

Classification Of Web Application Attacks

Talha Osman Saraç 150118006

Mehmet Akgün 150119061

Serdar Alsan 150120034

Mehmet Alıcı 150118060

Dataset

CSIC (Spanish National Research Council) 2010 Web Application Attacks

csic_database.csv (29.54 MB)


Detail

Compact

Column

About this file

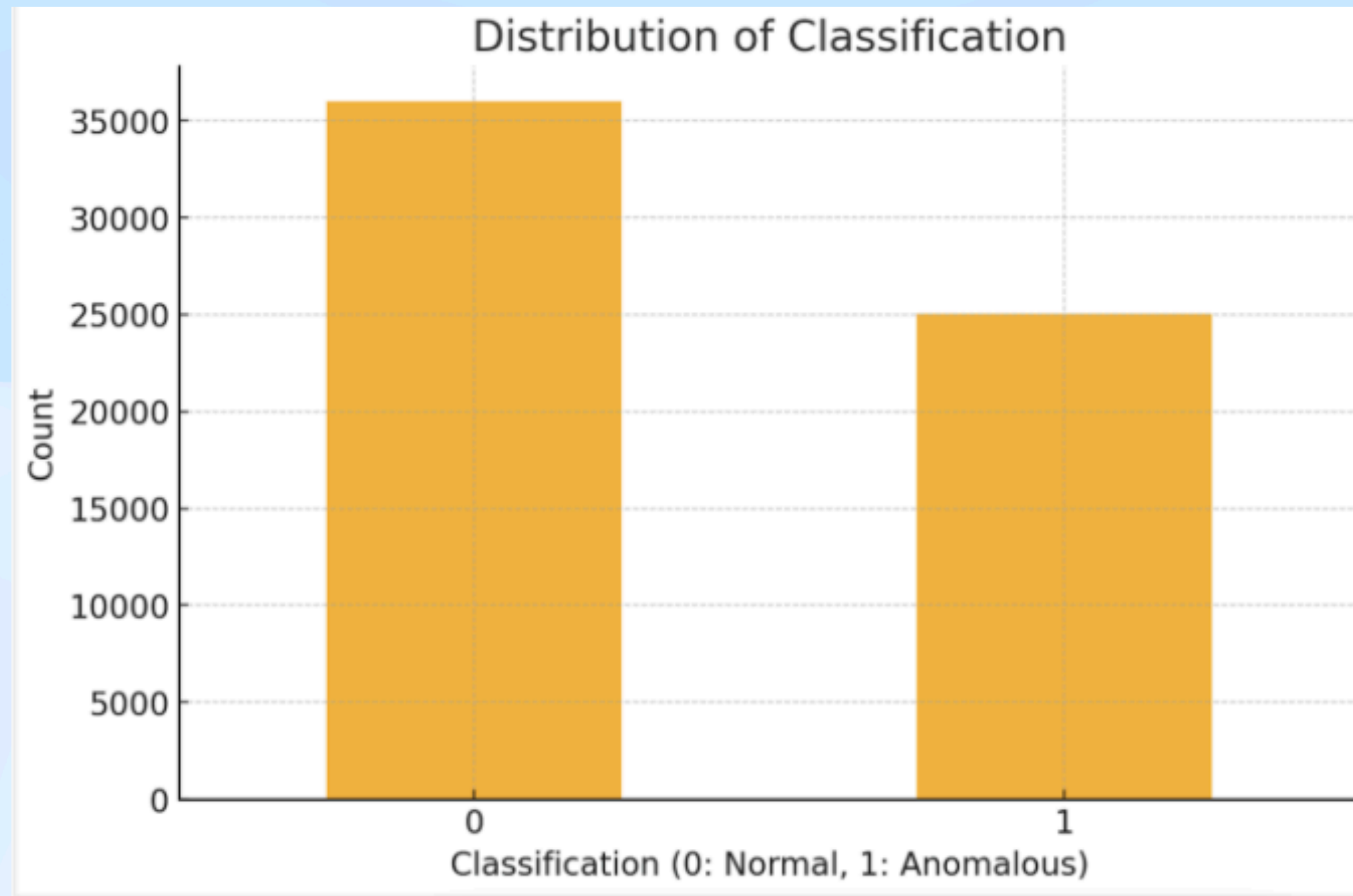
This file does not have a description yet.

⬆	≡	⬆ Method	≡	⬆ content-type	≡	⬆ connection	≡	⬆ lenght	≡	⬆ content	≡	# classification	≡	🔗 URL	≡
Normal	59%	GET	71%	[null]	71%	close	71%	[null]	71%	[null]	71%			http://localhost:80...	4%
Anomalous	41%	POST	29%	application/x-ww...	29%	Connection: close	29%	Content-Length: 4	2%	B2=Vaciar+carrito	2%			http://localhost:80...	4%
		Other (397)	1%					Other (16920)	28%	Other (16931)	28%			Other (56202)	92%
Normal		GET				close						0		http://localhost:8080/tienda1/index.jsp HTTP/1.1	
Normal		GET				close						0		http://localhost:8080/tienda1/publico/anadir.jsp?id=3&nombre=Vino+Rioja&precio=100&cantidad=55&B1=A%...	
Normal		POST		application/x-www-form-urlencoded		Connection: close		Content-Length: 68		id=3&nombre=Vino+Rioja&precio=100&cantidad=55&B1=A%F1adir+al+carrito		0		http://localhost:8080/tienda1/publico/anadir.jsp HTTP/1.1	
Normal		GET				close						0		http://localhost:8080/tienda1/publico/au	

Distribution of the dataset

Normal (0): 36,000 entries

Anomalous (1): 25,065 entries



Feature extraction

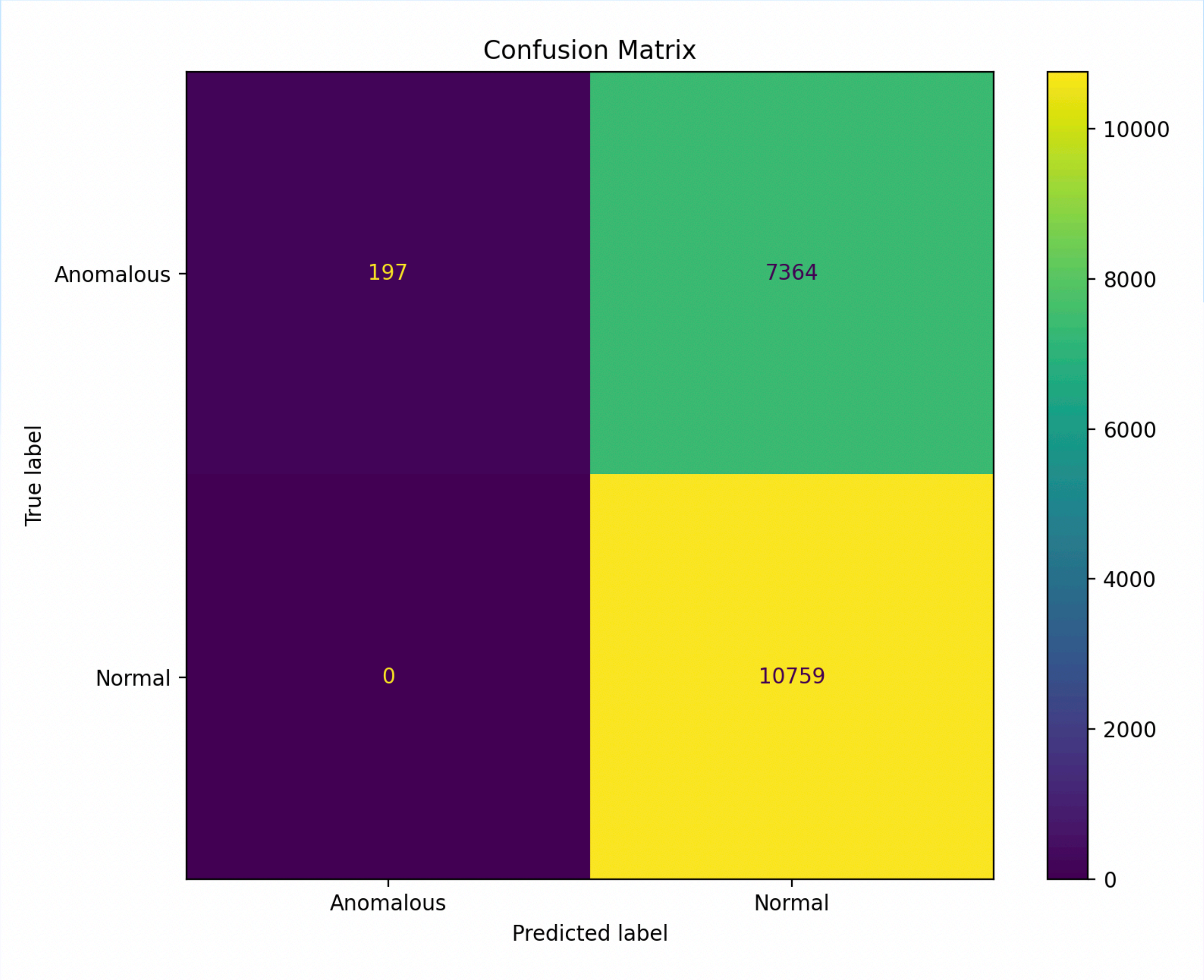
Using Features from URLs as zeros and ones

has_index_jsp	has_percent_login	has_anadir_jsp	has_entrar_login	has_pagar	has_menum	has_titulo	has_miembros	has_estilos	has_imagenes	has_caracter
1	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	1	0	1	0
0	0	0	0	0	0	0	1	0	1	0
0	0	0	0	0	0	0	1	0	1	0

Confusion Matrix Analysis

Gaussian Naive Bayes

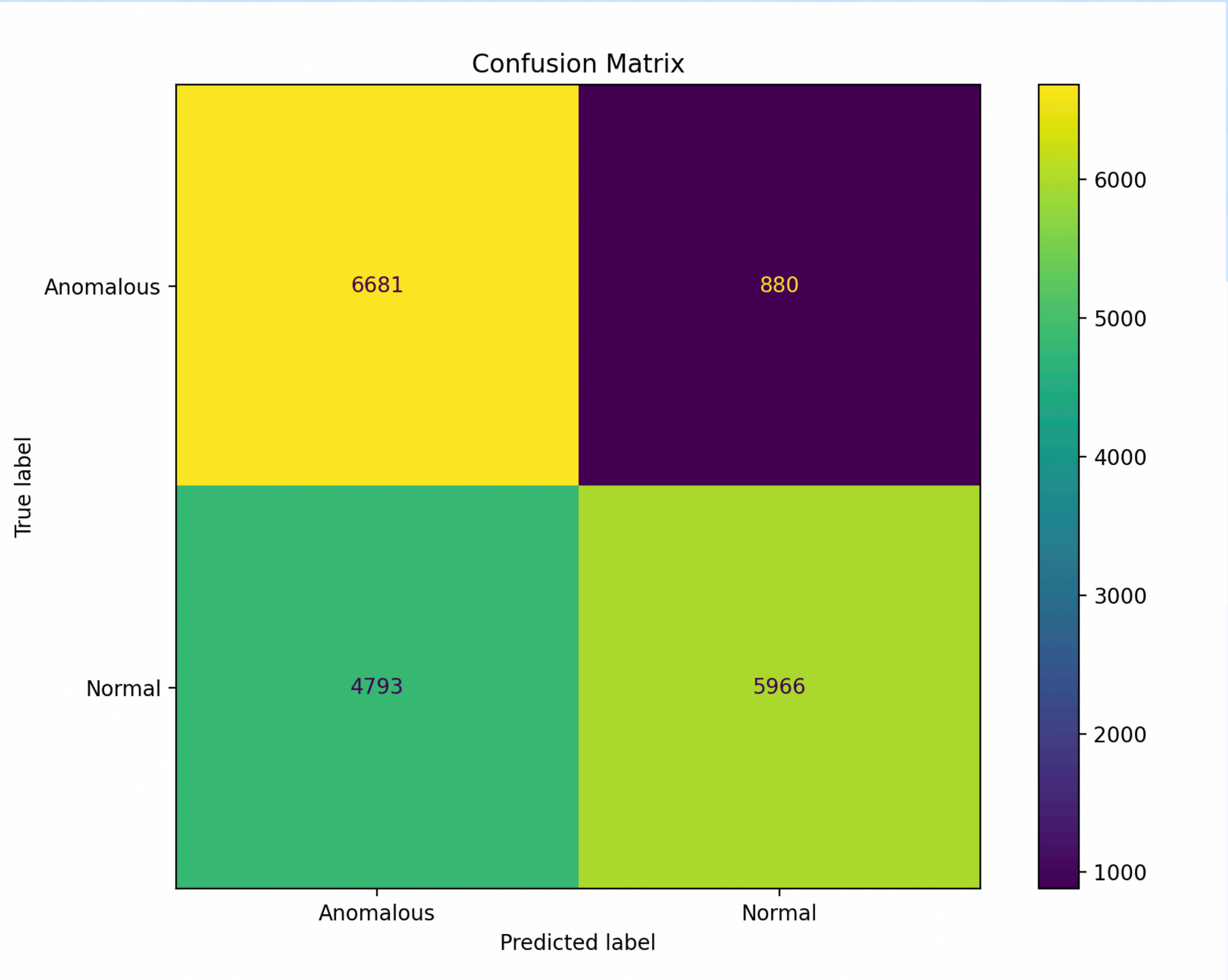
precision (macro avg)	recall (macro avg)	f1-score (attack)	support (all data)
0.80	0.51	0.05	18320



Confusion Matrix Analysis

Decision tree

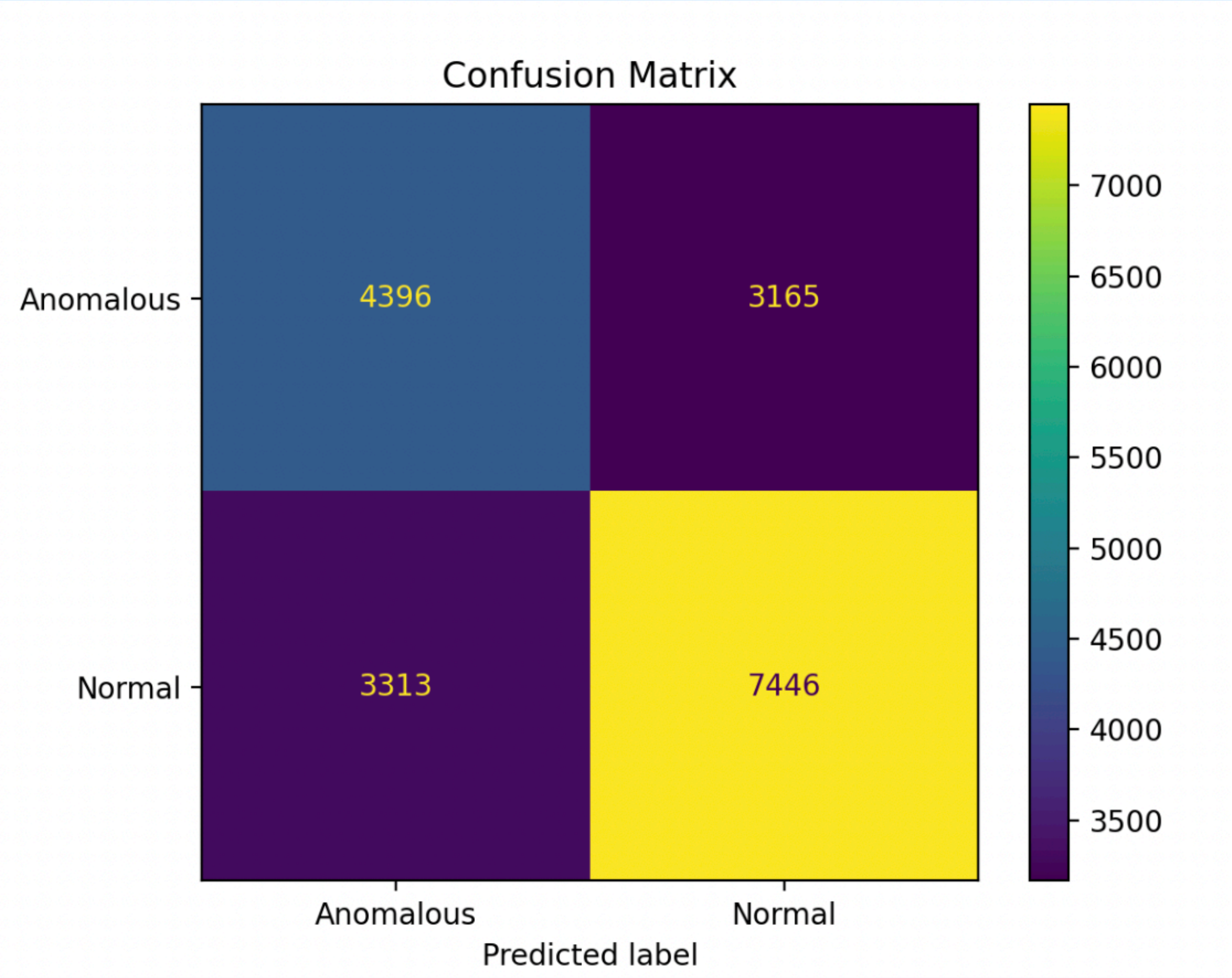
precision (macro avg)	recall (macro avg)	f1-score (attack)	support (all data)
0.73	0.72	0.70	18320



Confusion Matrix Analysis

Logistic Regression

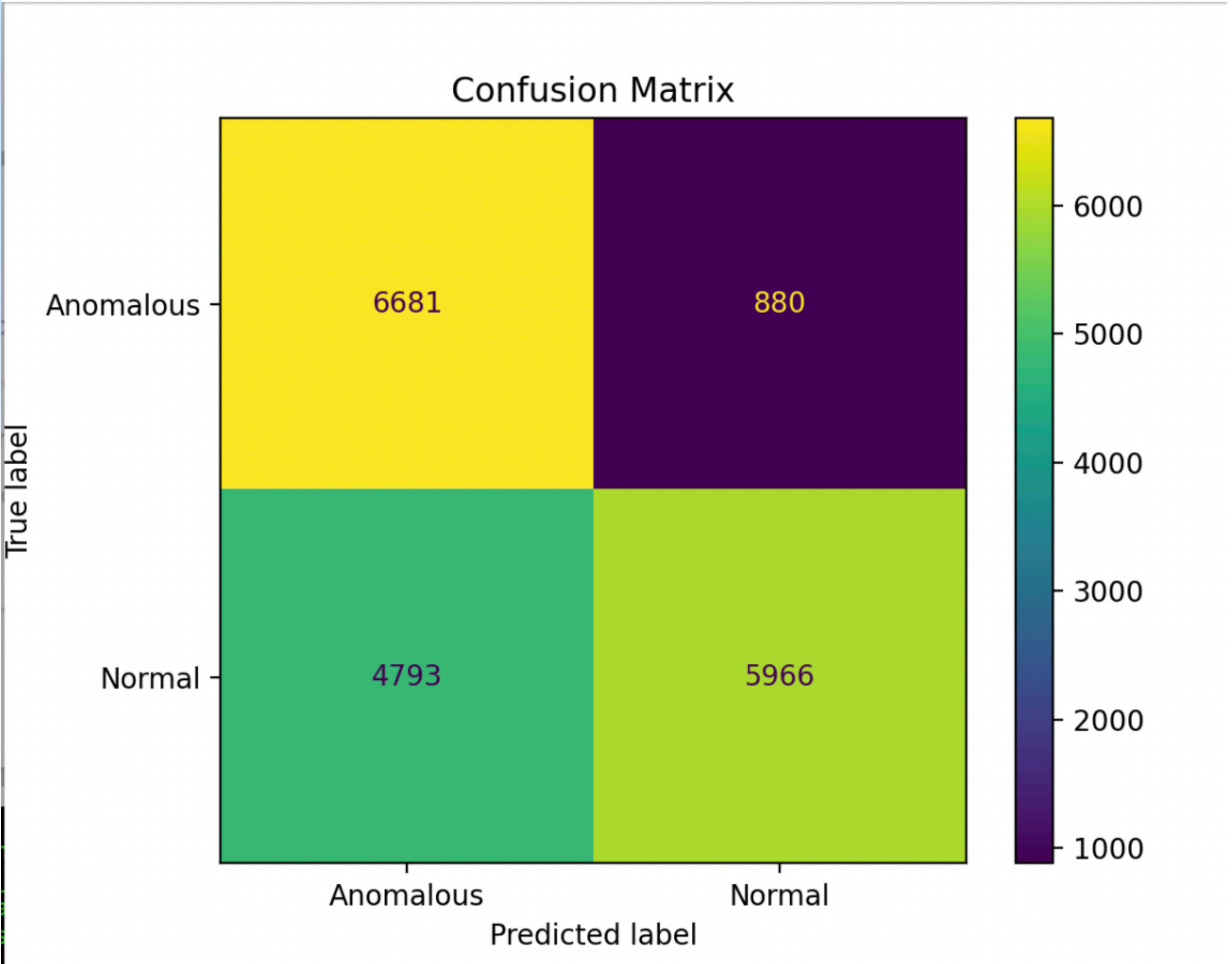
precision (macro avg)	recall (macro avg)	f1-score (attack)	support (all data)
0.64	0.64	0.58	18320



Confusion Matrix Analysis

Random Forest

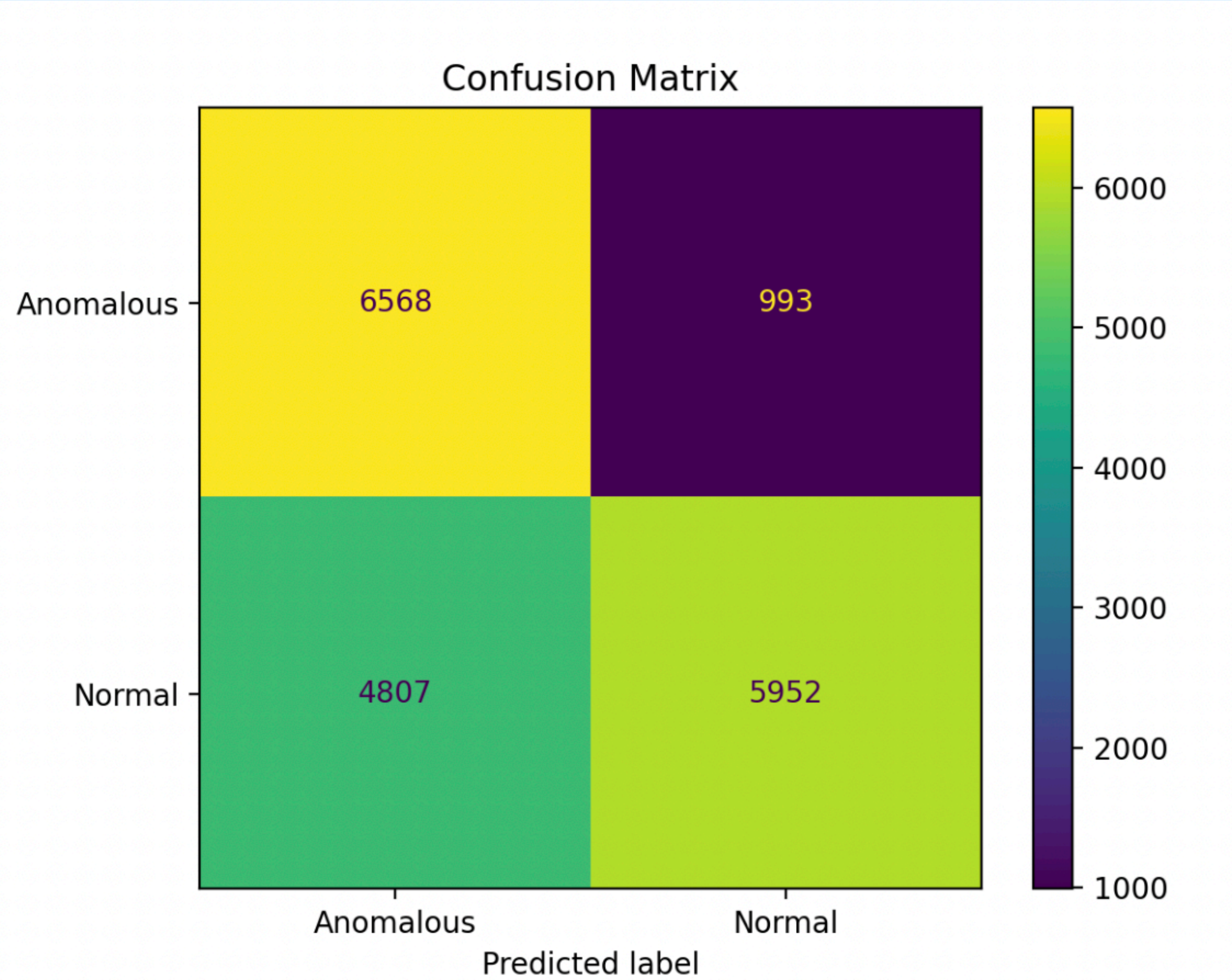
precision (macro avg)	recall (macro avg)	f1-score (attack)	support (all data)
0.73	0.72	0.70	18320



Confusion Matrix Analysis

Gradient Boost

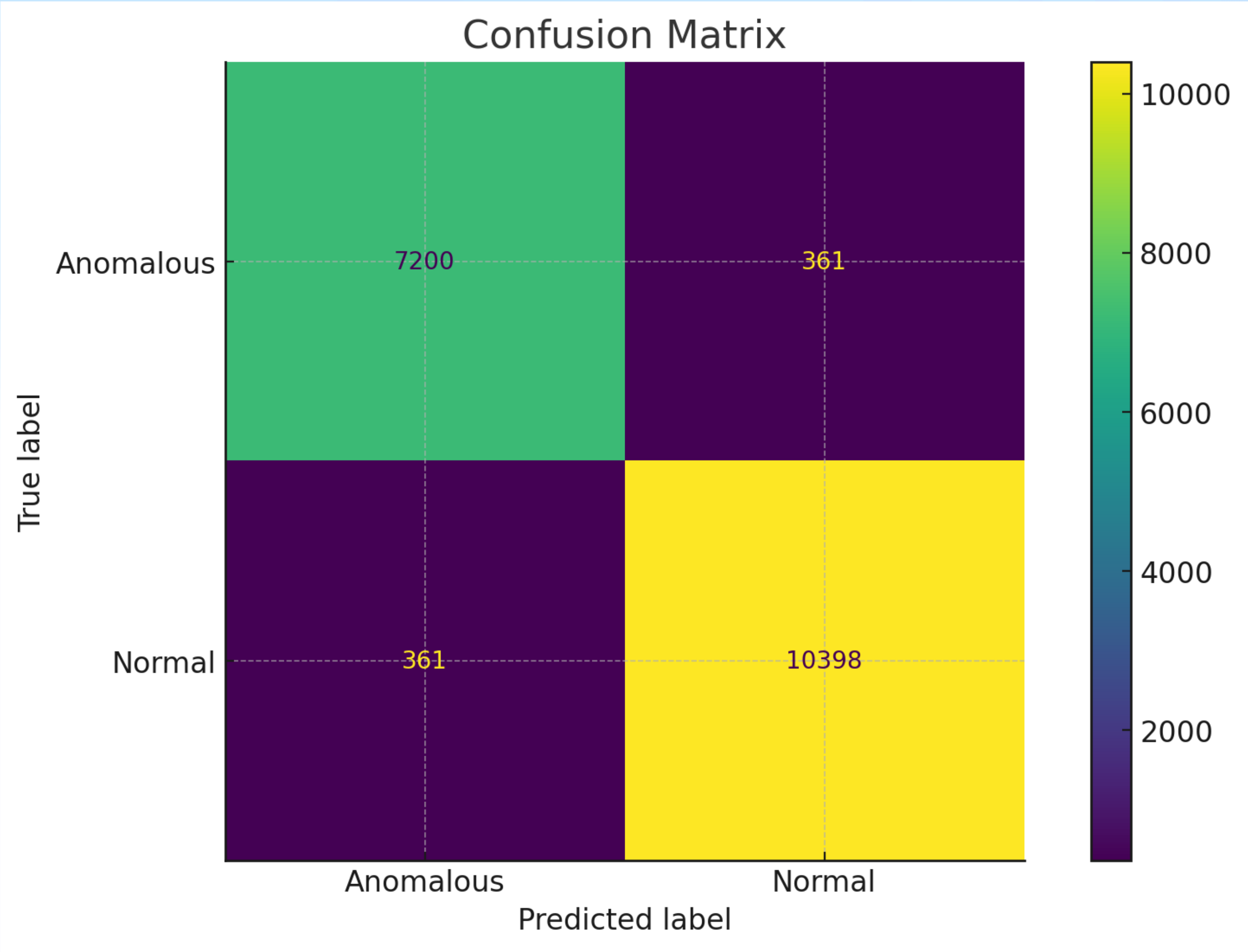
precision (macro avg)	recall (macro avg)	f1-score (attack)	support (all data)
0.72	0.71	0.69	18320



Confusion Matrix Analysis

LSTM (character-level LSTM-based binary classification model)

precision (macro avg)	recall (macro avg)	f1-score (attack)	support (all data)
0.84	0.85	0.83	18320



Performance Metrics

Comparison of Model Performance Metrics for Anomaly Classification

Model	Precision (macro avg)	Recall (macro avg)	F1-score (Anomaly)
Gaussian Naive Bayes	0.8	0.51	0.05
Logistic Regression	0.64	0.64	0.58
Random Forest	0.73	0.72	0.70
Gradient Boost	0.72	0.71	0.69
LSTM (Character Level)	0.84	0.85	0.83