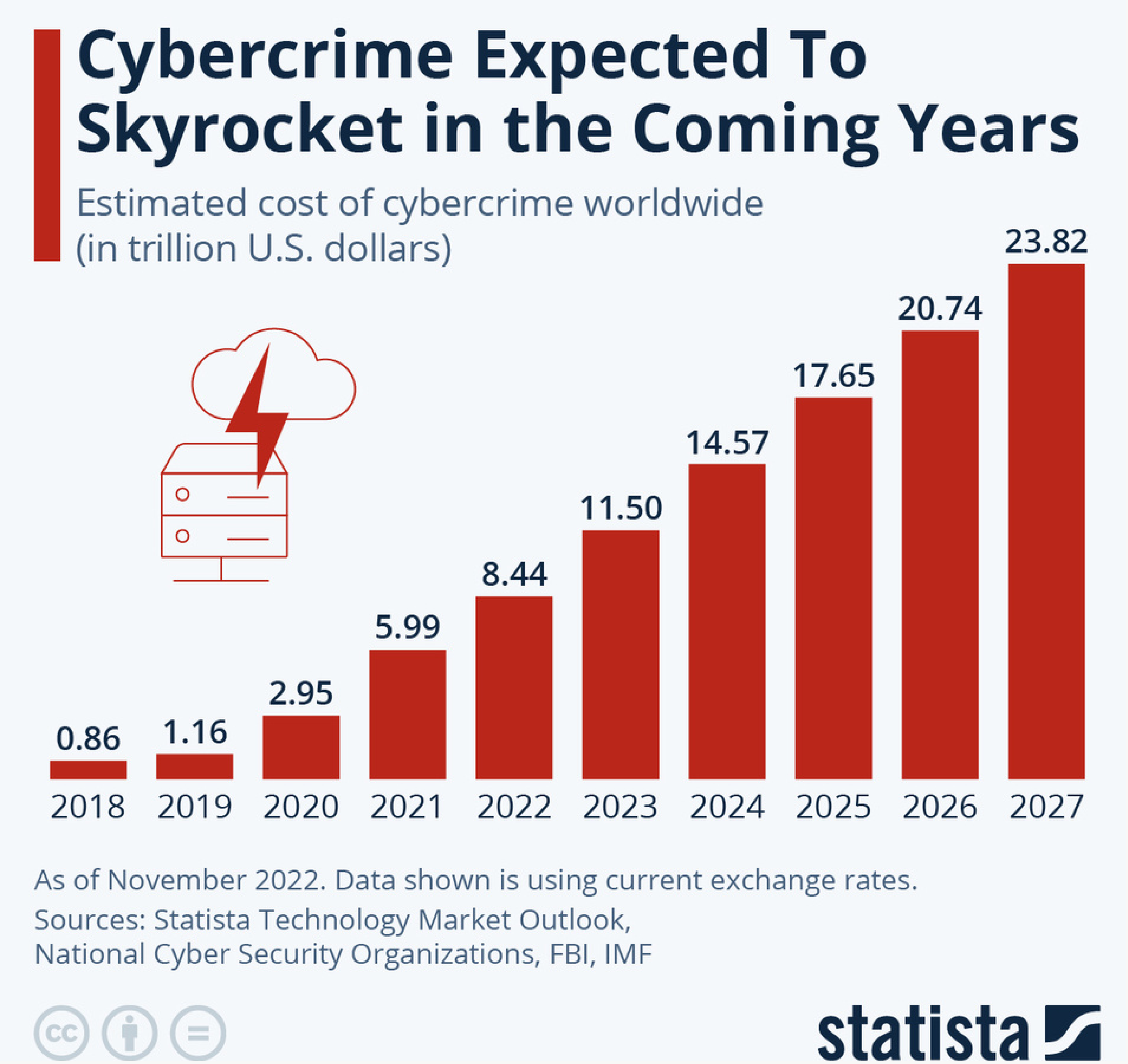
Machine learning and XAI in cybersecurity

Exploring the CICIDS 2019 Dataset and SHAP

–By JUNAID and SIVAMOHAN

Introduction:

Cybersecurity has become a critical concern in our interconnected digital world. The increasing sophistication of cyber threats poses a significant challenge to safeguarding sensitive information and critical infrastructure. Traditional rule-based security systems struggle to keep pace with the evolving nature of these threats. Machine Learning (ML) emerges as a promising solution, providing the capability to detect and mitigate cyber-attacks in real-time.



Cybercrime Expected To Skyrocket in the Coming Years

Estimated cost of cybercrime worldwide (in trillion U.S. dollars)

As of November 2022. Data shown is using current exchange rates.
Sources: Statista Technology Market Outlook, National Cyber Security Organizations, FBI, IMF

WHY XAI?

While ML enhances the capability to detect and prevent cyber threats, the ”black box” nature of many complex models raises concerns about **interpretability and trust**. eXplainable AI (XAI) addresses this challenge by providing insights into the decision-making processes of ML models. In the context of cybersecurity, understanding **why a model flags certain activities as malicious is crucial** for effective decision support and response. XAI techniques, such as feature **importance analysis** and **model-agnostic explanations,** contribute to the **transparency** and **trustworthiness** of ML-based cybersecurity systems.

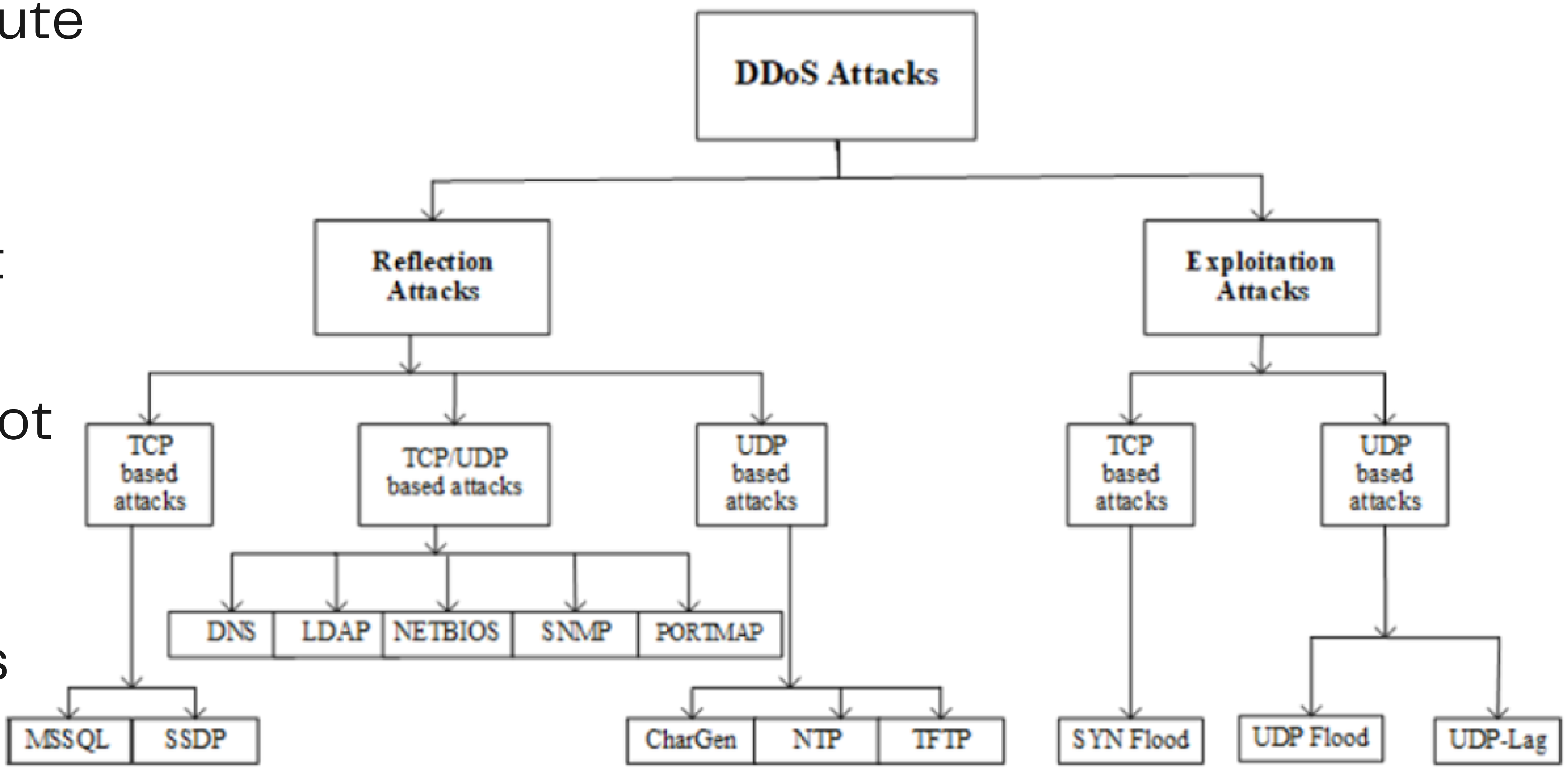
By enhancing the interpretability of ML models, XAI facilitates collaboration between human analysts and automated systems, resulting in a more effective and accountable defense against cyber threats.

Problem Statement

This research centers on utilizing machine learning to detect UDP and LDAP attacks within the CICIDS 2019 Dataset. The study encompasses the entire process, including model development, training, and evaluation. Additionally, eXplainable Artificial Intelligence (XAI) techniques are investigated to enhance the interpretability of the model's decision-making.

About Dataset:

The CICIDS (Canadian Institute for Cybersecurity Intrusion Detection Systems) 2019 Dataset is a synthetic dataset created for research and evaluation purposes. It was not collected from real-world network traffic but was generated to simulate various network scenarios and cyberattacks. The complete data set consists of many different types of attacks, including UDP and LDAP.



About LDAP

An LDAP Denial of Service attack targets LDAP servers by overwhelming them with a high volume of requests, causing the server to become unresponsive to legitimate users.
Disrupt the normal functioning of the LDAP service, leading to service downtime or degradation.

About UDP

A UDP flood attack overwhelms a target system by inundating it with a large volume of User Datagram Protocol (UDP) packets to random ports.
Disrupt the target system's normal operation by consuming its resources, potentially leading to slow responsiveness or unavailability.

Data preprocessing:

Data preprocessing is a crucial step in preparing the CICIDS 2019 dataset, and specifically the DrDoS LDAP and UDP subsets, for machine learning analysis. The dataset contains critical features such as Source IP address and Destination IP address, which were initially represented as non-integer values for our machine learning model. To address this, the IP addresses were segmented into four distinct integers—Source IP1,Source IP2, Source IP3, and Source IP4—by employing a period ('.') as the delimiter. For instance, if the Source IP column contains the value 104.110.151.222, it is split into 104, 110, 151,
and 222, respectively.

Data preprocessing:

The Time Stamp feature adds a layer of temporal context, capturing the evolving nature of the data by quantifying the temporal gaps from the dataset's starting point.Certain columns deemed less significant were removed from the dataset, specifically those whose names concluded with ”max” or ”min.” As a result, attributes such as ’Fwd Packet Length Max’ and ’Fwd Packet Length Min’ were excluded. Subsequently, the data underwent standardization using the formula $(X-\mu) / \sigma$ for each column. Prior to standardization, rows containing exceptionally high values that impeded the standardization process were eliminated

Data preprocessing:

Notably, the dataset exhibited substantial bias, with 1,612 instances labeled as ’attacked’ and 2,179,930 labeled as ’BENIGN’. To mitigate this imbalance, both undersampling for ’BENIGN’ and oversampling for ’attacked’ were implemented. This resulted in a balanced dataset containing 50,000 samples for each class. Both undersampling and oversampling techniques were implemented utilizing the ’resample’ function from the ’sklearn’ library.

PCA and Eigenvalues

01

Principal Component Analysis (PCA)

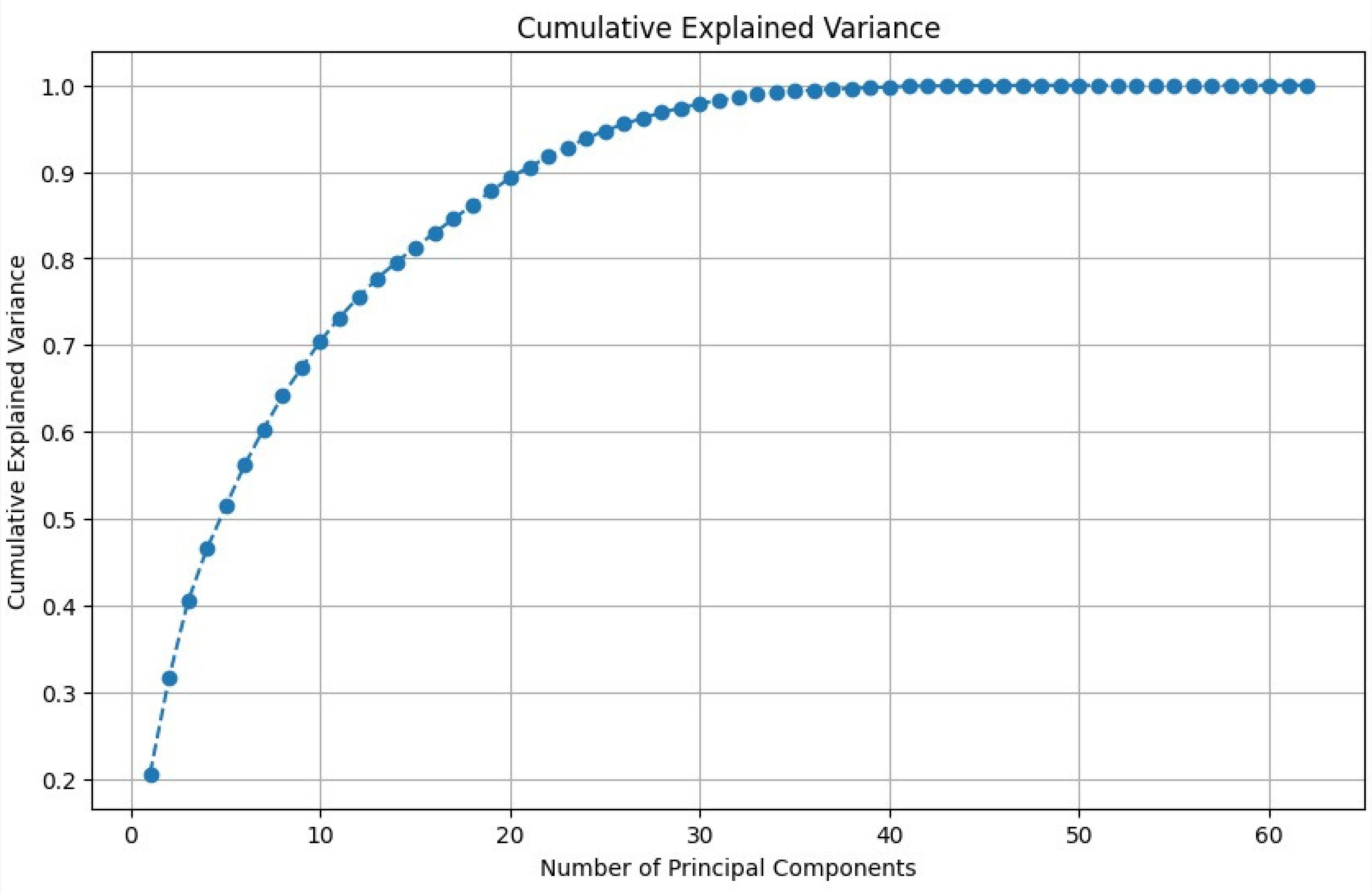
Principal Component Analysis (PCA) is a dimensionality reduction technique used to transform high-dimensional data into a lower-dimensional space. By identifying principal components, which are linear combinations of original features, PCA retains the most important information while reducing the dataset's complexity. The principal components are ordered by the amount of variance they explain, providing a concise representation of the data.

PCA and Eigenvalues

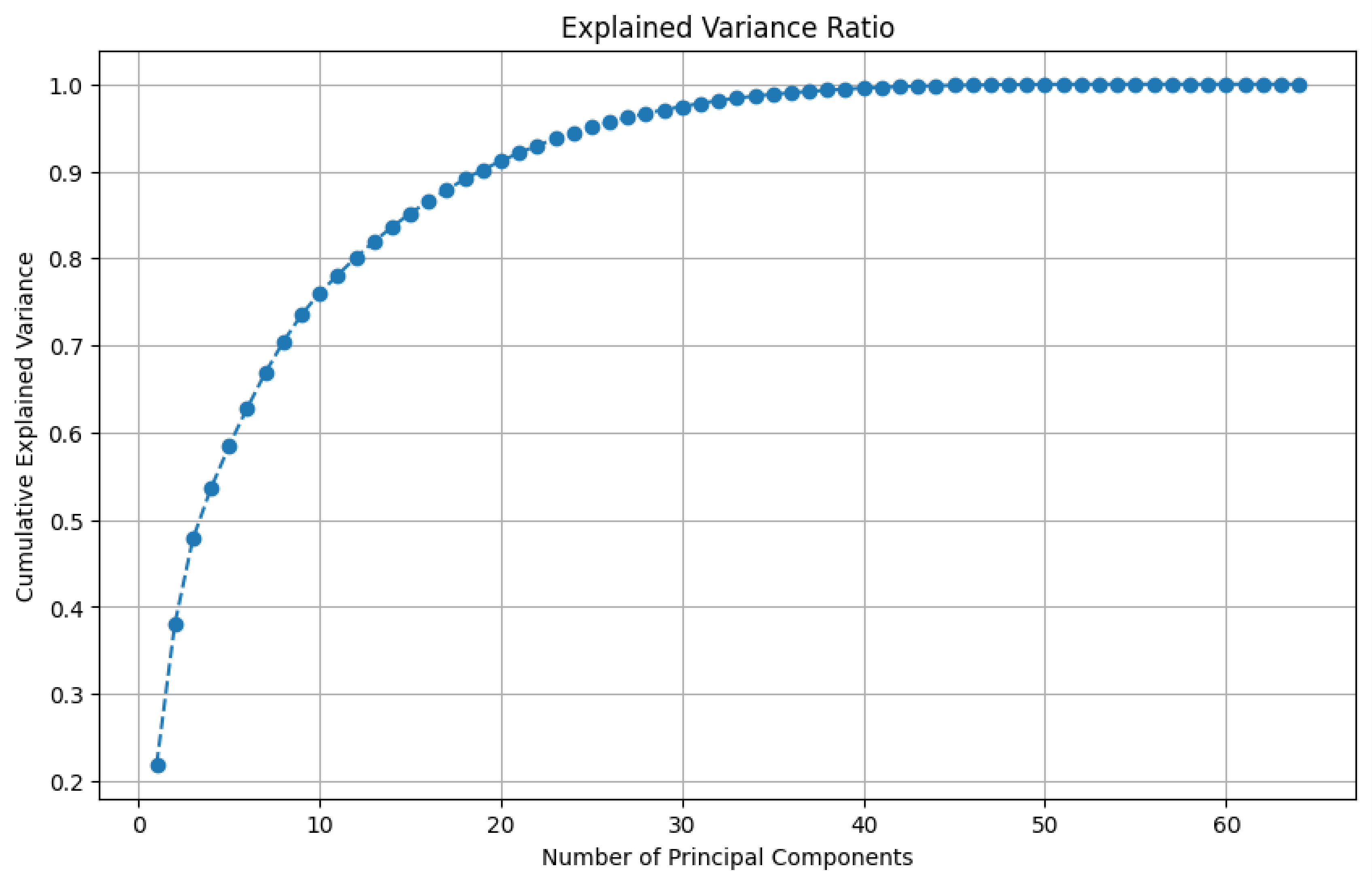
02 Eigenvalues:

Eigenvalues in PCA quantify the variance captured by each principal component. Each eigenvalue corresponds to a principal component and represents the amount of variance in the data along that direction. Higher eigenvalues indicate more significant variability. Eigenvalues help in selecting the most influential principal components, guiding the dimensionality reduction process. In essence, eigenvalues provide a quantitative measure of the importance of each principal component in the dataset.

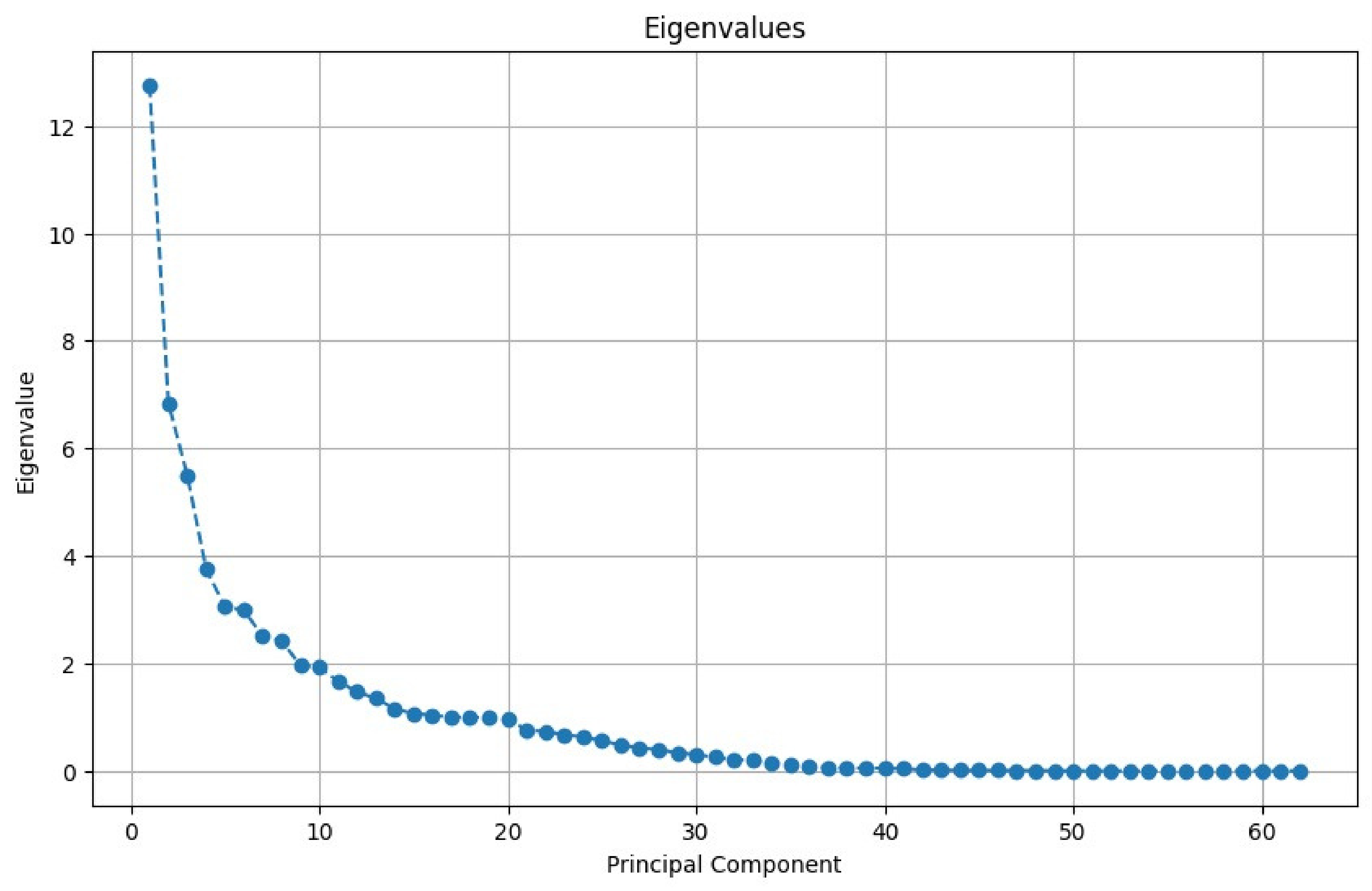
For LDAP



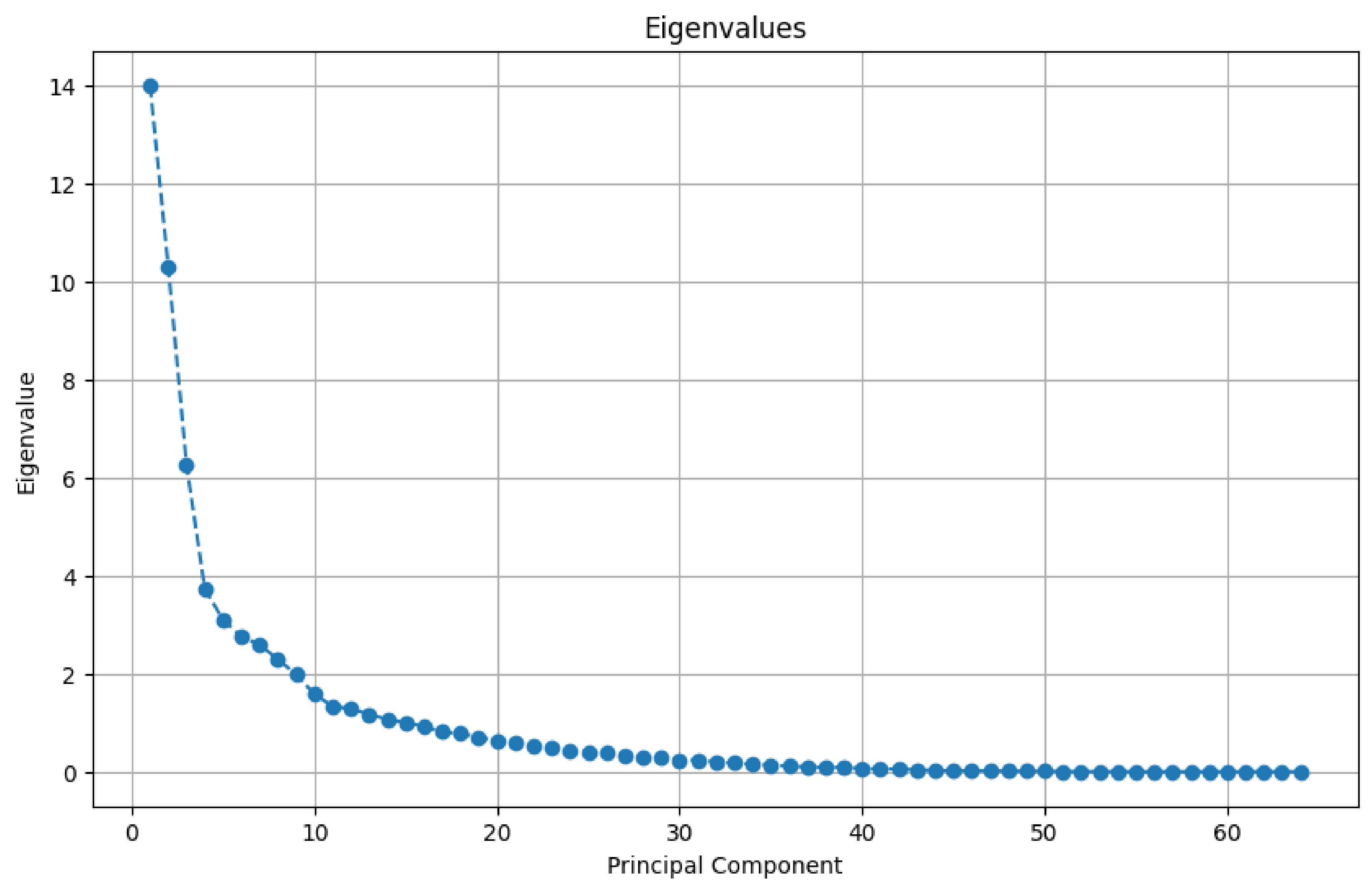
For UDP



For LDAP



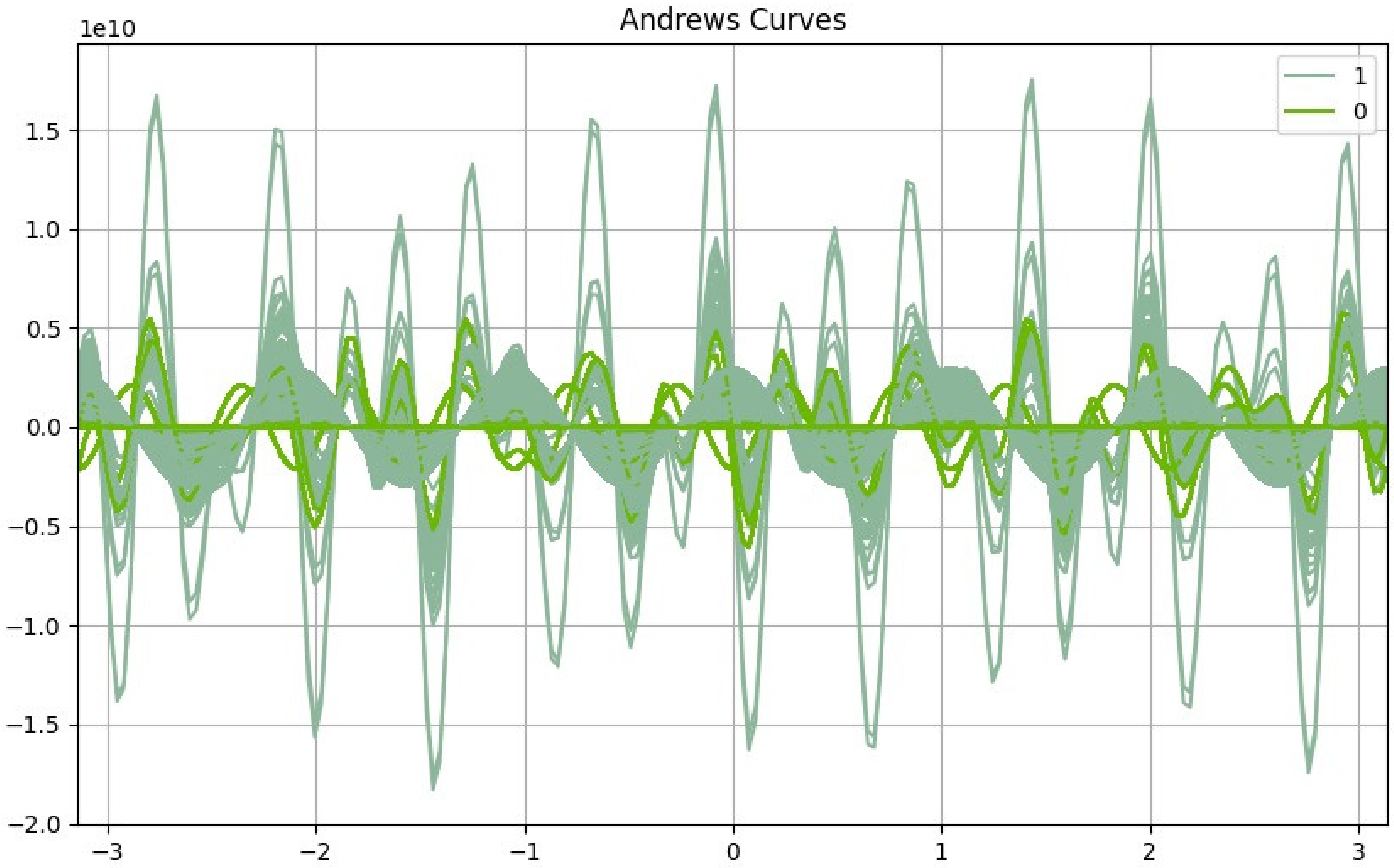
For UDP



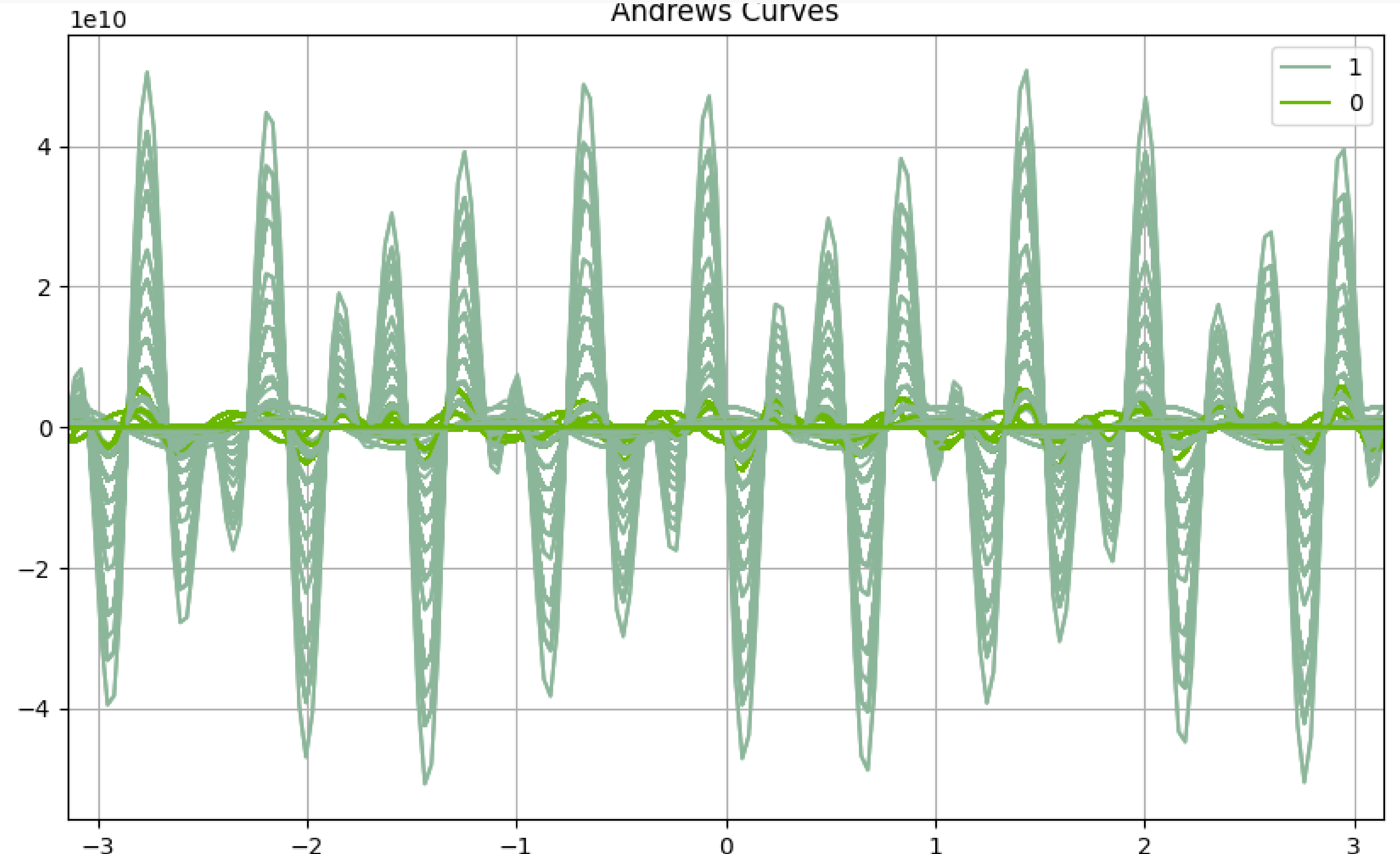
Andrew Curves

Andrews Curves, introduced by D. F. Andrews in 1972, are a visualization method that transforms multivariate data into curves. Each curve represents a row in the dataset, visualizing the values of individual variables through a Fourier series. They serve as a powerful tool for pattern recognition, enabling the intuitive identification of relationships and trends within the data. Andrews Curves are particularly useful for comparing groups, detecting outliers, and simplifying complex multivariate data for exploratory data analysis. Their two-dimensional representation aids in dimensionality reduction without losing essential patterns, and the method can reveal inherent clusters or groups within the dataset. Integration with machine learning allows for combining visual insights with analytical techniques, facilitating interpretation and understanding of complex relationships.

For LDAP



For UDP



Results :

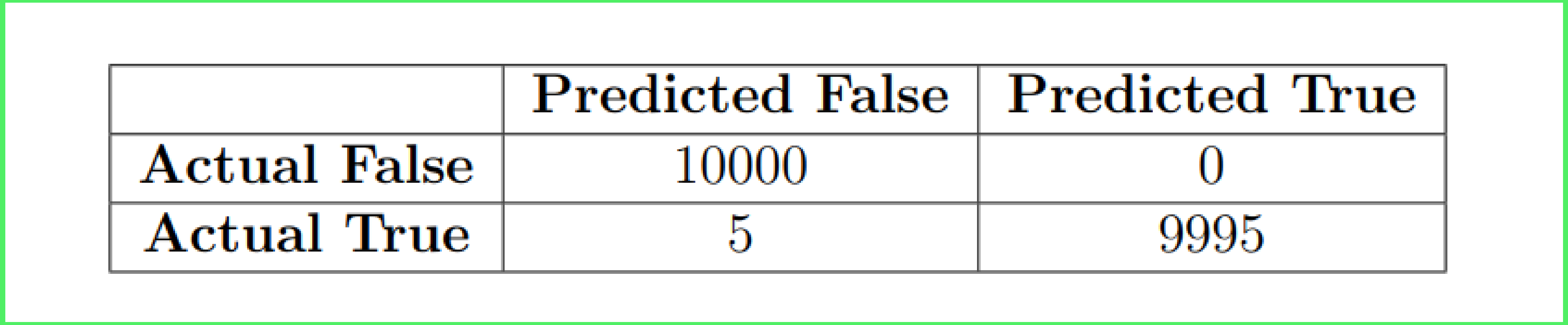
Model Accuracy
0.99972

F1 Score
0.9997

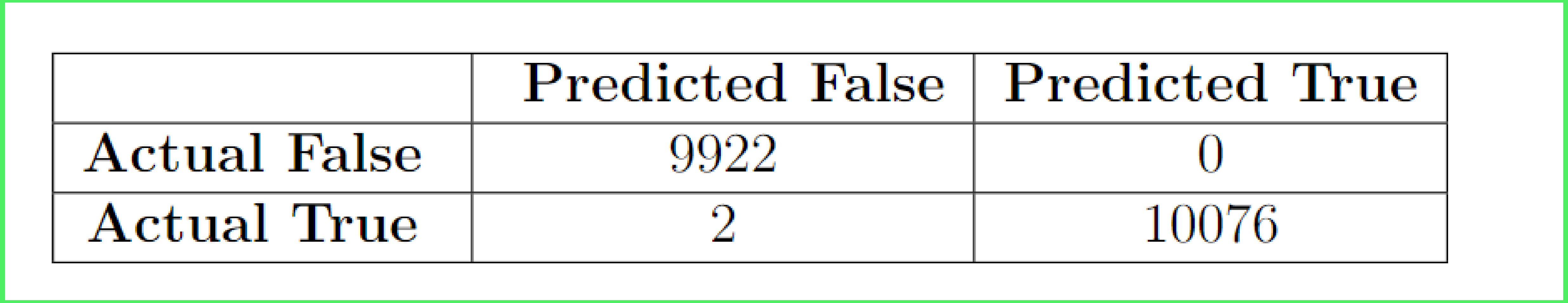
Model Accuracy
0.9999

F1 Score
0.9999

Machine Learning

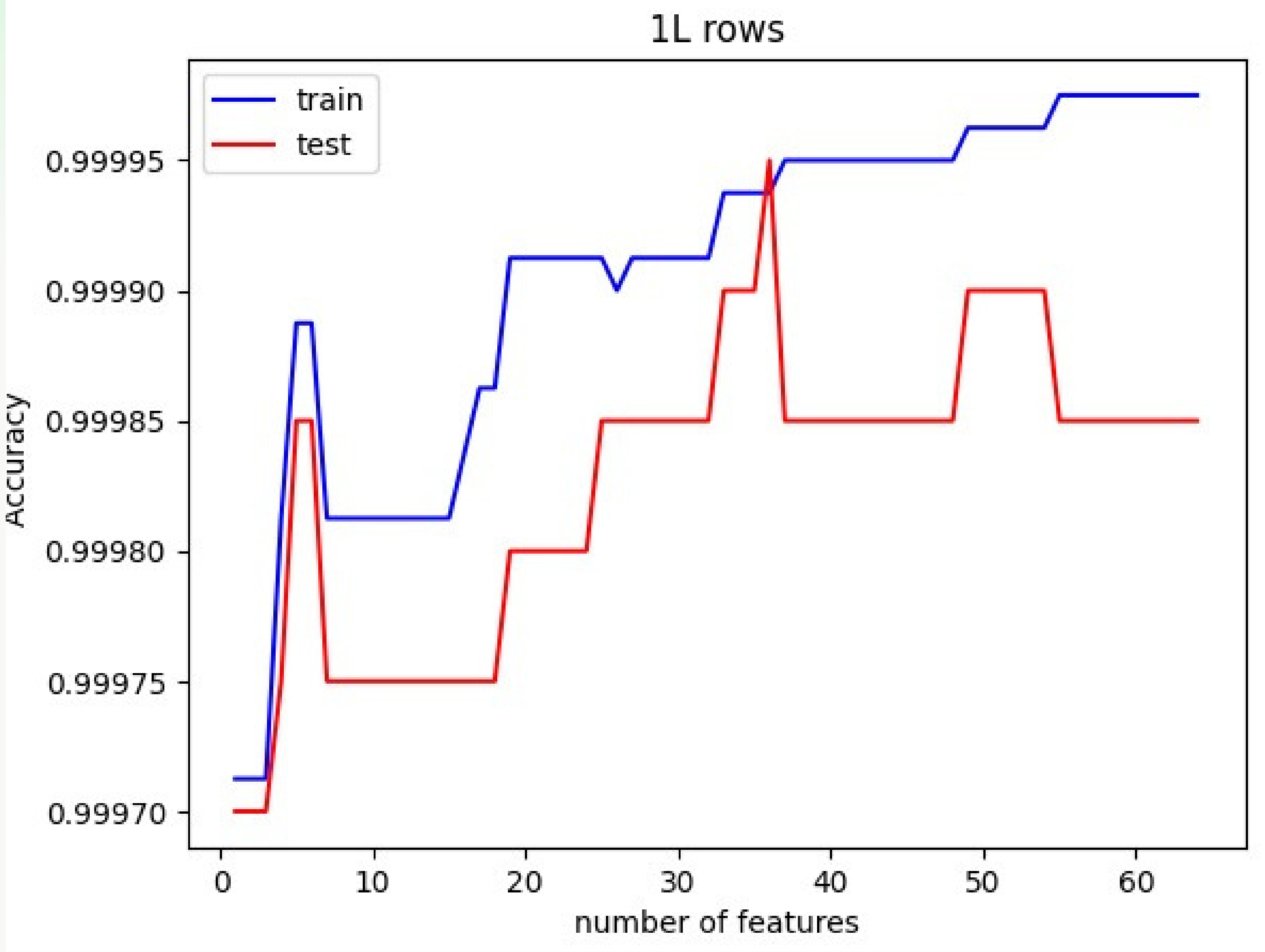


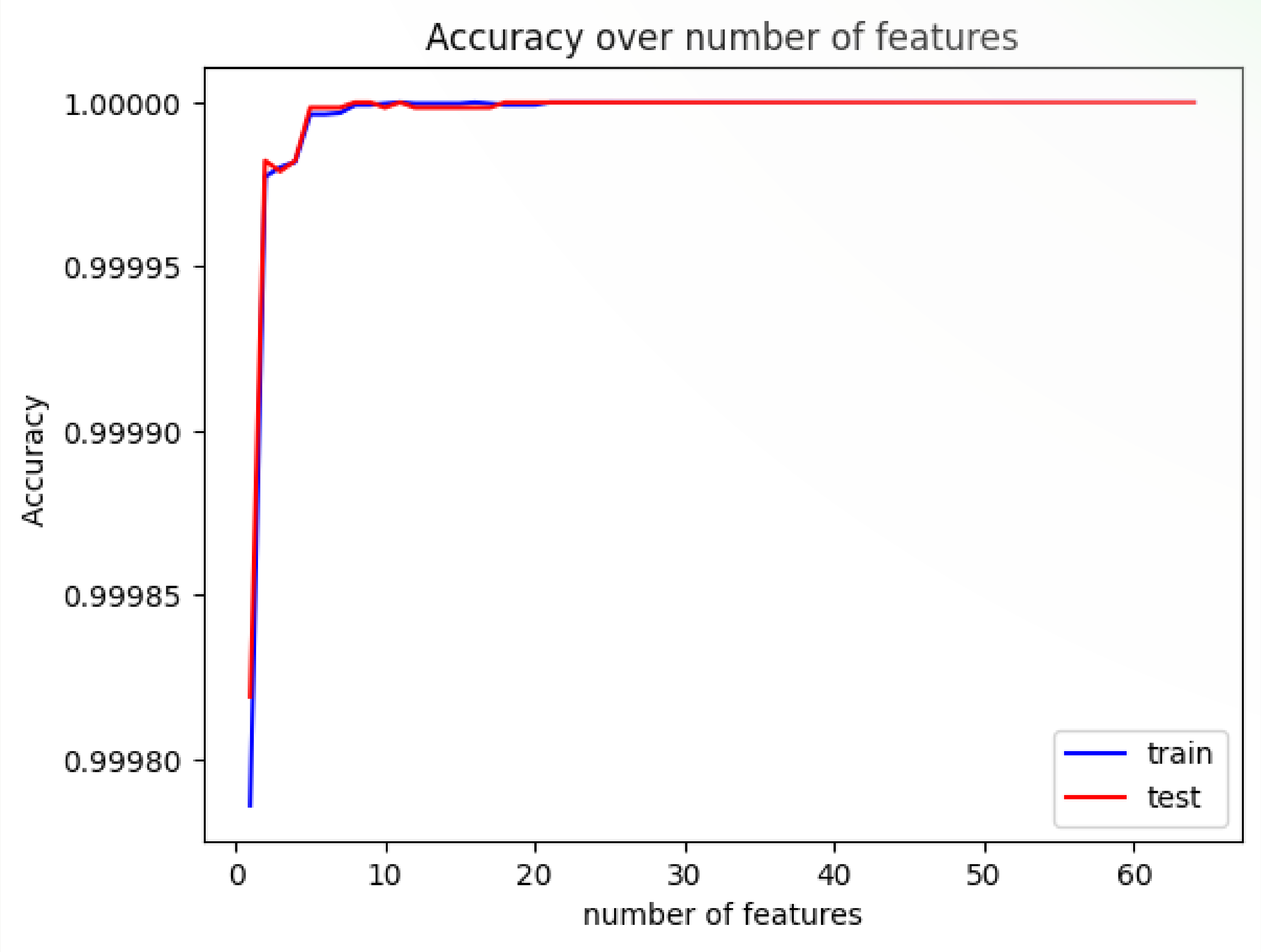
	Predicted False	Predicted True
Actual False	10000	0
Actual True	5	9995



	Predicted False	Predicted True
Actual False	9922	0
Actual True	2	10076

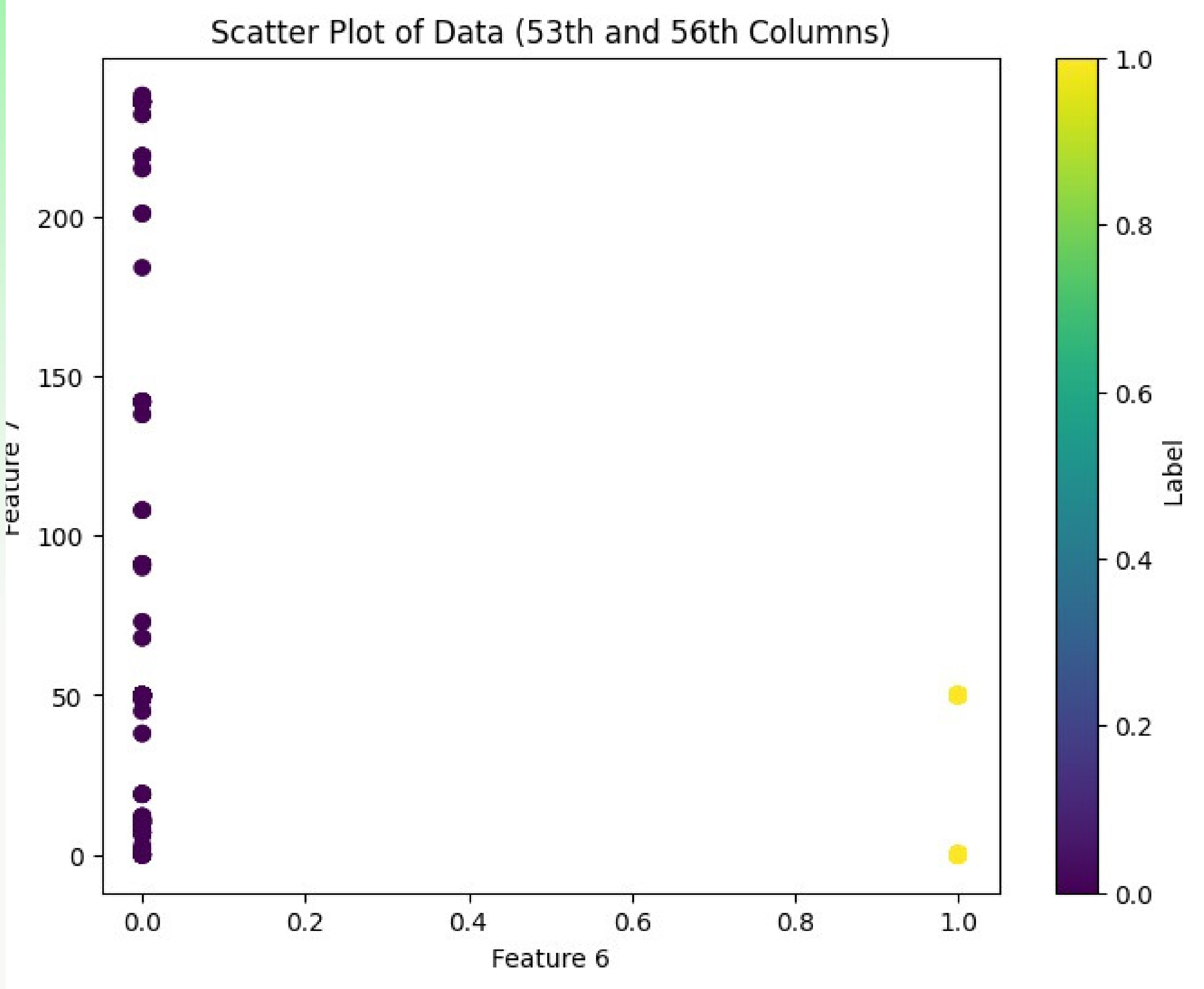
Fluctuation in accuracy between the training and test datasets as the number of features varies

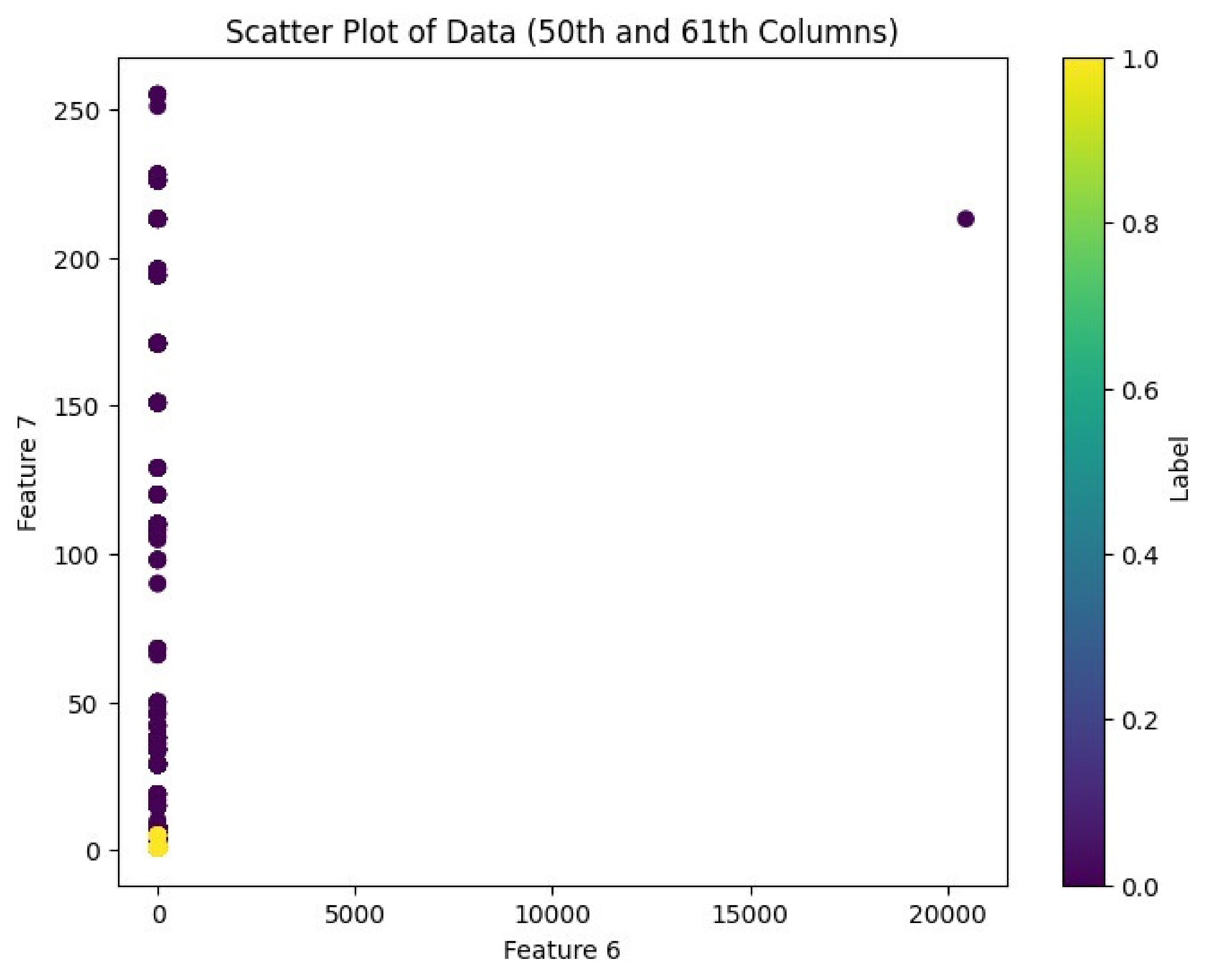




Why so good results?

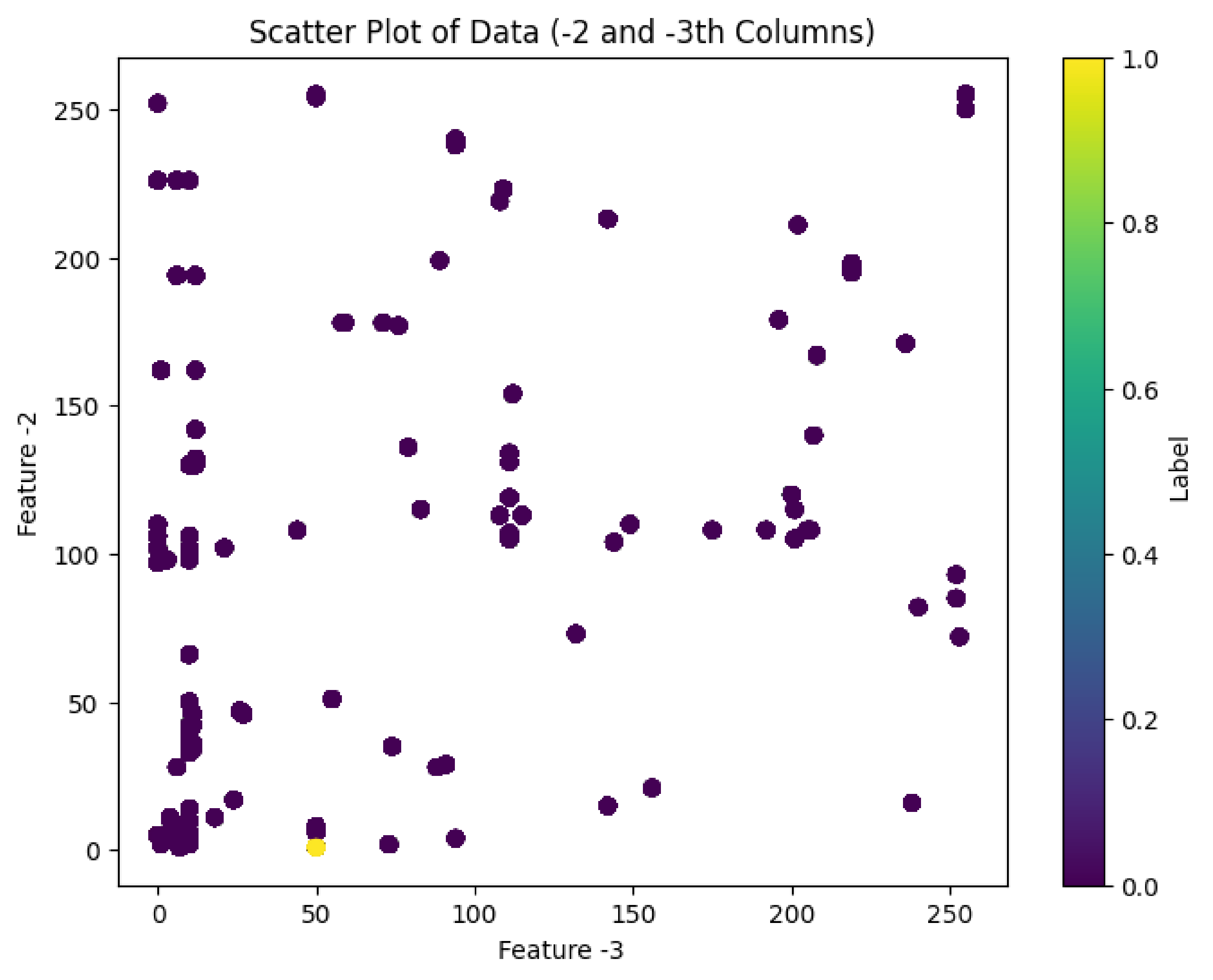
For LDAP

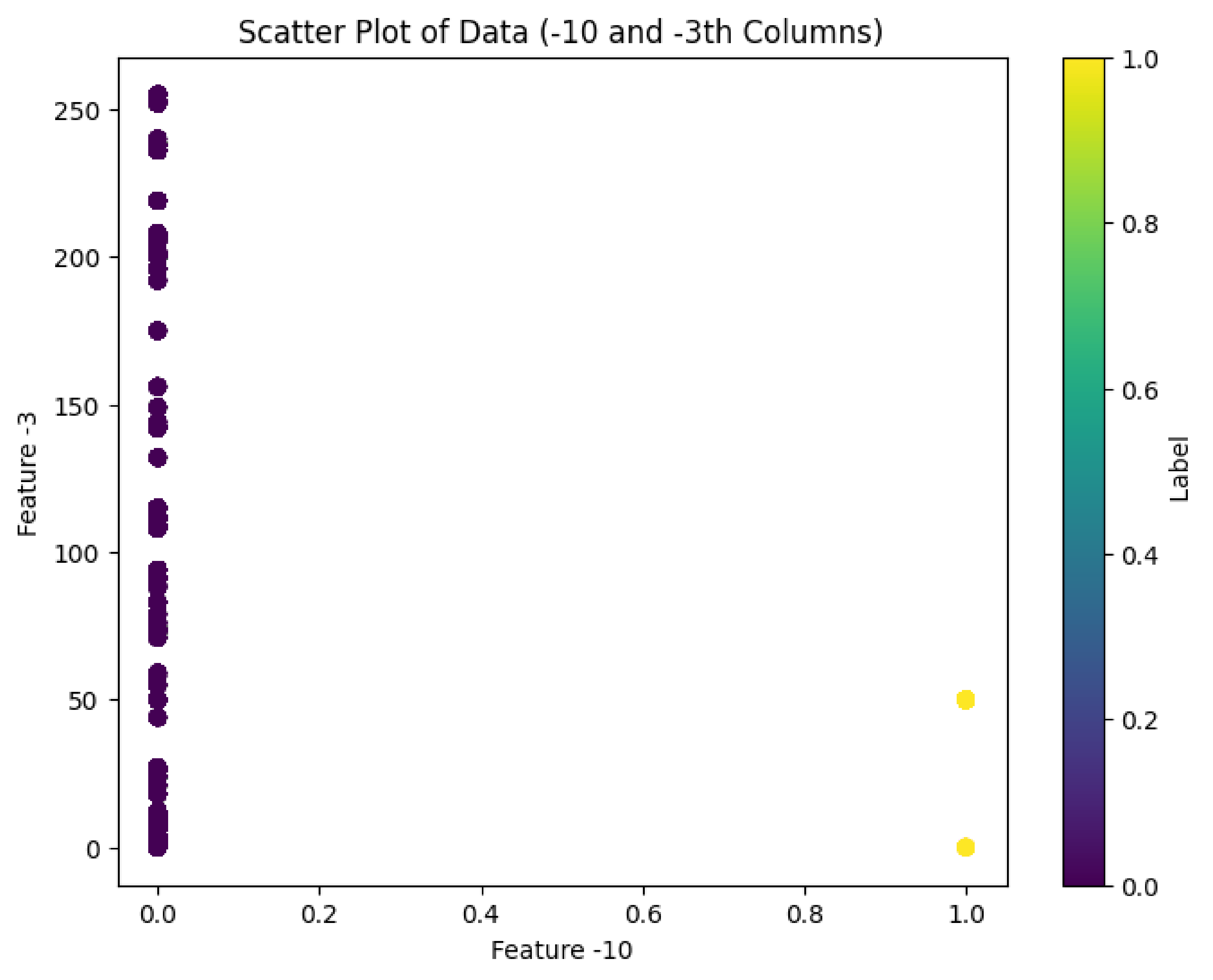




Why so good results?

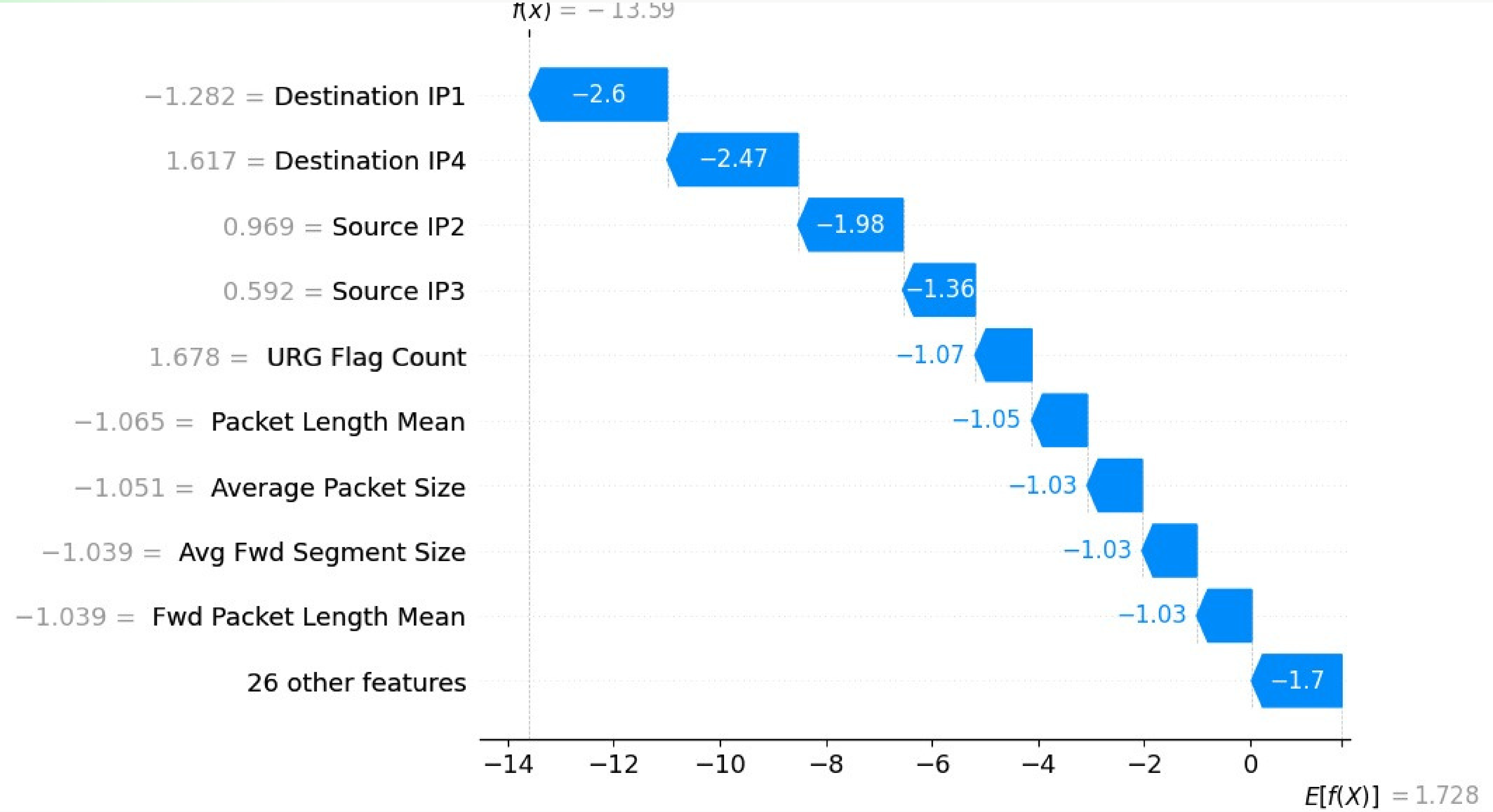
For UDP





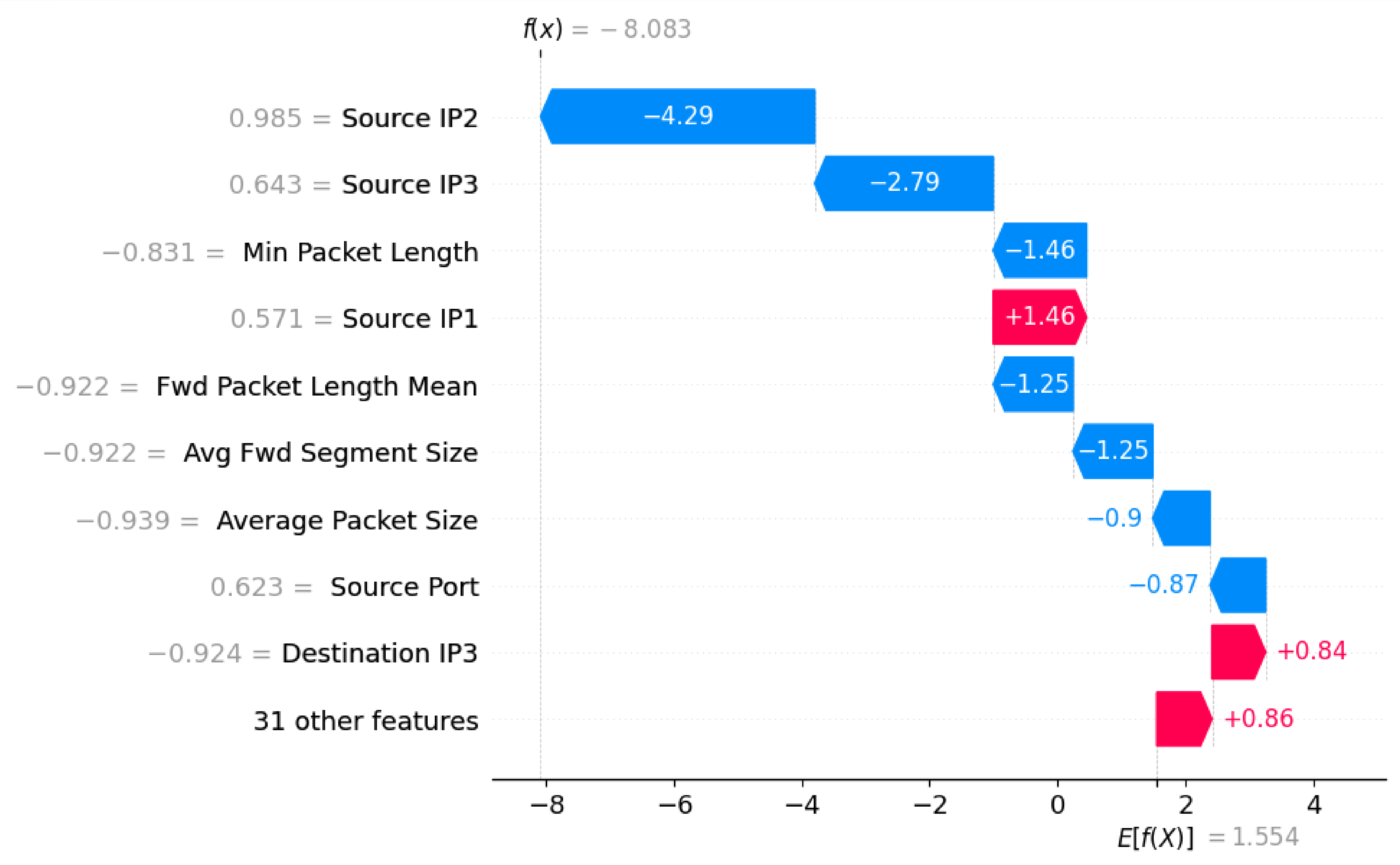
Application of SHAP(Waterfall Plot)

For LDAP

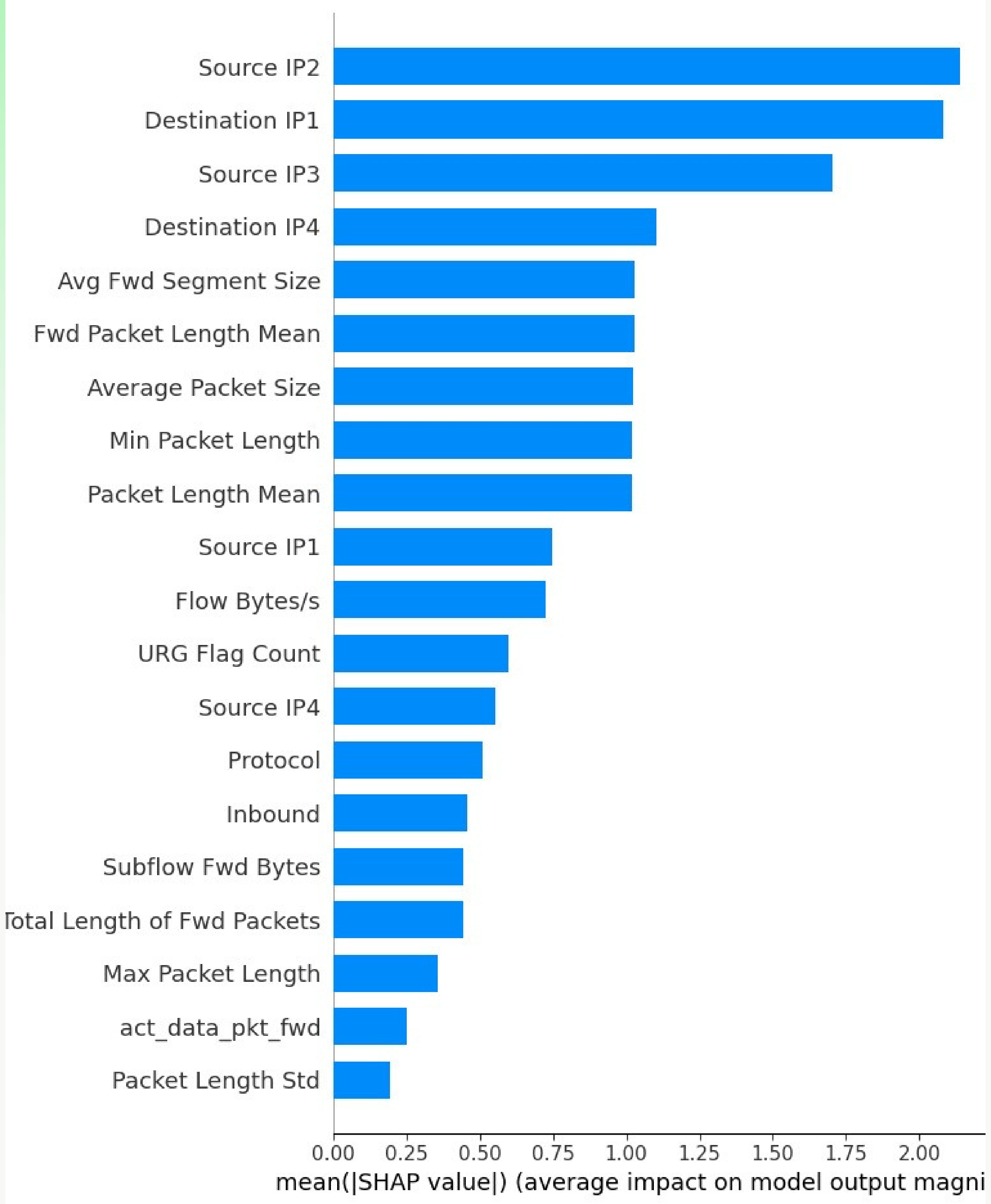


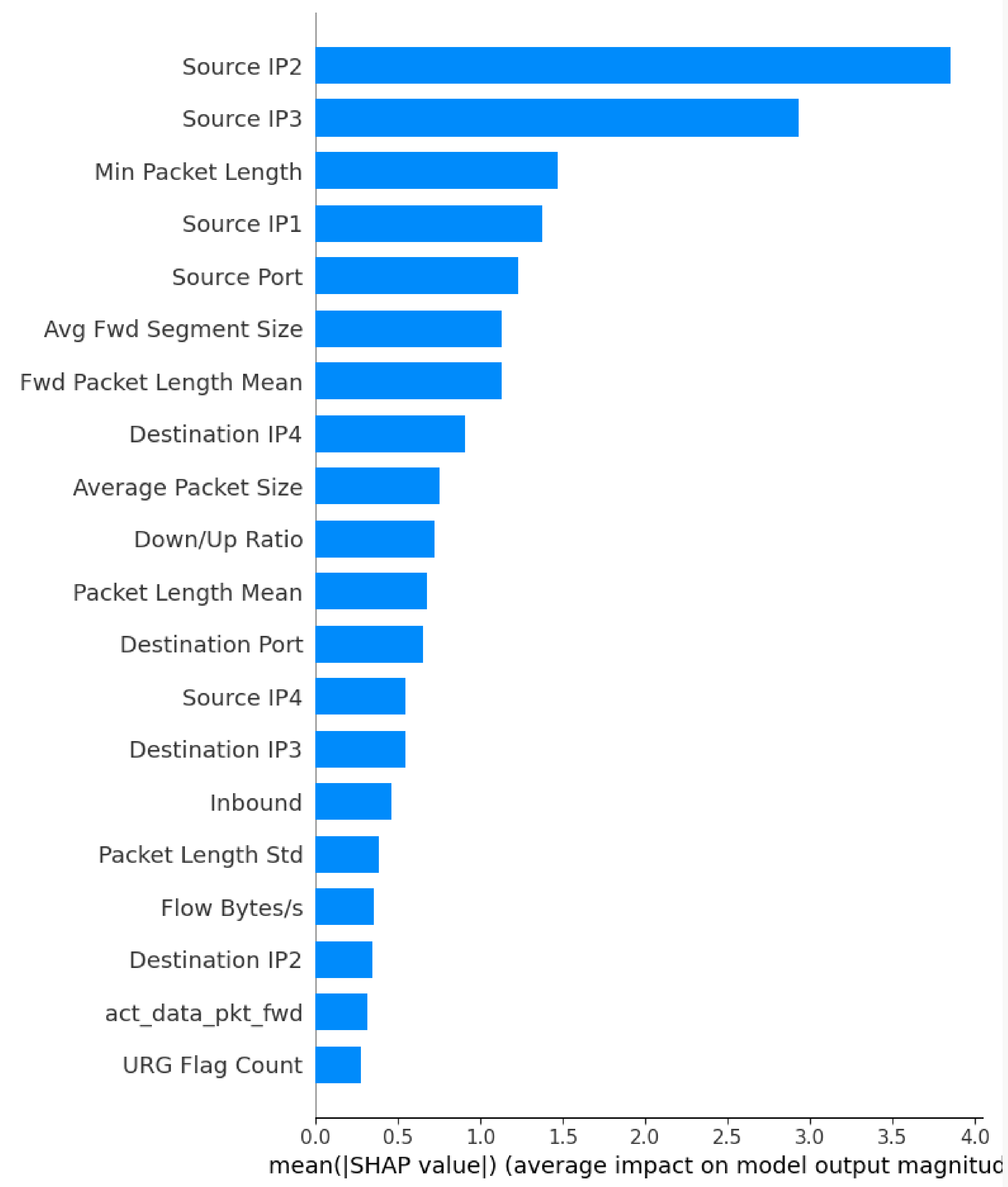
Application of SHAP(Waterfall Plot)

For UDP



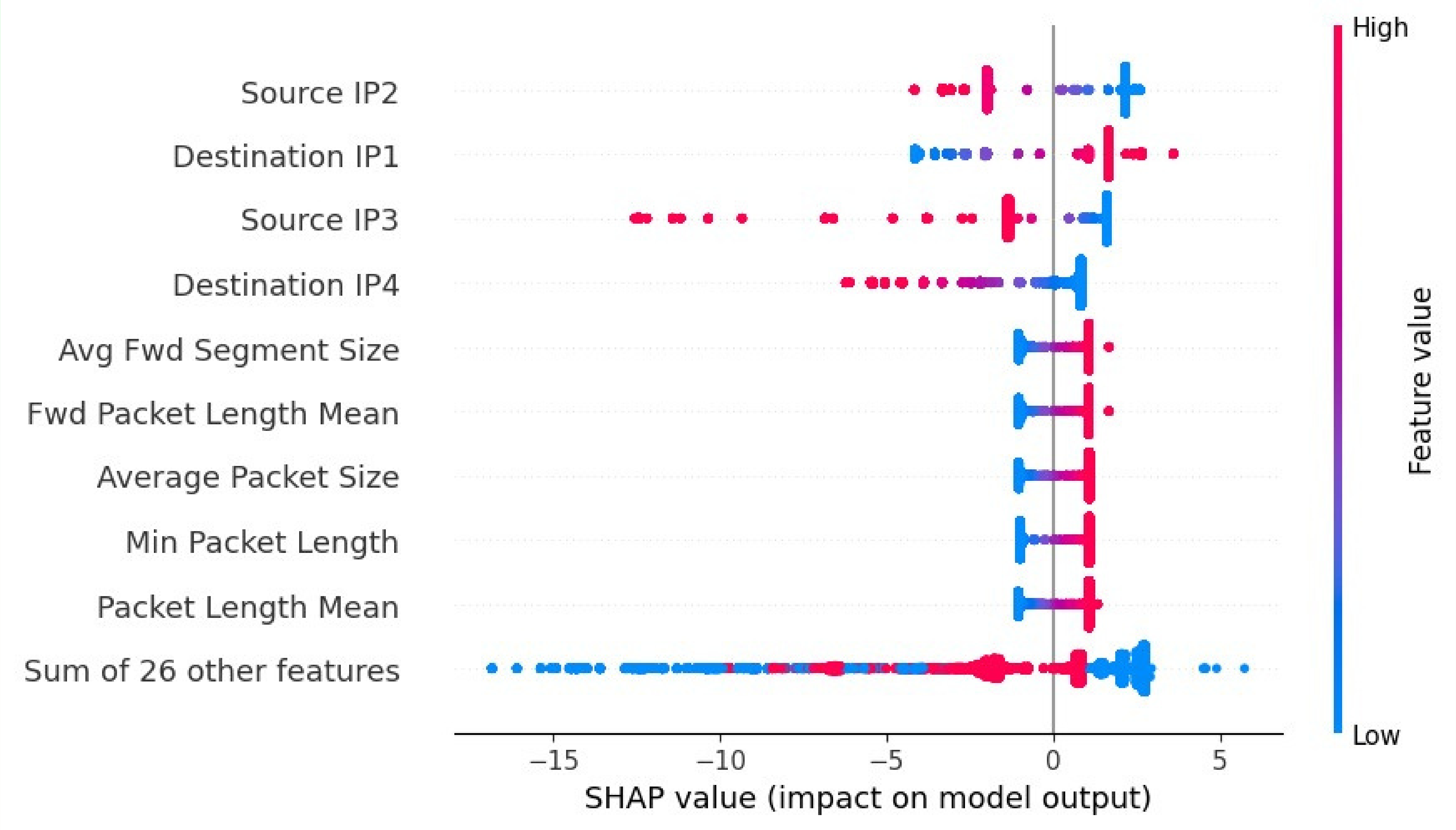
Application of SHAP(Mean SHAP Plot)





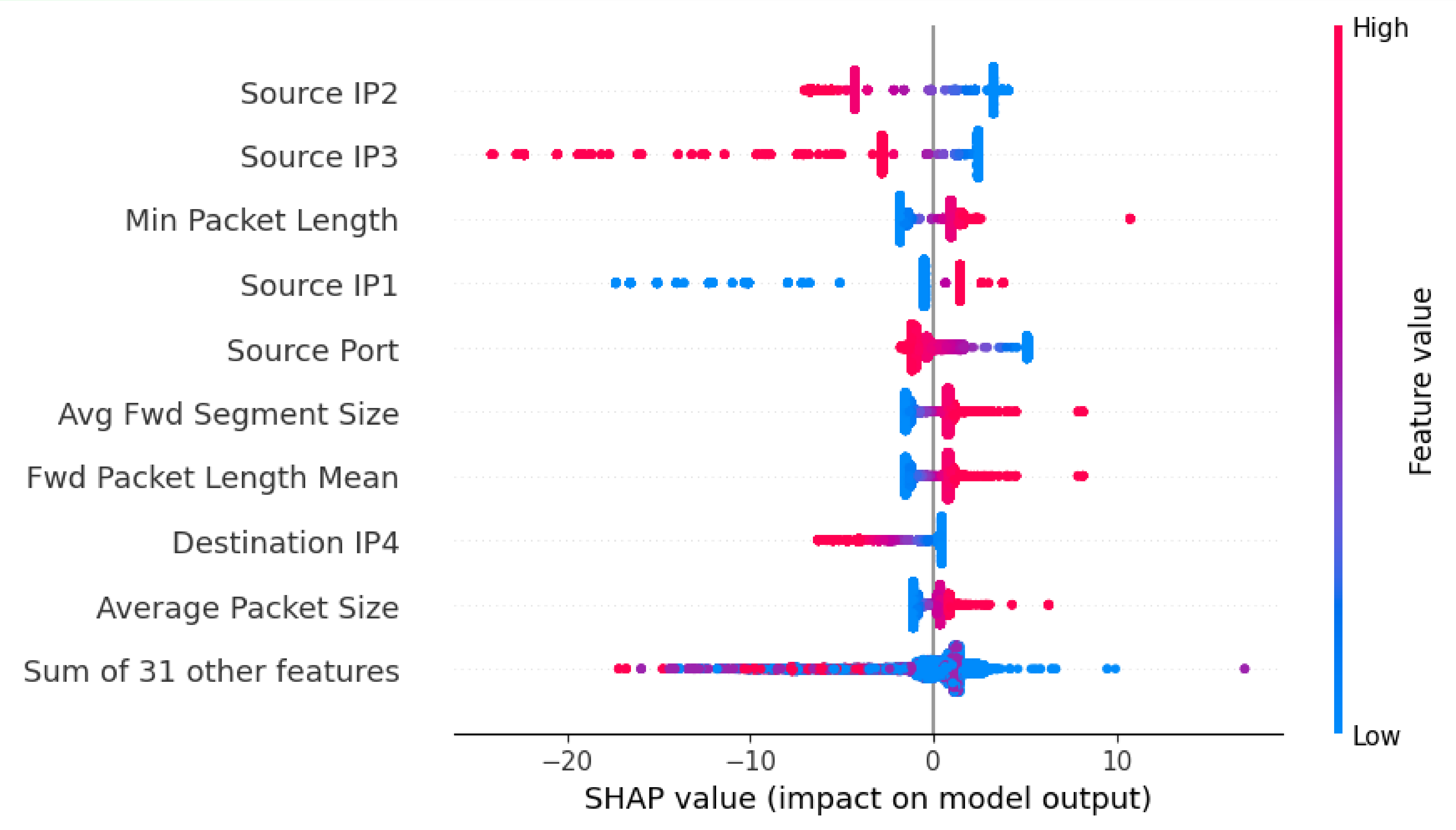
Application of SHAP(Beeswarm Plot)

For LDAP



Application of SHAP(Beeswarm Plot)

For UDP



Conclusion

The alignment of critical features identified not only through the implementation of the Random Forest classifier but also through SelectKBest, correlations, and the top features highlighted by SHAP (SHapley Additive exPlanations) holds paramount significance.This congruence strengthens the reliability and interpretability of the model's decision-making process. The utilization of SHAP in conjunction with black-box models serves as a pivotal approach to shedding light on the intricacies of the model's inner workings.

In essence, the integration of XAI methodologies in black-box models serves as a powerful tool for enhancing the transparency and comprehensibility of machine learning systems.This approach not only aids in model validation but also empowers stakeholders to makeinformed decisions based on a clearer understanding of the model's behavior and the relevance of its identified features.

Thank You