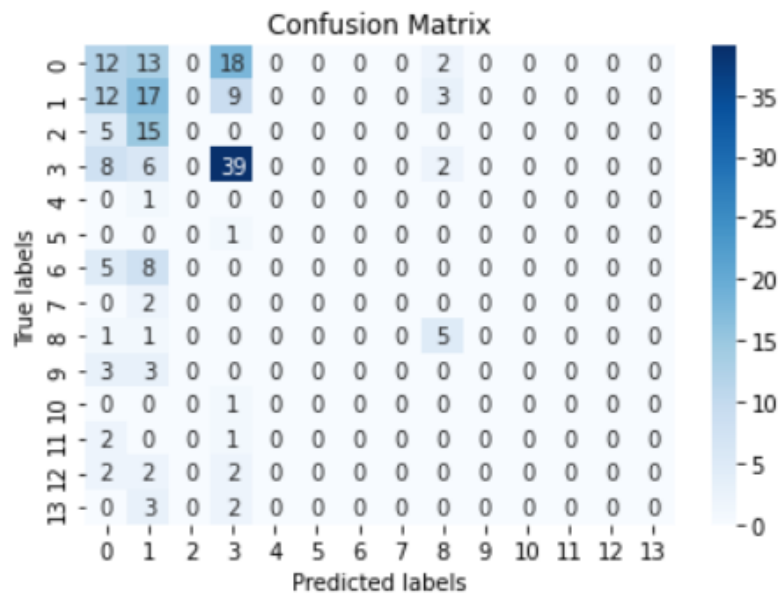


Figura 25. Matriz de confusión de variable "causas" prueba 2.

El presente código generado para la minería de datos de variable “causas” prueba 2 - caso 2, se encuentra presentado en el anexo 22.

Prueba 3 - Caso 3

En esta prueba se presentó las variables generadas por rangos en donde se consideró el rango de las variables “hora”, “dia” y “tipologia” con los rangos de valores del caso 3, tal como se presenta en el anexo 28.

En la Tabla 18. Exactitud y precisión del modelo de predicción de variable "causas" prueba 3., se presentó el informe de datos generados que permitieron establecer el valor de la exactitud del modelo aplicado; además dentro de la matriz de confusión presentada en la Figura 26, se evidencia la cantidad de los valores predichos, siendo acordes al porcentaje generado en la exactitud del modelo.

Tabla 18. Exactitud y precisión del modelo de predicción de variable "causas" prueba 3.

	precision	recall	f1-score	support
1	0,391304	0,2	0,264706	45
2	0,255102	0,609756	0,359712	41
3	0	0	0	20
4	0,534247	0,709091	0,609375	55
5	0	0	0	1
6	0	0	0	1
7	0	0	0	13
8	0	0	0	2
9	0,416667	0,714286	0,526316	7
10	0	0	0	6
11	0	0	0	1
12	0	0	0	3
13	0	0	0	6
16	0	0	0	5
accuracy	0,378641	0,378641	0,378641	0,378641
macro avg	0,114094	0,159509	0,125722	206
weighted avg	0,293049	0,378641	0,309999	206

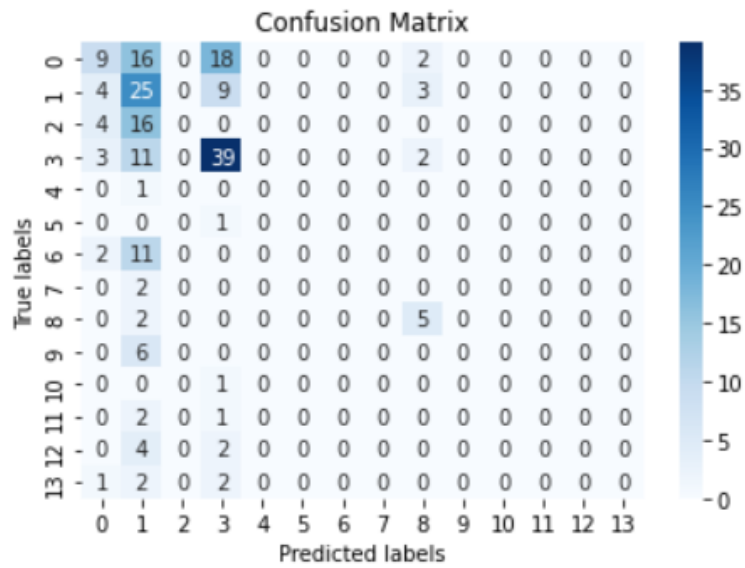


Figura 26. Matriz de confusión de variable "causas" prueba 3.

El presente código generado para la minería de datos de variable “causas” prueba 3 - caso 3, se encuentra presentado en el anexo 23.

Variable “nro_heridos”

Prueba 1 – Caso 1

Como se presenta en la Figura 27, se representaron todos los porcentajes de los 1028 registros contenidos dentro de la variable “nro_heridos”.

En esta prueba se utilizó la variable “nro_heridos” como atributo de referencia para la predicción de datos, en donde las variables restantes permanecieron con los datos originales identificados en el caso 1, tal como se presenta en el anexo 28.

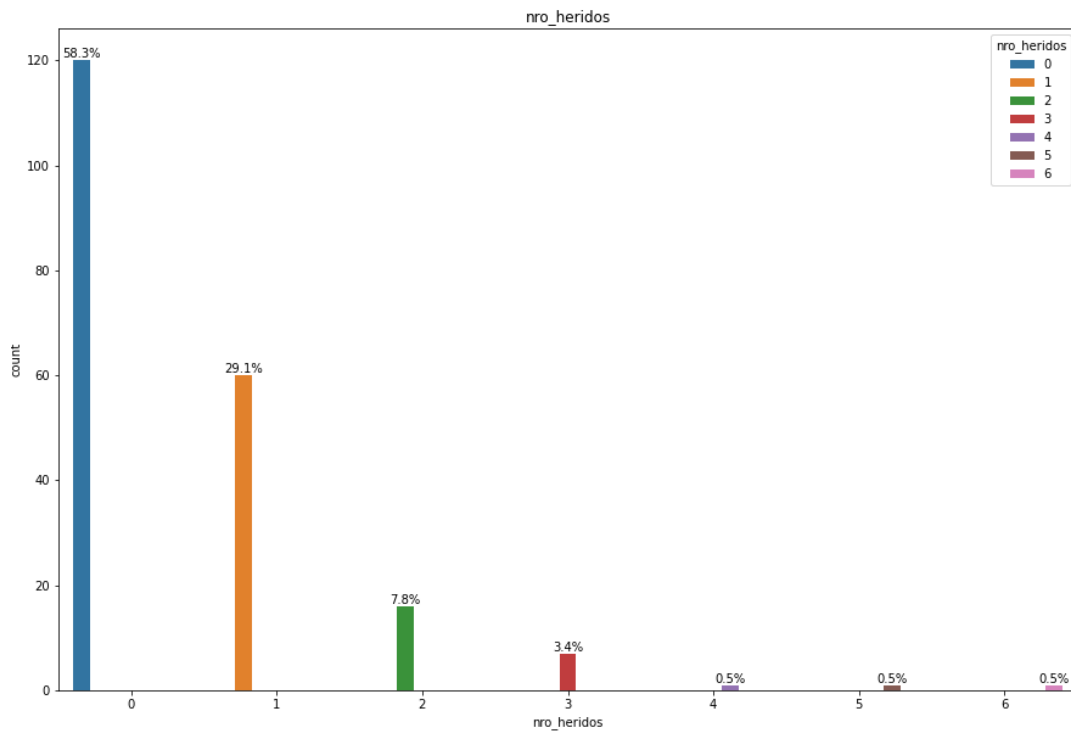


Figura 27. Representación de valores originales de variable "nro_heridos".

En la Tabla 19. Exactitud y precisión del modelo de predicción de variable "nro_heridos" prueba 1., se presentó el informe del valor de las nuevas métricas de clasificación del modelo; además, dentro de la matriz de confusión presentada en la Figura 28, se evidencia la cantidad de los valores predichos por el modelo, siendo acordes al porcentaje generado en la exactitud del modelo.

Tabla 19. Exactitud y precisión del modelo de predicción de variable "nro_heridos" prueba 1.

	precision	recall	f1-score	support
0	0,625668	0,975	0,762215	120
1	0,684211	0,216667	0,329114	60
2	0	0	0	16
3	0	0	0	7
4	0	0	0	1
5	0	0	0	1
6	0	0	0	1
accuracy	0,631068	0,631068	0,631068	0,631068
macro avg	0,187126	0,170238	0,155904	206
weighted avg	0,563752	0,631068	0,539867	206

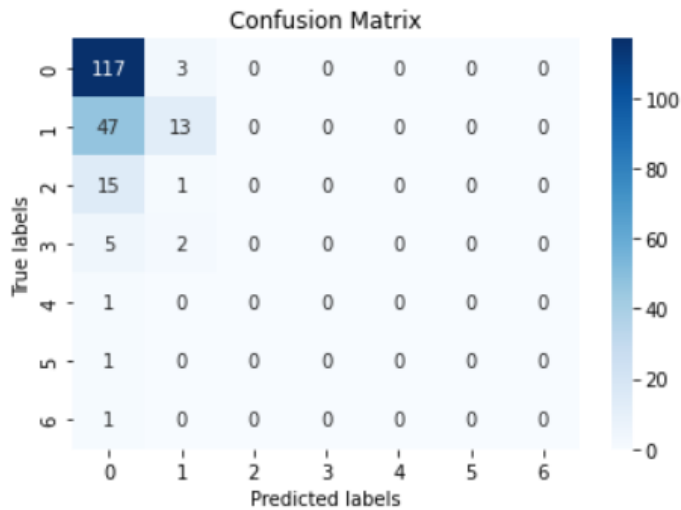


Figura 28. Matriz de confusión de variable "nro_heridos" prueba 1.

El presente código generado para la minería de datos de variable “nro_heridos” prueba 1 - caso 1, se encuentra presentado en el anexo 24.

Prueba 2 – Caso 2

En esta prueba se presentó las variables generadas por rangos en donde se consideró el rango de las variables “hora”, “dia” y “tipologia” con los rangos de valores del caso 2, tal como se presenta en el anexo 28.

En la Tabla 20. Exactitud y precisión del modelo de predicción de variable "nro_heridos" prueba 2., se presentó el informe de datos generados que permitieron establecer el valor de la exactitud del modelo aplicado; además dentro de la matriz de confusión presentada en la Figura 29, se evidencia la cantidad de los valores predichos, siendo acordes al porcentaje generado en la exactitud del modelo.

Tabla 20. Exactitud y precisión del modelo de predicción de variable "nro_heridos" prueba 2.

	precision	recall	f1-score	support
0	0,62766	0,983333	0,766234	120
1	0,722222	0,216667	0,333333	60
2	0	0	0	16
3	0	0	0	7
4	0	0	0	1
5	0	0	0	1
6	0	0	0	1
accuracy	0,635922	0,635922	0,635922	0,635922

	precision	recall	f1-score	support
macro avg	0,19284	0,171429	0,157081	206
weighted avg	0,575983	0,635922	0,543437	206

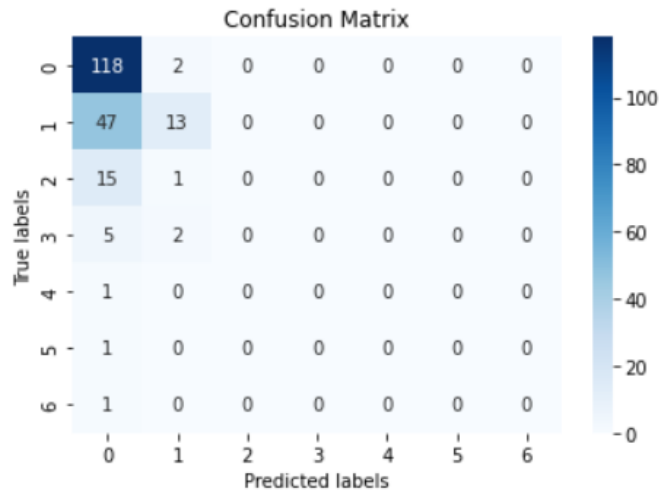


Figura 29. Matriz de confusión de variable "nro_heridos" prueba 2.

El presente código generado para la minería de datos de variable “nro_heridos” prueba 2 - caso 2, se encuentra presentado en el anexo 25.

Prueba 3 – Caso 3

En esta prueba se presentó las variables generadas por rangos en donde se consideró el rango de las variables “hora”, “dia” y “tipologia” con los rangos de valores del caso 3, tal como se presenta en el anexo 28.

En la Tabla 21. Exactitud y precisión del modelo de predicción de variable "nro_heridos" prueba 3., se presentó el informe de datos generados que permitieron establecer el valor de la exactitud del modelo aplicado; además dentro de la matriz de confusión presentada en la Figura 30, se evidencia la cantidad de los valores predichos, siendo acordes al porcentaje generado en la exactitud del modelo.

Tabla 21. Exactitud y precisión del modelo de predicción de variable "nro_heridos" prueba 3.

	precision	recall	f1-score	support
0	0,62766	0,983333	0,766234	120
1	0,722222	0,216667	0,333333	60
2	0	0	0	16
3	0	0	0	7

4	0	0	0	1
5	0	0	0	1
6	0	0	0	1
accuracy	0,635922	0,635922	0,635922	0,635922
macro avg	0,19284	0,171429	0,157081	206
weighted avg	0,575983	0,635922	0,543437	206

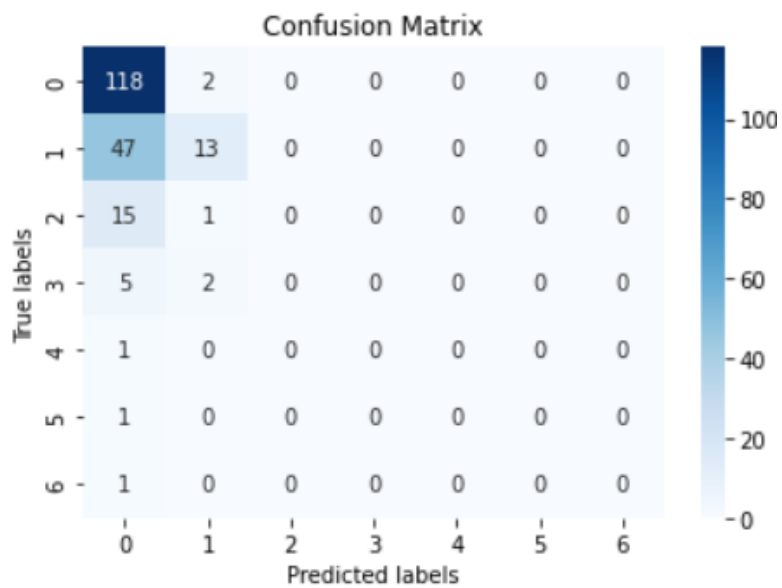


Figura 30. Matriz de confusión de variable "nro_heridos" prueba 3.

El presente código generado para la minería de datos de variable “nro_heridos” prueba 3 - caso 3, se encuentra presentado en el anexo 26.

Variable “nro_fallecidos”

Prueba 1 – Caso 1

En esta prueba se utilizó la variable “nro_fallecidos” como atributo de referencia para la predicción de datos, en donde las variables restantes permanecieron con los datos originales identificados en el caso 1.

En la Tabla 22. Exactitud y precisión del modelo de predicción de variable “nro_fallecidos” prueba 1., se presentó el informe del valor de las nuevas métricas de clasificación del modelo; además dentro de la matriz de confusión presentada en la Figura 31, se evidencia la cantidad de los valores predichos por el modelo, siendo acordes al porcentaje generado en la exactitud del modelo.

Tabla 22. Exactitud y precisión del modelo de predicción de variable “nro_fallecidos” prueba 1.

	precision	recall	f1-score	support
0	0,961165	1	0,980198	198
1	0	0	0	8
accuracy	0,961165	0,961165	0,961165	0,961165
macro avg	0,480583	0,5	0,490099	206
weighted avg	0,923838	0,961165	0,942132	206

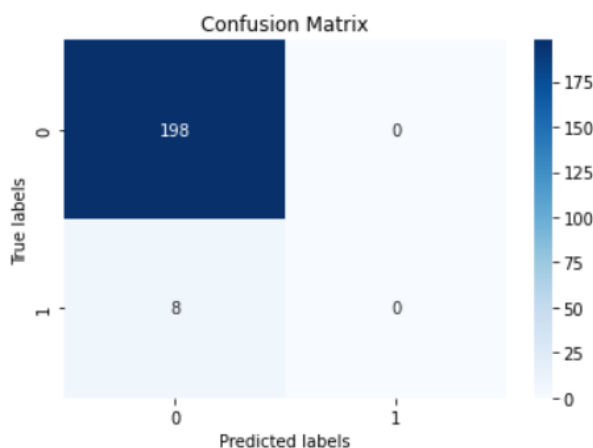


Figura 31. Matriz de confusión de variable “nro_fallecidos” prueba 1.

A mismo el presente código generado para la minería de datos de variable “nro_fallecidos” prueba 1 - caso 1, se encuentra presentado en el anexo 27.

Anexos

1. **Anexo 1.** *Representación de datos originales de variable “dia”.*
2. **Anexo 2.** *Árbol generado de variable “dia” prueba 1.*
3. **Anexo 3.** *Código Python en entorno Google Colab de variable “dia” prueba 1.*
4. **Anexo 4.** *Código Python en entorno Google Colab de variable “dia” prueba 2.*
5. **Anexo 5.** *Código Python en entorno Google Colab de variable “dia” prueba 3.*
6. **Anexo 6.** *Representación de valores originales de variable “hora”.*
7. **Anexo 7.** *Código Python en entorno Google Colab de variable “hora” prueba 1.*
8. **Anexo 8.** *Código Python en entorno Google Colab de variable “hora” prueba 2.*
9. **Anexo 9.** *Código Python en entorno Google Colab de variable “hora” prueba 3.*
10. **Anexo 10.** *Código Python en entorno Google Colab de variable “hora” prueba 4.*
11. **Anexo 11.** *Código Python en entorno Google Colab de variable “hora” prueba 5.*
12. **Anexo 12.** *Representación de valores originales de variable “tipologia”.*
13. **Anexo 13.** *Código Python en entorno Google Colab de variable “tipologia” prueba 1.*
14. **Anexo 14.** *Código Python en entorno Google Colab de variable “tipologia” prueba 2.*
15. **Anexo 15.** *Código Python en entorno Google Colab de variable “tipologia” prueba 3.*
16. **Anexo 16.** *Código Python en entorno Google Colab de variable “tipologia” prueba 4.*
17. **Anexo 17.** *Representación de valores originales de variable “parroquia_urbana”.*
18. **Anexo 18.** *Código Python en entorno Google Colab de variable “parroquia_urbana” prueba 1.*
19. **Anexo 19.** *Código Python en entorno Google Colab de variable “parroquia_urbana” prueba 2.*
20. **Anexo 20.** *Código Python en entorno Google Colab de variable “parroquia_urbana” prueba 3.*
21. **Anexo 21.** *Código Python en entorno Google Colab de variable “causas” prueba 1.*

- 22. Anexo 22.** *Código Python en entorno Google Colab de variable “causas” prueba 2.*
- 23. Anexo 23.** *Código Python en entorno Google Colab de variable “causas” prueba 3.*
- 24. Anexo 24.** *Código Python en entorno Google Colab de variable “nro_heridos” prueba 1.*
- 25. Anexo 25.** *Código Python en entorno Google Colab de variable “nro_heridos” prueba 2.*
- 26. Anexo 26.** *Código Python en entorno Google Colab de variable “nro_heridos” prueba 3.*
- 27. Anexo 27.** *Código Python en entorno Google Colab de variable “nro_fallecidos” prueba 1.*
- 28. Anexo 28.** *Casos para selección de rangos de variables.*