TECNOLOGICO NACIONAL DE MEXICO

INSTITUTO TECNOLOGICO DE NUEVO LAREDO

INTELIGENCIA ARTIFICIAL 2

INGENIERIA EN SISTEMAS COMPUTACIONALES

DOCENTE: LUIS DANIEL CASTILLO GARCIA

U2 – PRACTICA 3 – VISUALIZACION DE DATOS

JOEL RODRIGUEZ MUÑOZ

NUMERO DE CONTROL: 19100244

NUEVO LAREDO TAMAULIPAS.

29 octubre de 2024

Visualización de Datos

Utilizando el mismo set de información vistas en la explicación:

- Obtener los atributos con mayor correlación
- Graficas dichos atributos
- Ordenar los datos por un atributo de libre elección
- Guardar el conjunto ordenado en formato CSV
- Entregar el CSV en conjunto con el PDF

Obtención de los atributos con mayor correlación

```
from sklearn.preprocessing import LabelEncoder
labelencoder =LabelEncoder()
df["class"]=labelencoder.fit_transform(df["class"])
df
```

[23] ✓ 0.1s Python

	duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	...	dst_host_srv_count	dst_host_same_srv_rate	dst_
0	0.0	tcp	ftp_data	SF	491.0	0.0	0	0.0	0.0	0.0	...	25.0	0.17	
1	0.0	udp	other	SF	146.0	0.0	0	0.0	0.0	0.0	...	1.0	0.00	
2	0.0	tcp	private	S0	0.0	0.0	0	0.0	0.0	0.0	...	26.0	0.10	
3	0.0	tcp	http	SF	232.0	8153.0	0	0.0	0.0	0.0	...	255.0	1.00	
4	0.0	tcp	http	SF	199.0	420.0	0	0.0	0.0	0.0	...	255.0	1.00	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
125968	0.0	tcp	private	S0	0.0	0.0	0	0.0	0.0	0.0	...	25.0	0.10	
125969	8.0	udp	private	SF	105.0	145.0	0	0.0	0.0	0.0	...	244.0	0.96	
125970	0.0	tcp	smtp	SF	2231.0	384.0	0	0.0	0.0	0.0	...	30.0	0.12	
125971	0.0	tcp	klogin	S0	0.0	0.0	0	0.0	0.0	0.0	...	8.0	0.03	
125972	0.0	tcp	ftp_data	SF	151.0	0.0	0	0.0	0.0	0.0	...	77.0	0.30	

125973 rows × 42 columns

Obtenemos las 3 correlaciones mas altas con respecto a “duration” que en este caso son:

- dst_host_diff_srv_rate 0.254195
- dst_host_same_src_port_rate 0.228737
- rerror_rate 0.200682

```
# Calcular la matriz de correlación solo con columnas numéricas
corr_matrix = df.corr(numeric_only=True)

# Seleccionar las correlaciones con respecto a la columna 'duration' y ordenarlas
duration_corr=corr_matrix["duration"].sort_values(ascending=False)

print(duration_corr)
```

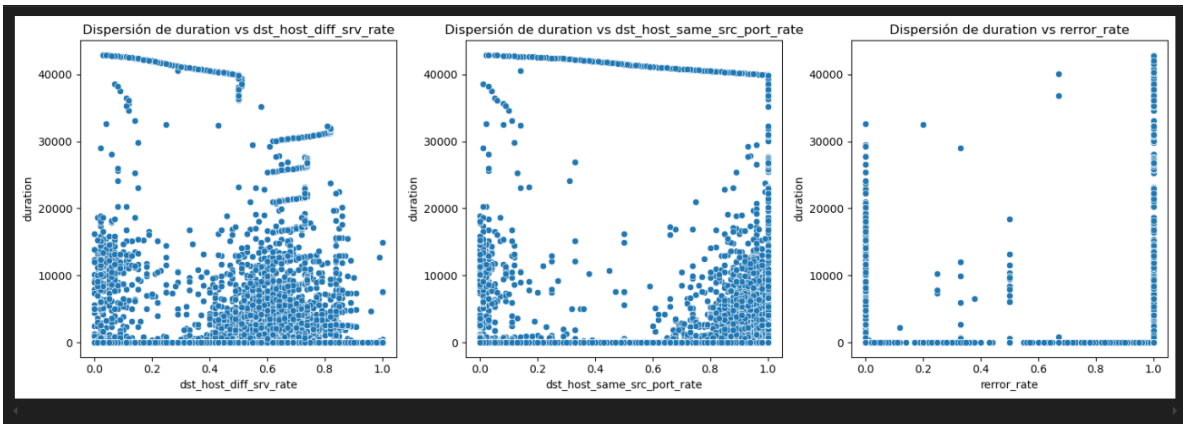
[24] ✓ 0.2s

```
duration                     1.000000
dst_host_diff_srv_rate       0.254195
dst_host_same_src_port_rate  0.228737
rerror_rate                  0.200682
srv_rerror_rate              0.199961
dst_host_srv_rerror_rate     0.199024
dst_host_rerror_rate         0.173815
num_file_creations           0.099116
su_attempted                 0.087183
same_srv_rate                0.074681
src_bytes                    0.070737
num_access_files             0.070420
root_shell                   0.052791
```

```python
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Selecciona los atributos con mayor correlación con 'duration'
# Por ejemplo, seleccionamos los 3 atributos más correlacionados (ajusta el número según lo necesites)
top_correlated = duration_corr.index[1:4]  # Excluyendo 'duration' en sí mismo

# Crear gráficos de dispersión
plt.figure(figsize=(15, 5))
for i, attr in enumerate(top_correlated, 1):
    plt.subplot(1, 3, i)
    sns.scatterplot(data=df, x=attr, y="duration")
    plt.title(f'Dispersión de duration vs {attr}')
    plt.xlabel(attr)
    plt.ylabel("duration")

plt.tight_layout()
plt.show()
```

[25] ✓ 0.8s

Graficamos la correlación que tiene el atributo de “Duration” con los 3 correlaciones mayores que tenemos



Guardamos el conjunto de datos en un CSV en una ruta específica

```python
# Ordenar el DataFrame por el atributo 'duration' en orden ascendente
# Puedes cambiar 'class' por cualquier otro atributo
df_sorted = df.sort_values(by="duration", ascending=True)

# Especifica la ruta completa donde deseas guardar el archivo CSV
ruta_guardado = "C:\\Users\\YoelR\\Desktop\\IA2\\Practica 3\\NSL-KDD\\datos_ordenados.csv"
df_sorted.to_csv(ruta_guardado, index=False)

print(f"El archivo CSV se ha guardado en {ruta_guardado}")
```

[26] ✓ 1.4s

```
El archivo CSV se ha guardado en C:\Users\YoelR\Desktop\IA2\Practica 3\NSL-KDD\datos_ordenados.csv
```

+ Code + Markdown

Repositorio:

https://github.com/YoelRM/IA2.git