National Technical University of Athens
School of Electrical and Computer Engineering
MSc: Data Science and Machine Learning

DGA-Based Botnet Traffic Detection with Federated Learning Methods and Utilization of eXplainable Artificial Intelligence (XAI) for Classifier Interpretation

Presentation Outline

1. Introduction
2. Theoretical Background
3. Federated Learning
4. Explainable Artificial Intelligence (XAI)
5. Dataset – Data Preprocessing
6. Deep Learning Model
7. Results

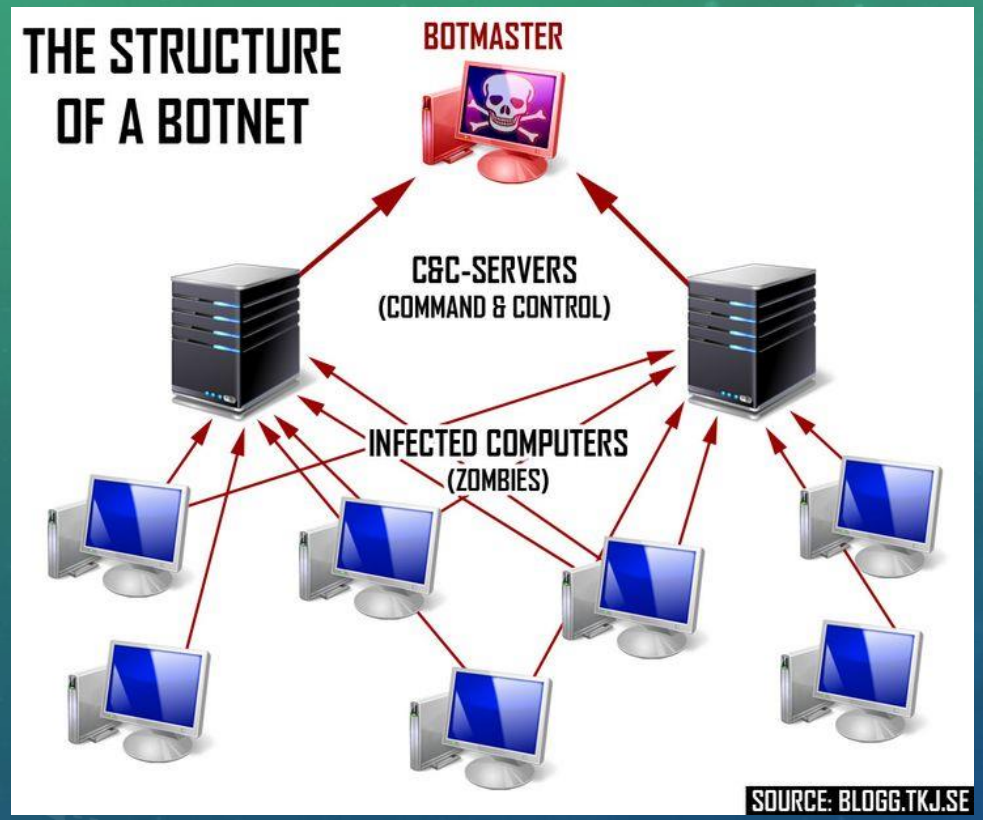
1. Introduction

Subject of Diploma Thesis

- Detection of malicious names generated by Domain Generation Algorithms (DGAs)
- Utilizing deep learning algorithms to distinguish between benign and malicious domain names (DGA names)
- Federated Learning approach to train classifiers
- Using eXplainable Artificial Intelligence (XAI) algorithms to interpret classification decisions

2. Theoretical Background

Structure of a Botnet



- Bots: Devices infected with malware and remotely controlled by attackers
- Botnet: Network of infected devices
- Botmaster: Botnet administrator
- Command & Control (C&C) server: Server through which the attacker instructs each bot to perform malicious actions

Domain Generation Algorithms - DGAs (1/2)

- DGA's:

  - Establish communication between the bots and their C&C servers

  - They receive a seed as input, which is common to bots and C&C servers

  - Generate a pseudorandom set of Domain Names
- C&C servers register some of the generated domain names
- The bots perform DNS Queries until the IP addresses are resolved and connect to the C&C server

Domain Generation Algorithms - DGAs (2/2)

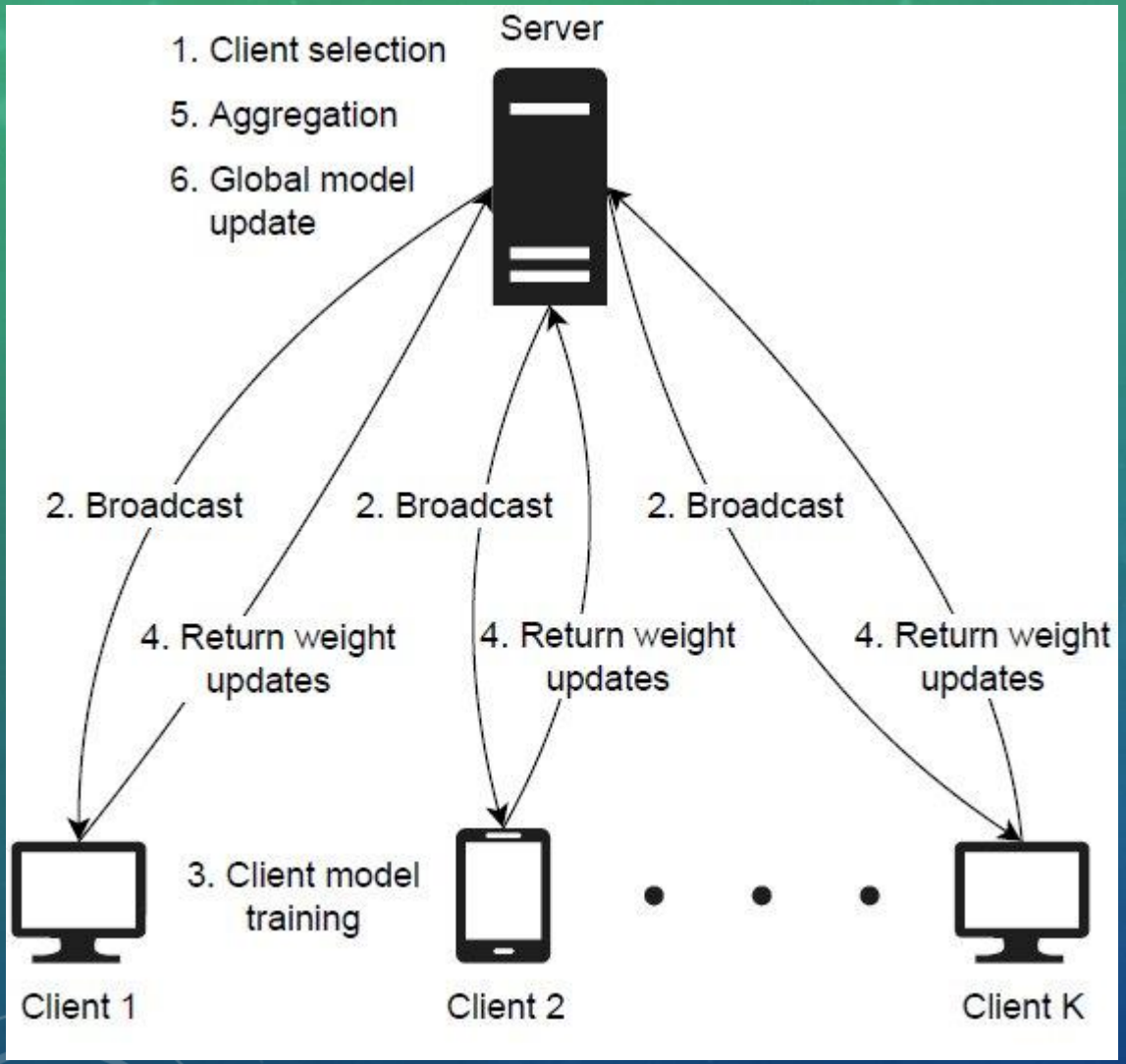
- DGA domain names are detected more efficiently with machine learning algorithms
- Most Popular Machine Learning Algorithms :
  - Decision Trees
  - Deep Learning

3. Federated Learning

Federated Learning (1/2)

- A technique in which multiple clients collaborate to train a common model without exchanging training datasets
- Coordination via a central server
- Advantages:
  - Data privacy protection
  - Develop a common model that is trained using data that the client would not have access to
- Federated Averaging algorithm (FedAvg)

Federated Learning (2/2)



Iterative algorithm where each iteration is called round. First, the server initializes the weights of the model, and then each round consists of the following steps :

1. Client selection
2. Broadcast
3. Client model training
4. Return weights
5. Aggregation
6. Global model update

4. Explainable Artificial Intelligence

eXplainable Artificial Intelligence - XAI (1/3)

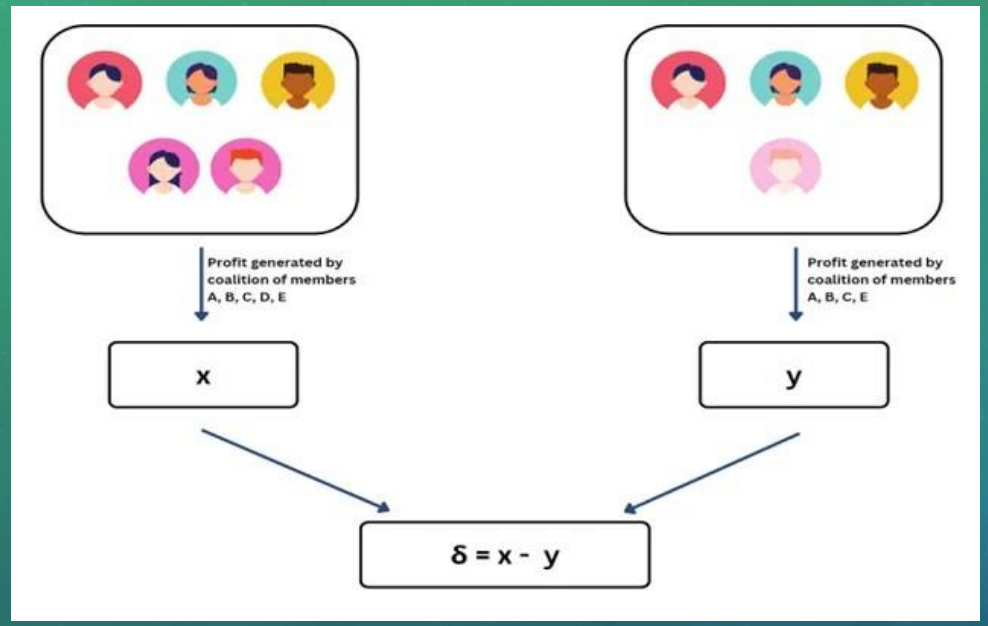
- Understanding the decisions of the deep learning model
- Contribution of features
- Local explainability: understanding and interpreting the model's decision for a particular instance
- Global explainability: total contribution to several instances
- XAI Algorithms
  - Permutation Feature Importance
  - LIME (Local Interpretable Model Agnostic Explanation)
  - SHAP (Shapley Additive exPlanations)

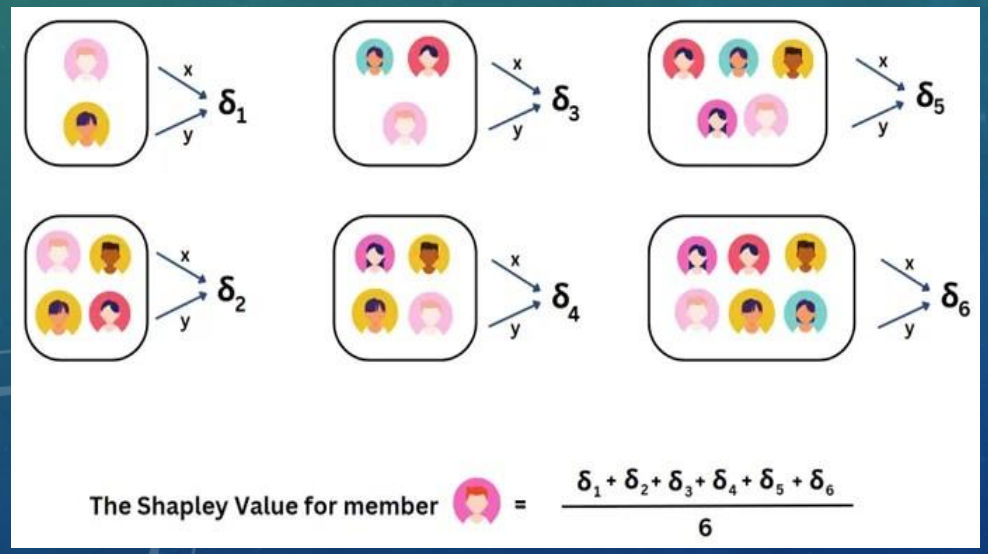
eXplainable Artificial Intelligence - XAI (2/3)

SHAP (SHapley Additive exPlanations)

- Game Theory– Coalition games
- Post-hoc: algorithms applied after the model training
- Model-agnostic: application on any model
- Contribution of each feature is given via Shapley value

eXplainable Artificial Intelligence - XAI (3/3)





- 5 players in a project to make a profit
- Calculating each player's contribution (Shapley value) to profit making
- The difference between the profit generated when the player is present and when the player is absent
- Calculation in each subgroup (coalition) the player belongs to and the final Shapley value is their average

5. Dataset – Data Preprocessing

Dataset

- Legitimate names: subset of Alexa Top 1 Million
- Malicious names:
  - Selected DGA families: pykspa, dircrypt, ramnit, qakbot, kraken, corebot, locky, banjori, cryptolocker, simda, ramdo

Data Preprocessing

Extraction of statistical features from domain names indicating linguistic and statistical properties :

Feature Name	Description
Domain Name Length	Number of characters
Consonants Freq	Number of consonants
Max Consonants Seq Length	Length of the maximum consonants sequence
Min Consonants Seq Length	Length of the minimum consonants sequence
Vowels Freq	Number of vowels
Max Vowels Seq Length	Length of the maximum vowels sequence
Min Vowels Seq Length	Length of the minimum vowels sequence
Digits Freq	Number of digits
Max Digits Seq Length	Length of the maximum digits sequence
Min Digits Seq Length	Length of the minimum digits sequence
Hyphen Freq	Number of hyphens
Max Hyphen Seq Length	Length of the maximum hyphens sequence
Min Hyphen Seq Length	Length of the minimum hyphens sequence
Characters Frequency	Number of a character
Max Letter Seq Length	Length of the maximum letters sequence
Min Letter Seq Length	Length of the minimum letters sequence

6. Deep Learning Model

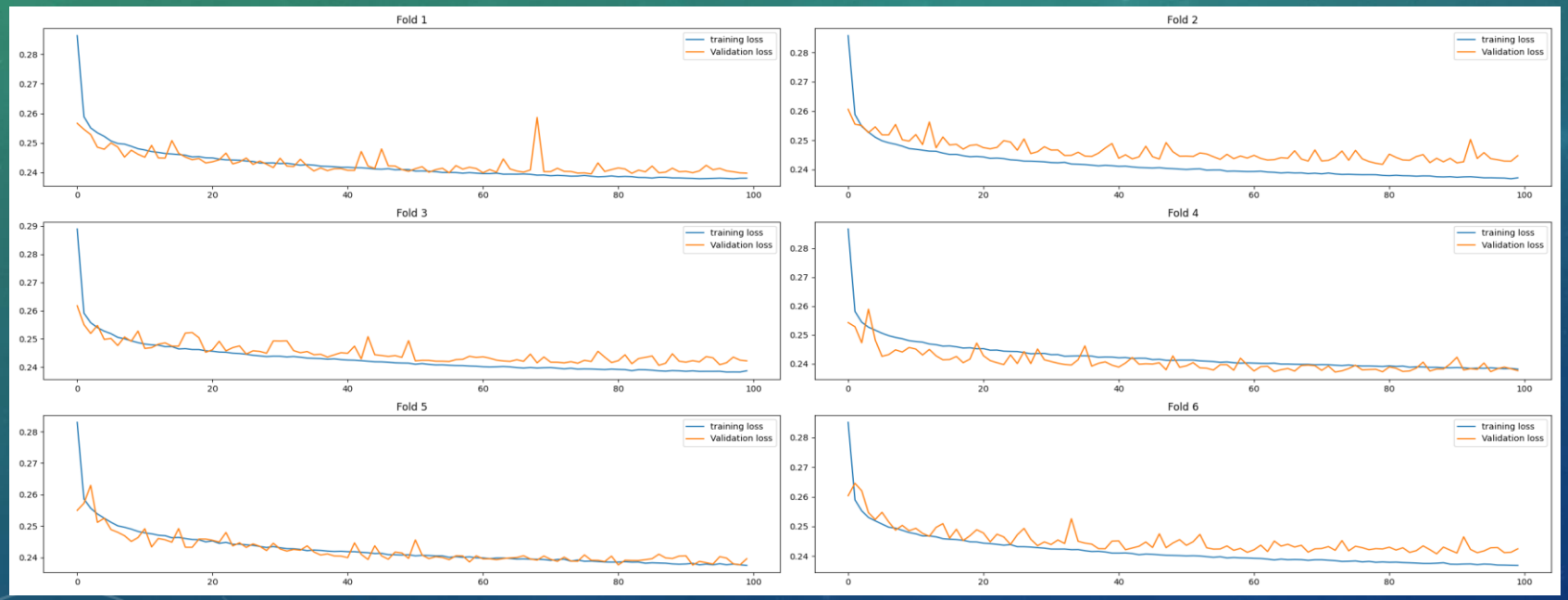
Deep Learning Model

- Multi Layer Perceptron (MLP) determined after hyperparameter tuning (KerasTuner):
  - Input Layer: 16 neurons
  - Hidden layer 1: 128 neurons - ReLU activation function
  - Hidden layer 2: 32 neurons - ReLU activation function
  - Hidden layer 3: 16 neurons - ReLU activation function
  - Output layer: 1 neurons - sigmoid activation function
- Loss function: binary cross entropy
- Optimizer: Adam

7. Results

Non-Federated Learning

- Train model using the total training dataset
- Stratified k-fold cross validation (k=6)
- Min-Max normalization



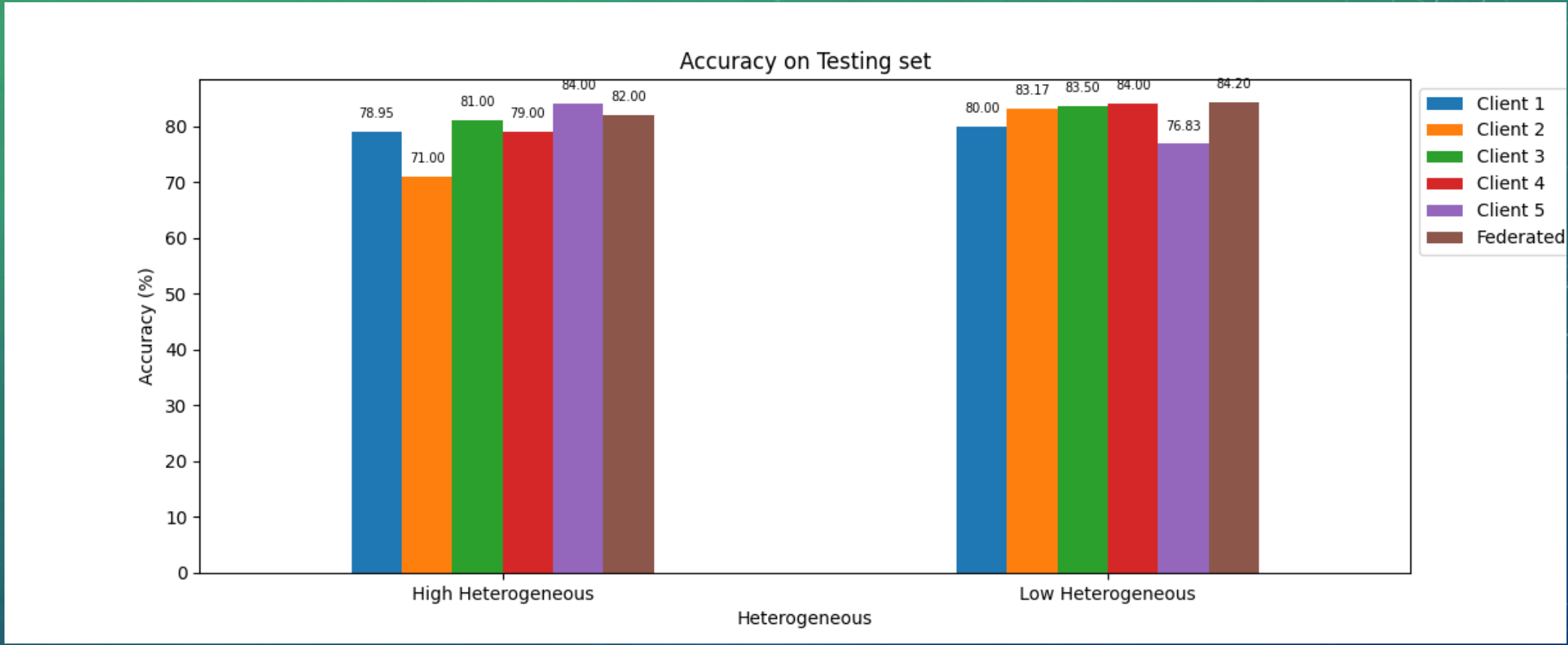
Metric	Average Value
Accuracy	85
Precision	84.5
Recall	70.6
F1-Score	76.8

Federated Learning (1/2)

- Examination of the performance of models when trained with local datasets in a Federated Learning environment (5 clients - 1 aggregated server)
- Federated Averaging
- Impact of data heterogeneity :
  - High heterogeneous
  - Low heterogeneous

Distribution	High Heterogeneous					Low Heterogeneous				
Client	Client 1	Client 2	Client 3	Client 4	Client 5	Client 1	Client 2	Client 3	Client 4	Client 5
Ramnit	100%					20%	20%	20%	20%	20%
Kraken	100%					20%	20%	20%	20%	20%
Simda		100%				20%	20%	20%	20%	20%
Banjori		100%				20%	20%	20%	20%	20%
Pykspa			100%			20%	20%	20%	20%	20%
Ramdo			100%			20%	20%	20%	20%	20%
Qakbot				100%		20%	20%	20%	20%	20%
Cryptolocker				100%		20%	20%	20%	20%	20%
DirCrypt					100%	20%	20%	20%	20%	20%
Corebot					100%	20%	20%	20%	20%	20%
Locky					100%	20%	20%	20%	20%	20%

Federated Learning (2/2)

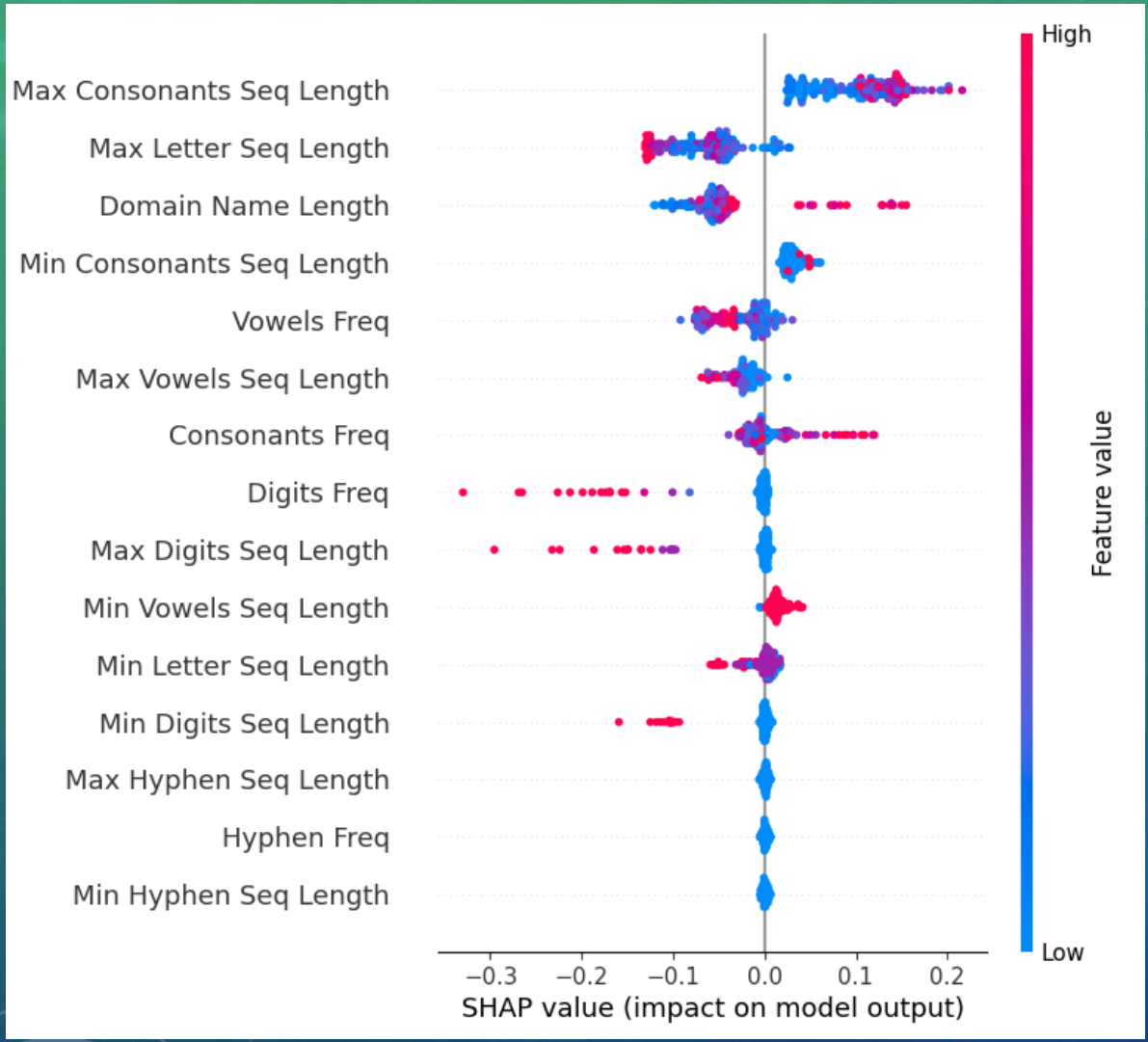


- The federated model achieves better performance than the individual models in the majority of cases
- Greater improvement in the accuracy of individual models for the case of high heterogeneity

XAI

- Exploring the broader interpretability of the models within the SHAP Framework
- Explainability Background Instances (XBI's)
  - Calculate Shapley values
  - Apply K-Means to train dataset (K=50)
- Explainability Test Instances (XTI's)
  - Extraction of interpretations
  - 200 random instances from test set
- Each client calculates its own Shapley values
- Averaging the Shapley values of the client's models to obtain interpretations for the aggregated model

Federated Model Results (1/7)



Summary Plot:

1) Features are sorted according to their contribution

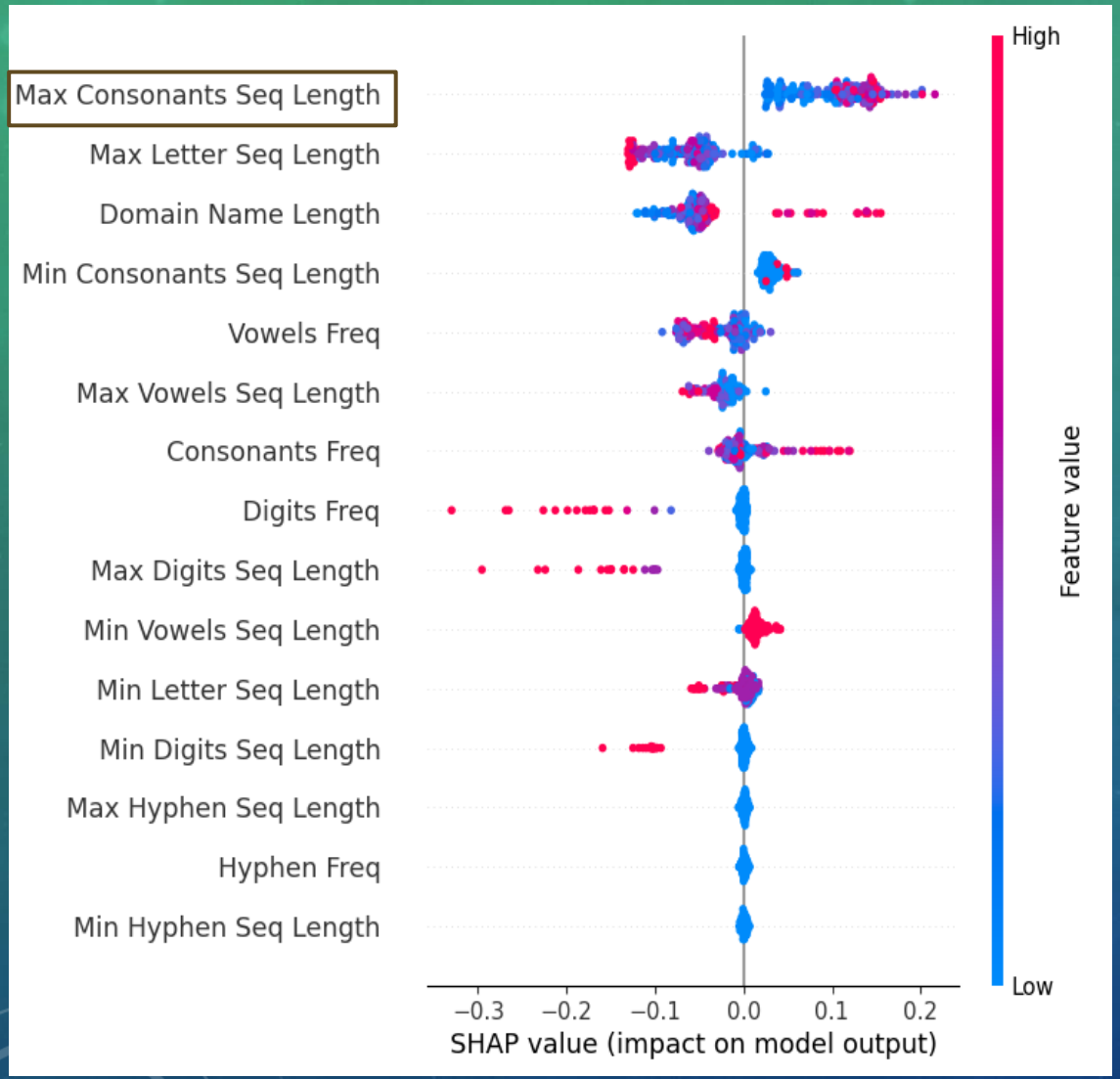
2) Each XTI is represented as a dot

3) The colour represents the value of the features from low (blue) to high (red)

5 most significant features:

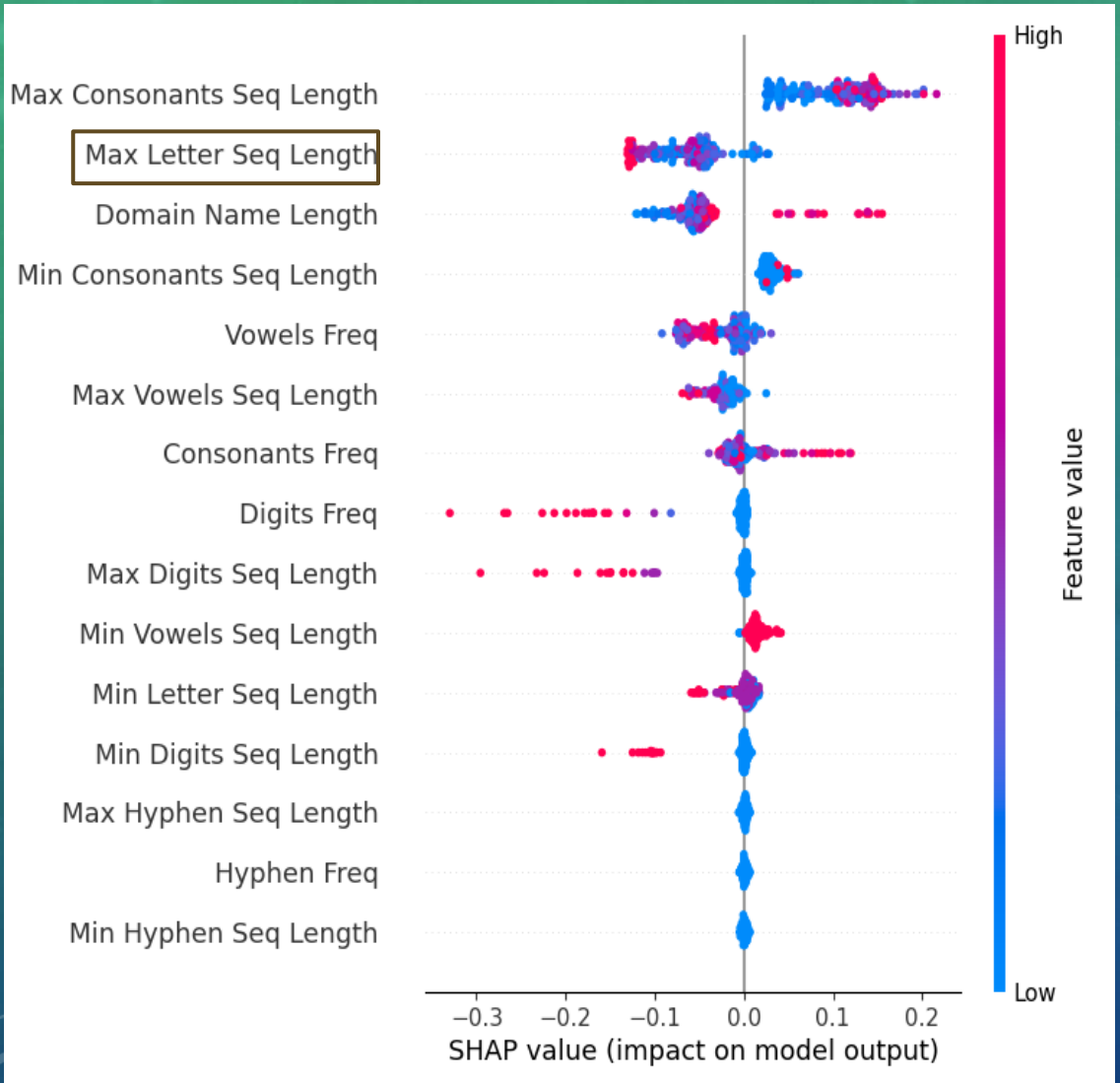
- Max Consonants Seq Length
- Max Letter Seq Length
- Domain Name Length
- Min Consonants Seq Length
- Vowels Frequency

Federated Model Results (2/7)



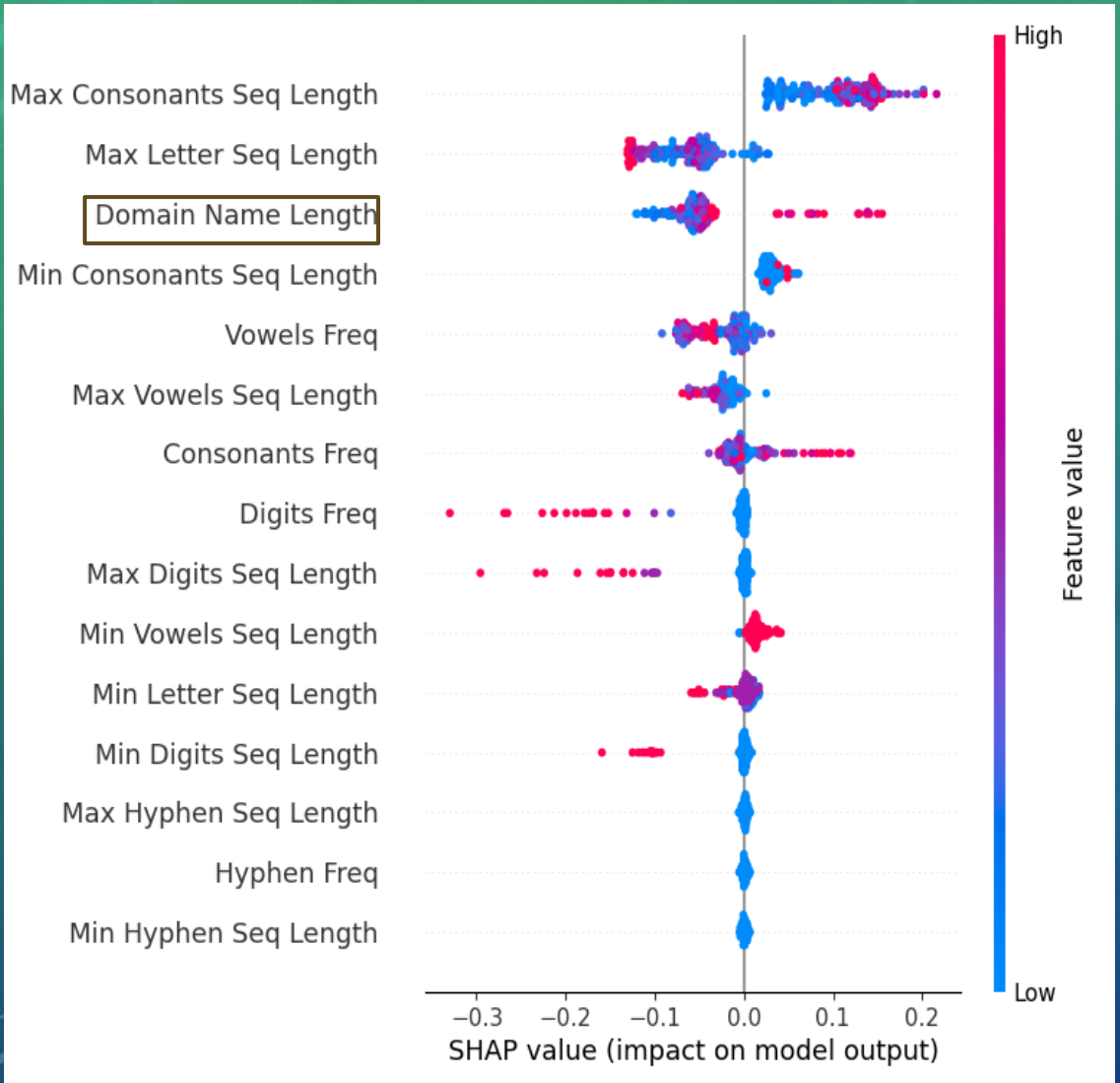
- Max Consonants Seq Length: length of the maximum consonants sequence
- Most important feature
- Influences the classification to the DGA class

Federated Model Results (3/7)



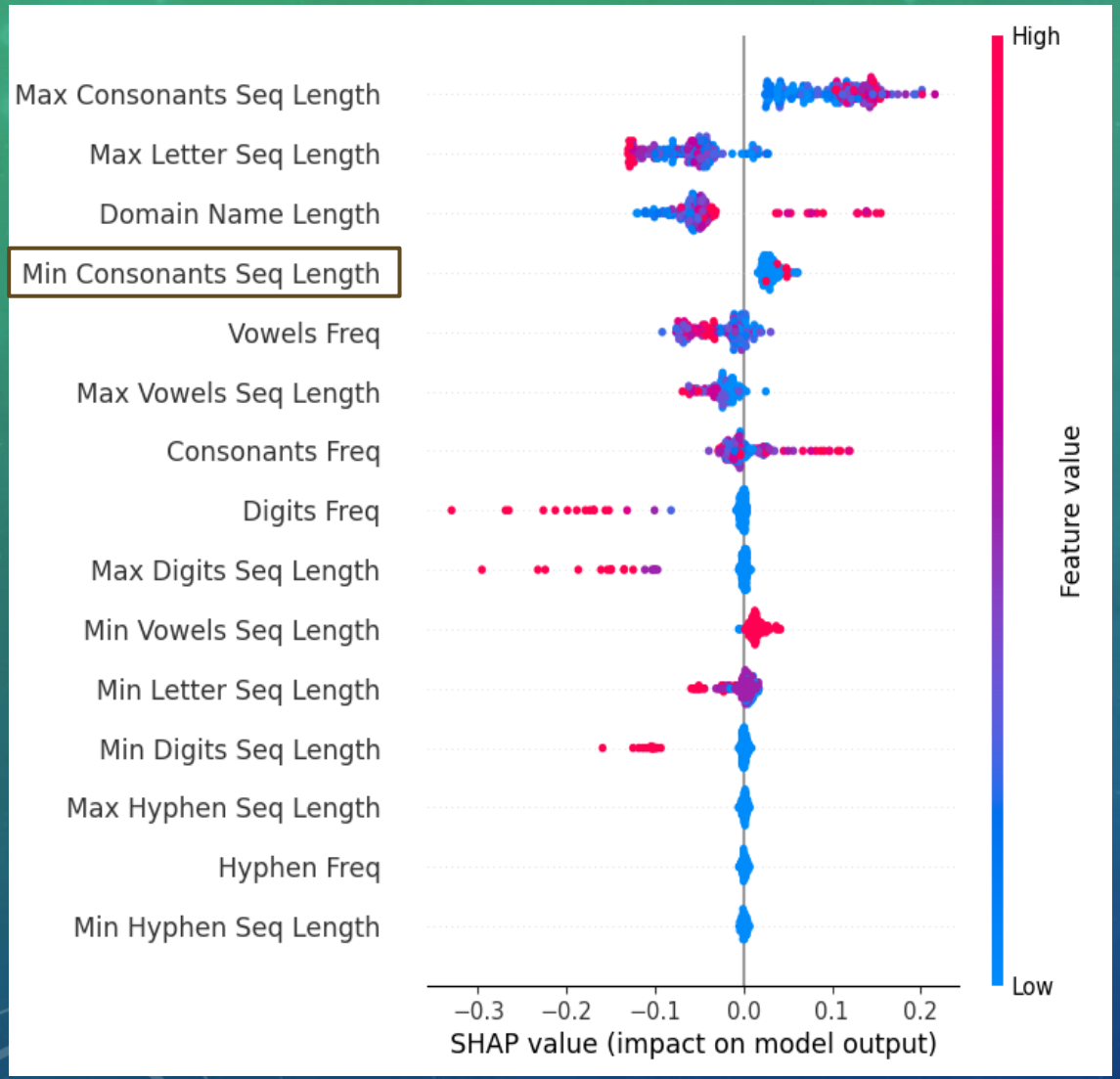
- Max Letter Seq Length : length of maximum letter sequence
- Influences the classification to the legitimate class
- More consecutive letters mean more legitimate domain names
- Low values do not affect the decision of the model

Federated Model Results (4/7)



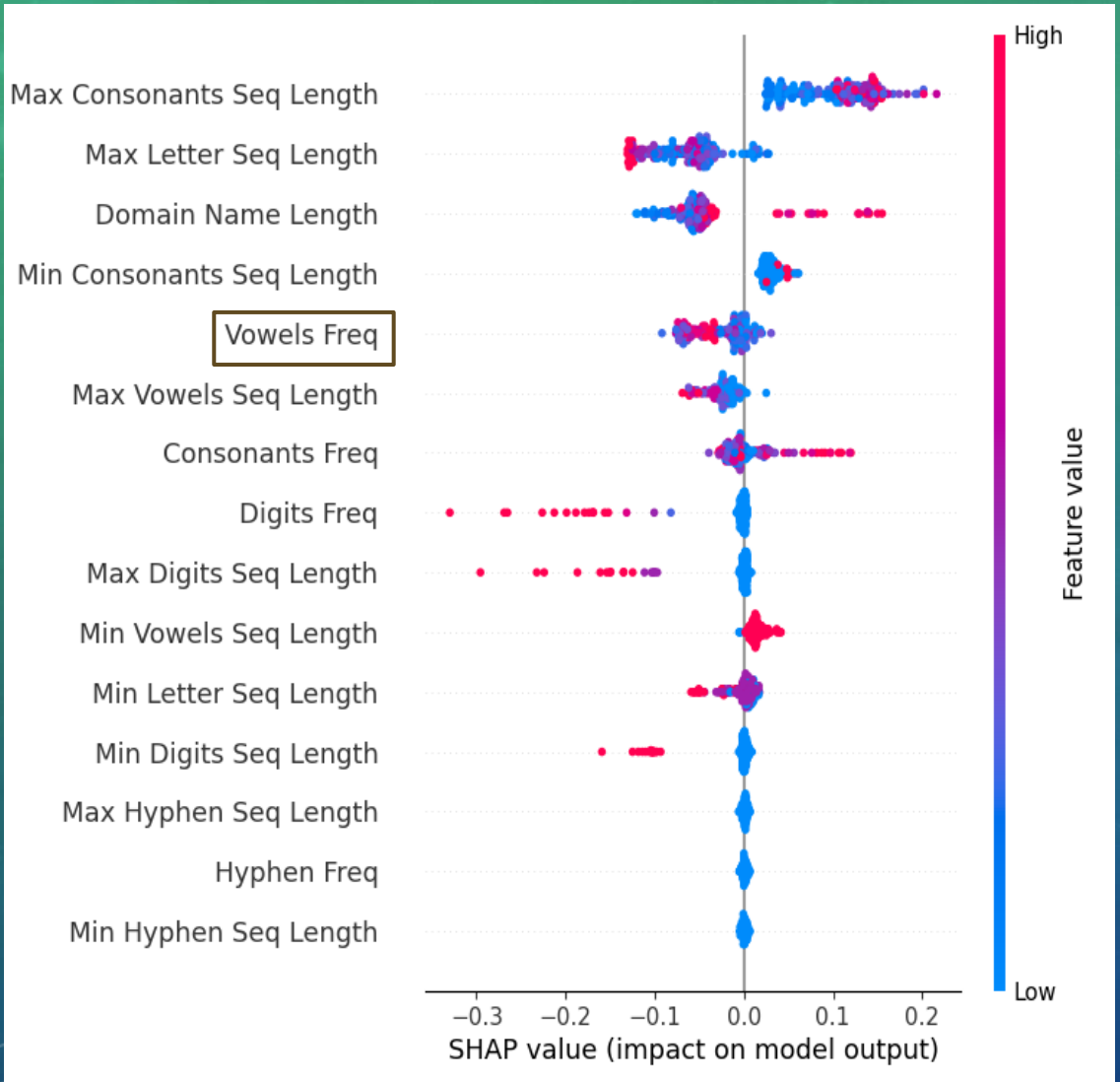
- Domain Name Length : total name length
- High values lead to the DGA class
- Low values lead to the ligitimate class

Federated Model Results (5/7)



- Min Consonants Seq Length : length of the minimum sequence of consonants
- Influences the classification to the DGA class

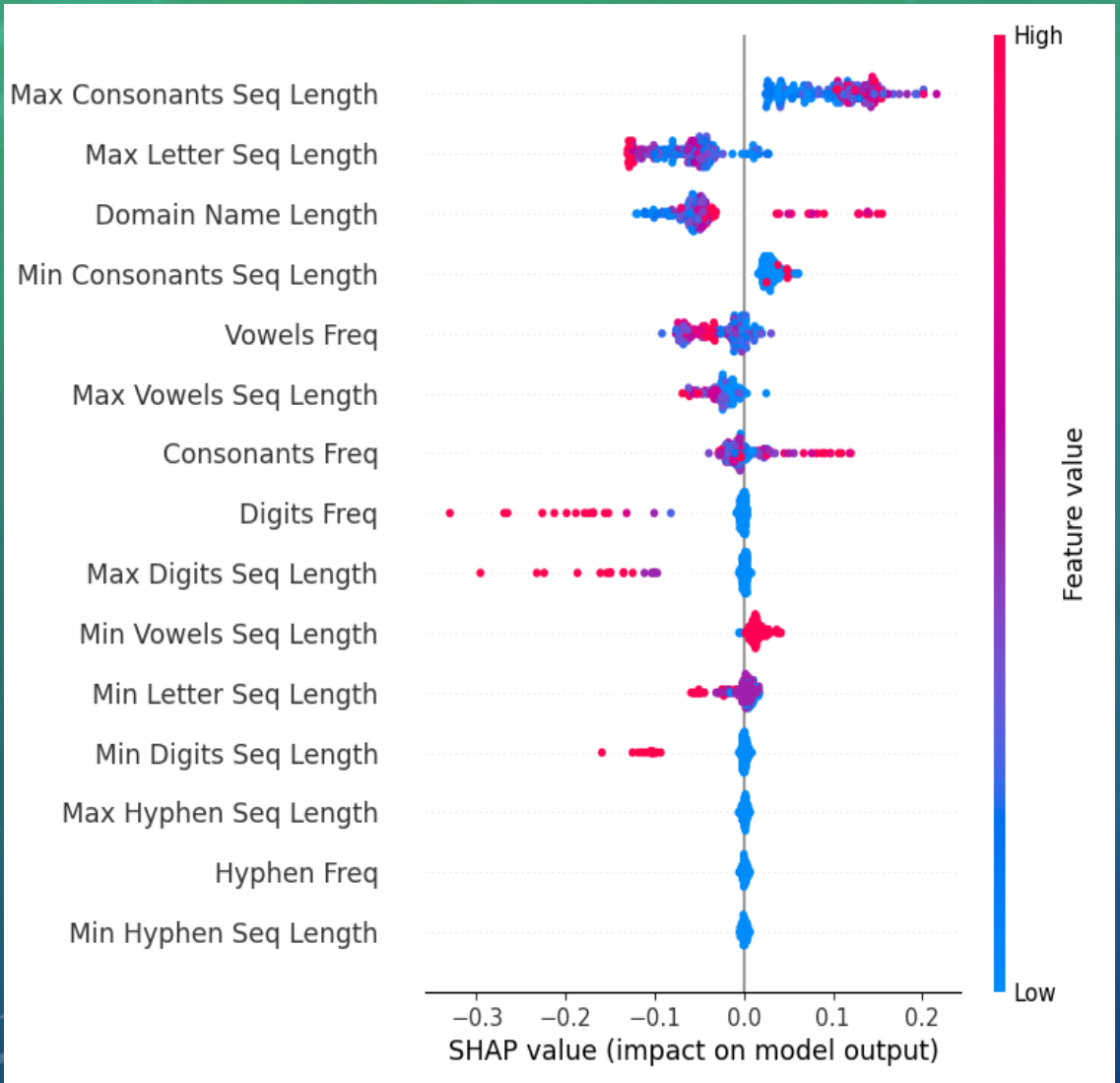
Federated Model Results (6/7)



- Vowels Freq : number of vowels
- Influences the classification to the legitimate class
- Most DGA families have low average vowel frequency, except Banjori and Ramdo

Family	AVG Vowels Freq
Ramnit	4
Kraken	3
Simda	5
Banjori	11
Pykspa	6
Ramdo	9
Qakbot	4
Cryptolocker	4
DirCrypt	5
Corebot	4
Locky	4

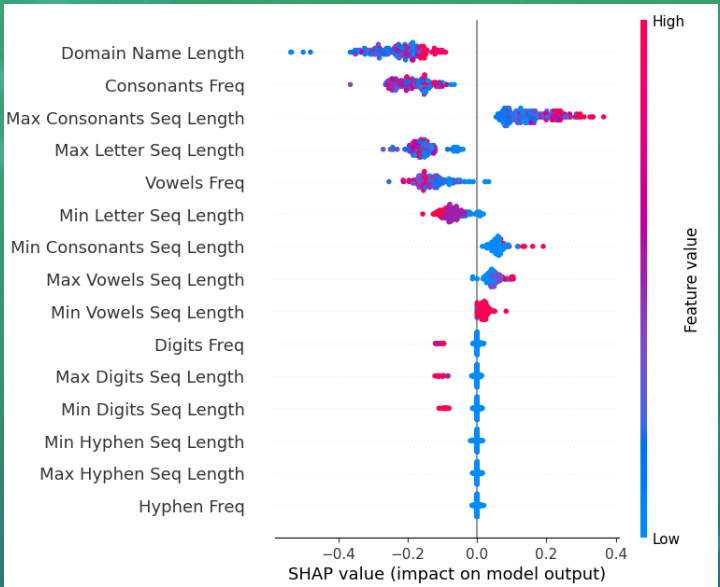
Federated Model Results (7/7)



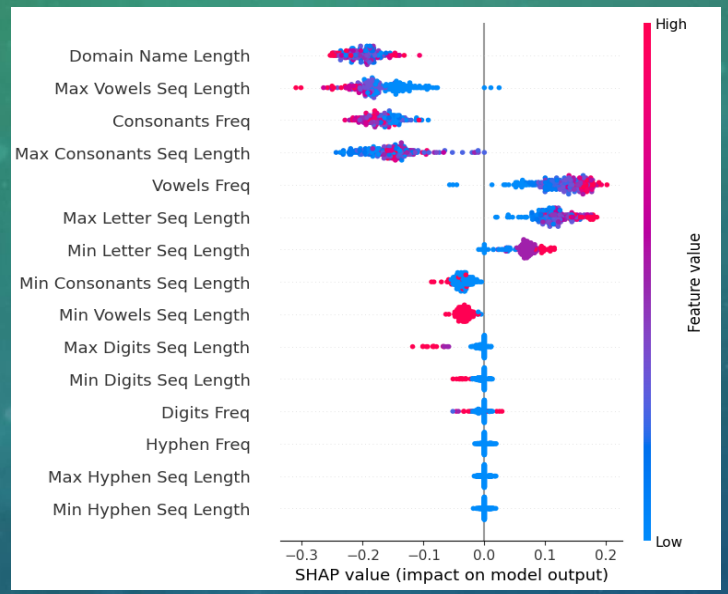
- Features with the special hyphen character do not affect the decision of the model
- The model is mainly influenced by the length of the sequence of letters and the length of the domain name
- Large sequences of digits seem to drive the model towards the class legit
- The features related to consonants are also highly ranked

Clients Results (1/4)

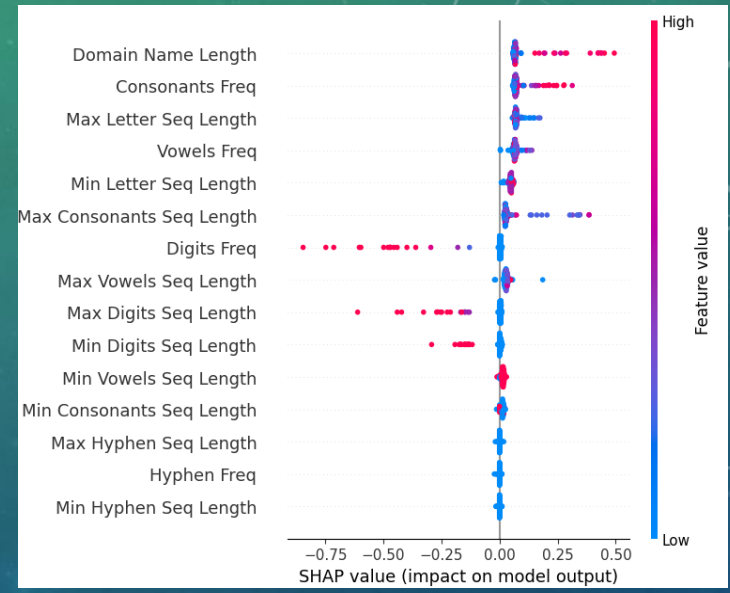
Client 1



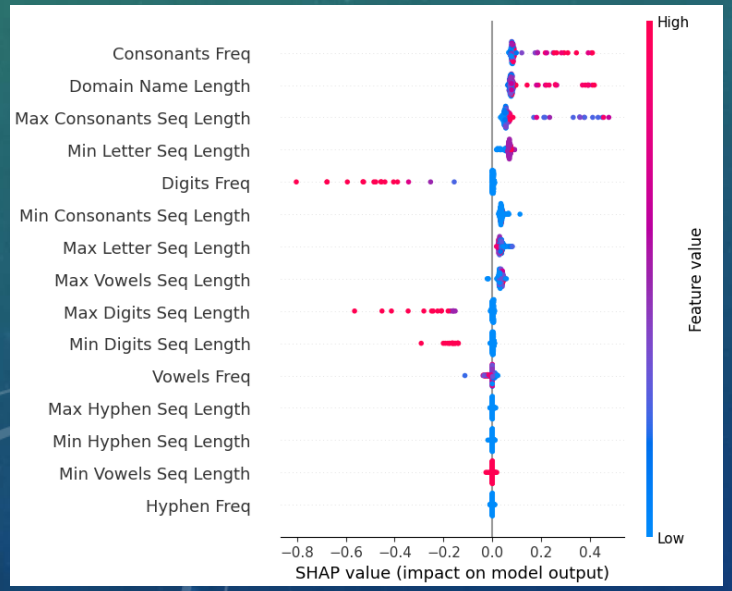
Client 2



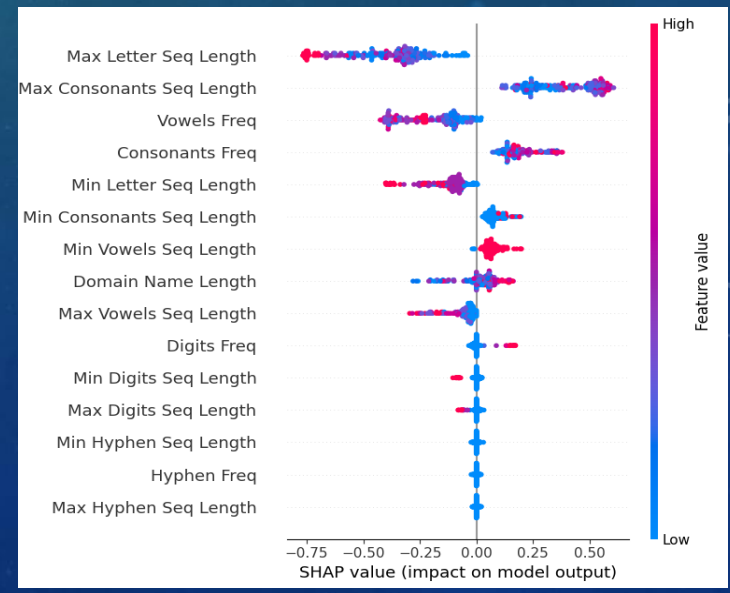
Client 3



Client 4



Client 5

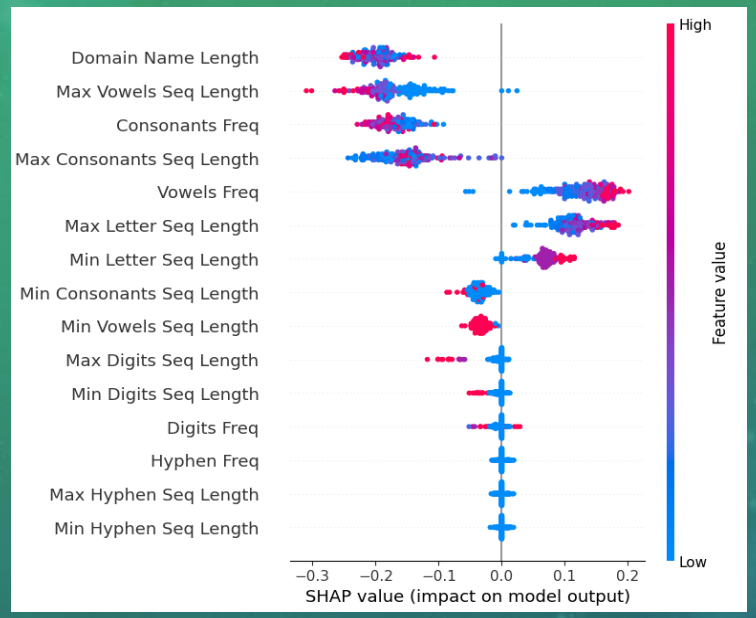


Clients Results (2/4)

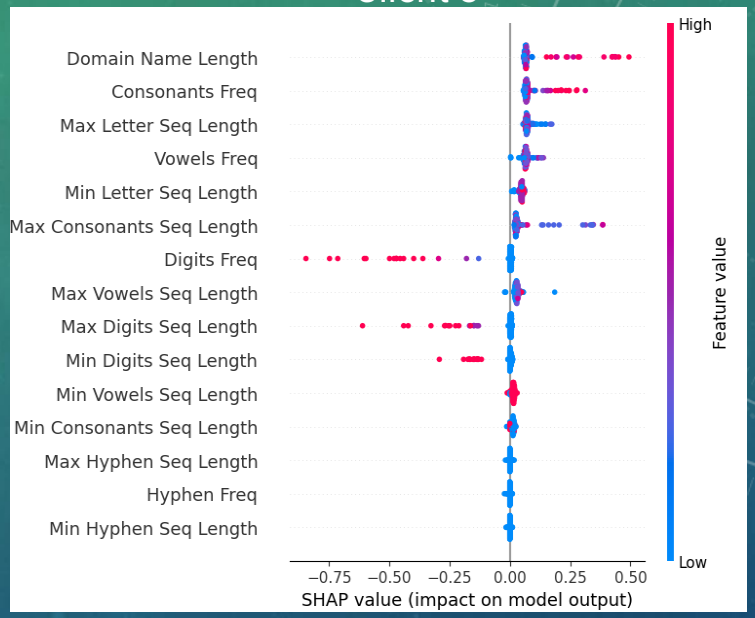
- The most important features that we extracted in the federated model are also found in the clients in the highest importance positions
- Consonant-related features have a high importance position as in the federated model
- Features with the special hyphen character do not affect any client

Clients Results (3/4)

Client 2



Client 3



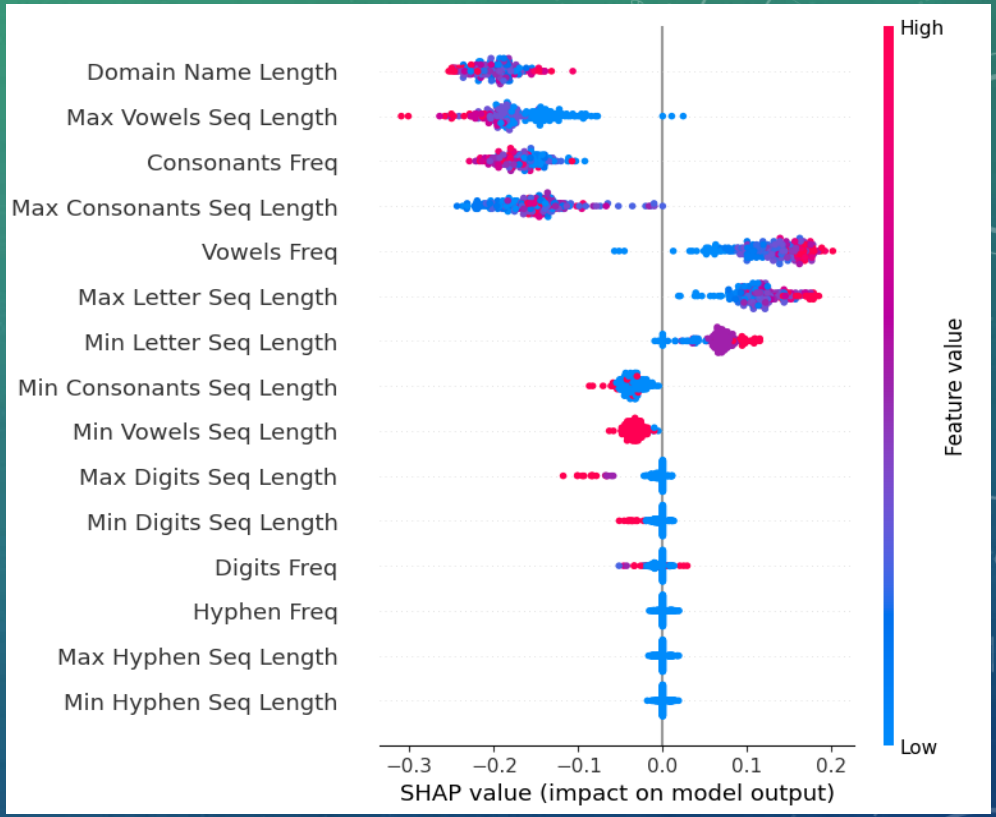
- Vowel frequency is high in the Banjori and Ramdo families
- Clients 2 and 3 drive the data into the DGA class because of these families
- Other clients push their data to the legitimate class

Family	AVG Vowels Freq
Ramnit	4
Kraken	3
Simda	5
Banjori	11
Pykspa	6
Ramdo	9
Qakbot	4
Cryptolocker	4
DirCrypt	5
Corebot	4
Locky	4

Clients Results (4/4)

Client	Family	AVG Max Consonants Seq Length
Client 1	Ramnit	7
	Kraken	6
Client 2	Simda	2
	Banjori	3
Client 3	Pykspa	3
	Ramdo	3
Client 4	Qakbot	6
	Cryptolocker	7
Client 5	DirCrypt	6
	Corebot	5
	Locky	5

Client 2



- max consonants sequence length pushes data to the DGA class on all clients except client 2
- Client 2 includes the Simda and Banjori families, which have the lowest average value for this feature
- For other clients, large values in the attribute lead to the DGA class

Summary

- Detecting legitimate from malicious DGA domain names
- In both cases of heterogeneity, the federated learning model performs satisfactorily.
- The XAI SHAP algorithm combined with deep machine learning models help to better understand malicious network traffic detection methods from DGA.