Explainable AI For Network Data Survey

Explainable Artificial Intelligence (XAI) is a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms.

* Required

Survey Objectives

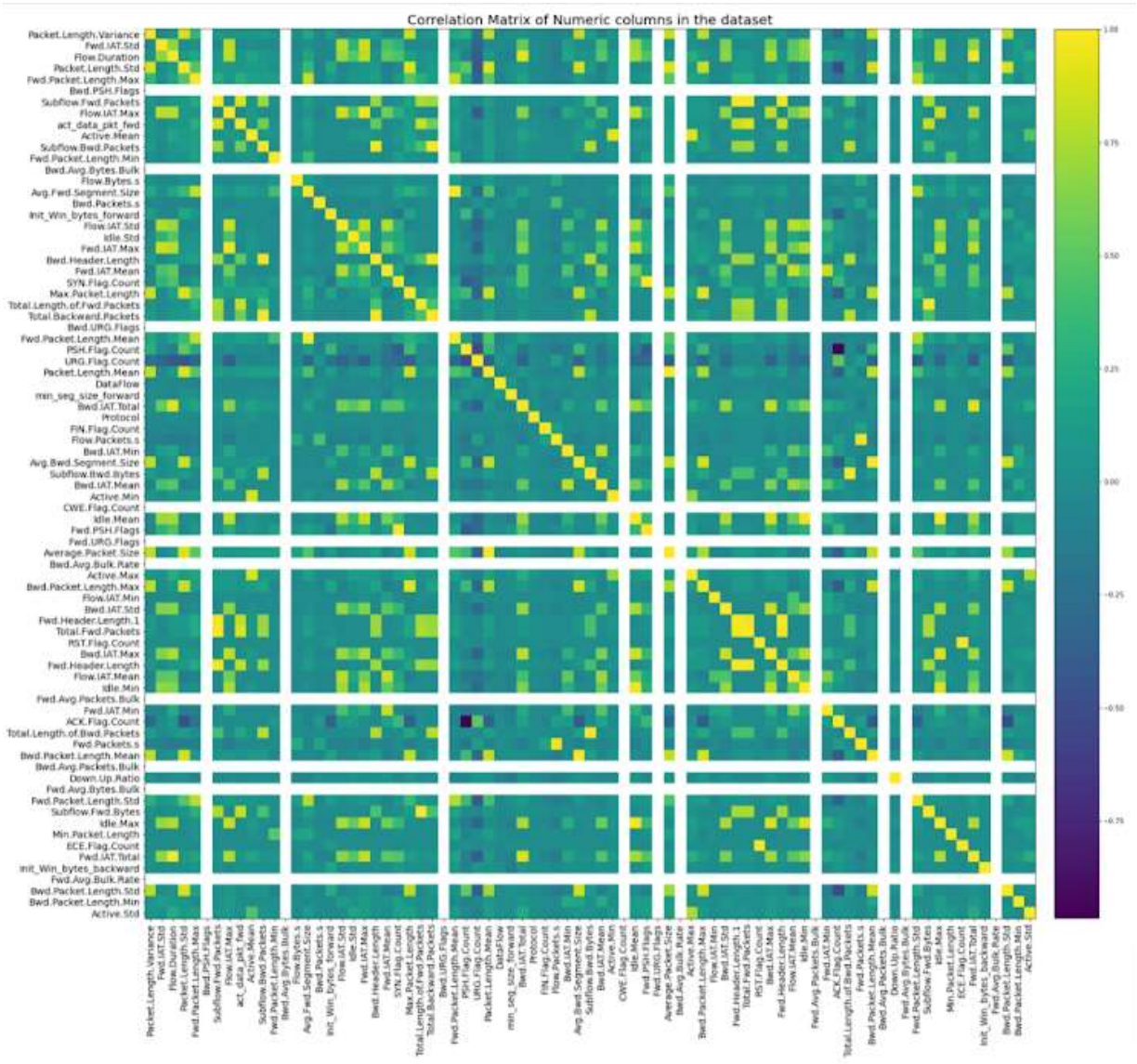
In this survey, you will be shown several visualizations of network data predictions using Explainable AI (XAI) tools to gauge their usability. Each visualization presented was designed from one of three machine learning models. Most questions will ask your overall sentiment toward each visualization in terms of simplicity and understanding. At the end of the survey, we would like you to pick which group of visualizations you think represents a trustworthy predictive model to use on Network Data.

The XAI tool used for these visualizations is the open-source XAI library, SHAP (https://shap.readthedocs.io/). The library uses Shapley (SHAP) values to explain the impact of each feature in a dataset based on the predicted outcomes.

Dataset Overview

The dataset used for our analysis consists of network data with 87 recorded features such as flow rate, packet length, and ACK count. Our models are trained to predict the application type (i.e. Amazon, Google, or YouTube) being accessed using some combination of the recorded features. A description for each feature of the dataset can be found on it's Kaggle page: https://www.kaggle.com/code/jeevan98/categorizing-dataset

1. This XAI visualization depicts the dataset's feature correlations. Despite the large feature set, is this visualization.... *



(Please select up to 2 answers)

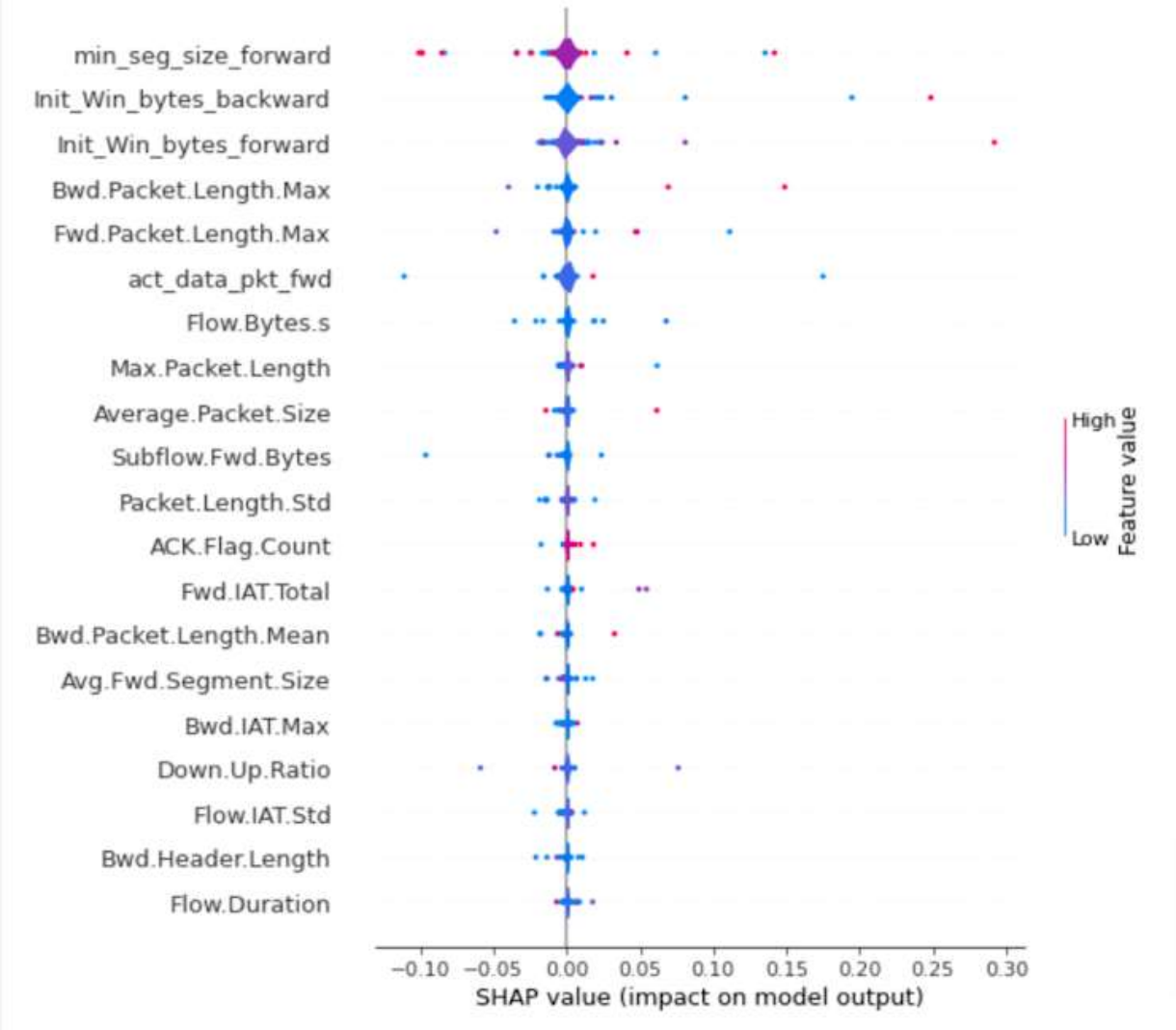
Check all that apply.

- [ ] Easy to understand and simple
- [ ] Difficult to understand and complex
- [ ] Interesting XAI visualization, but should be more concise
- [ ] Don't think this is an XAI visualization...?

Model A

The following visualizations represent the relationships between a predicted Application Type (Amazon) connection versus all other features in the dataset using Model A.

2. What level of significance do you think the features in the plot below, impact Model A's ability in predicting the Protocol connection to Amazon? *



This is a standard violin plot but with outliers drawn as points. The color represents the average feature value at that position, so red regions have mostly high valued feature values while blue regions have mostly low feature values.

Mark only one oval.

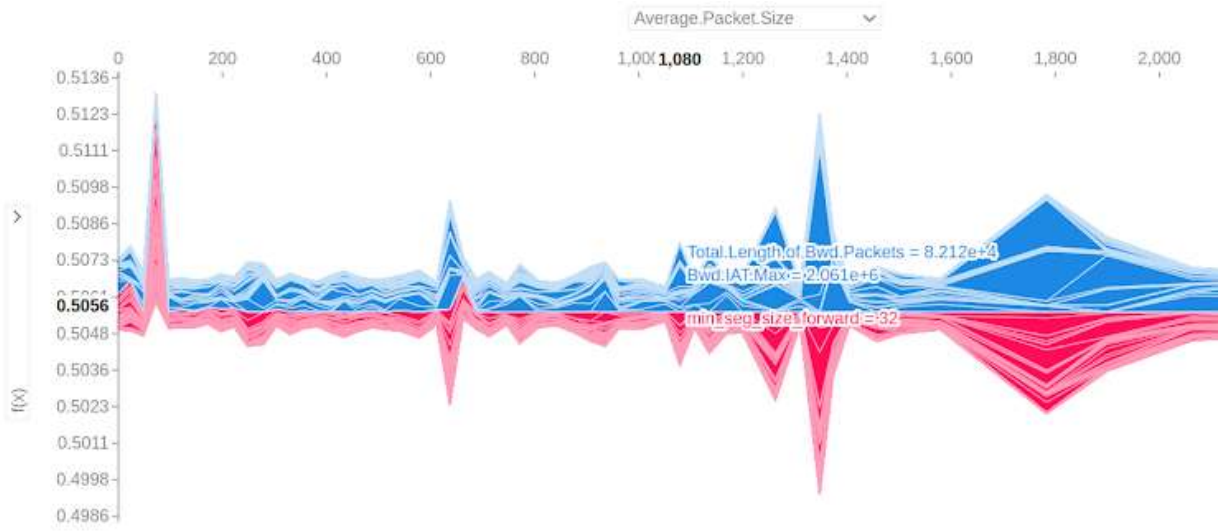
- ◯ Most features impact significantly
- ◯ Little features impact significantly
- ◯ No features impact significantly

3. What is your overall sentiment towards the simplicity or complexity of the previous visualization above? *

Mark only one oval.

	1	2	3	4	5	
Simple (easy to understand)	◯	◯	◯	◯	◯	Complex (not a good representation)

4. How much does the Average Packet Size feature impact the features in the predicted Protocol to Amazon? *



Mark only one oval.

- ◯ The average packet size greatly impacts the features presented
- ◯ The average packet size impacts half of the observed connections with the features presented
- ◯ The average packet size has little impact to the observed connections with the features presented

5. What is your overall sentiment towards the simplicity or complexity of the previous visualization above? *

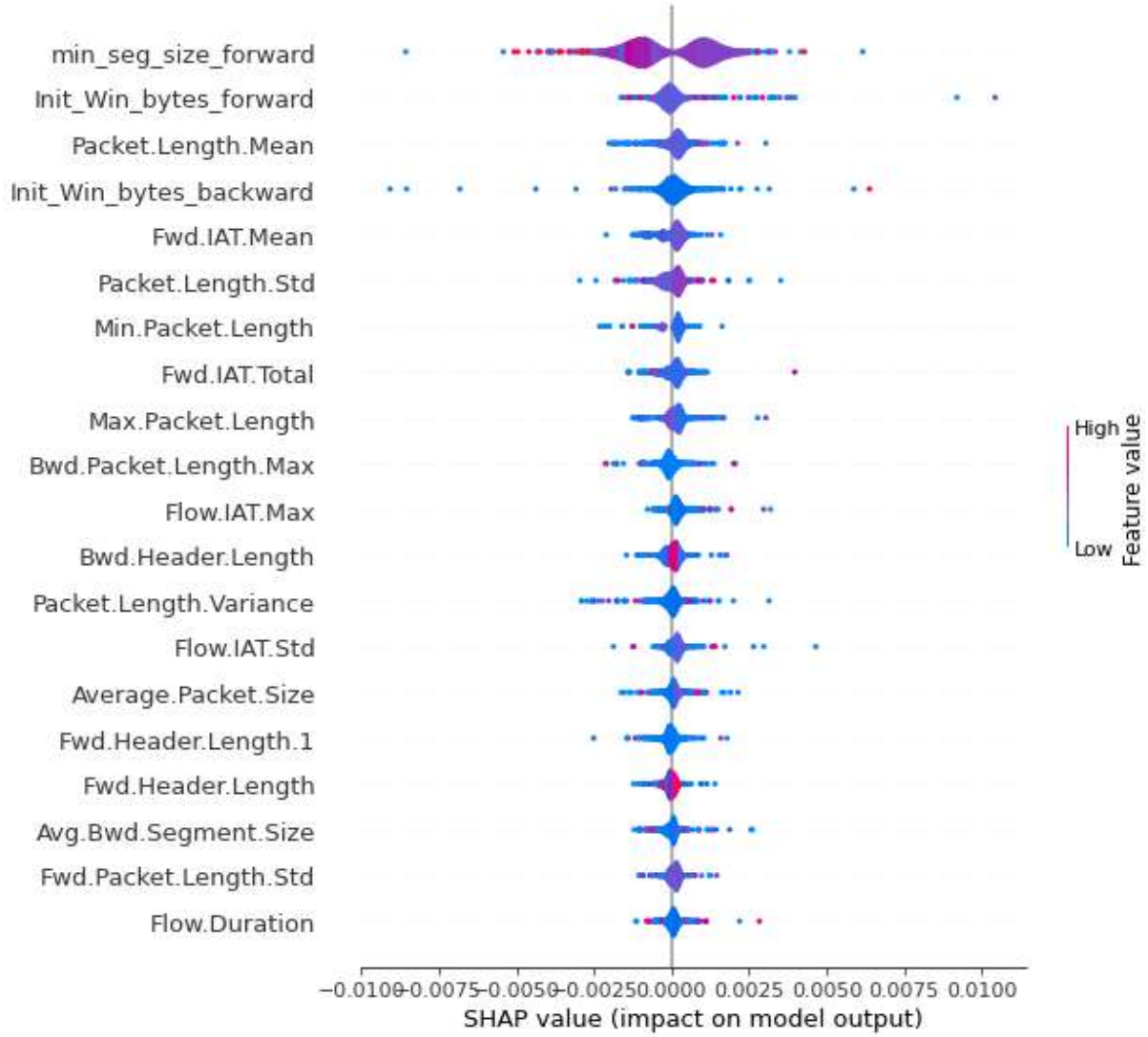
Mark only one oval.

	1	2	3	4	5	
Simple (easy to understand)	◯	◯	◯	◯	◯	Complex (not a good representation)

Model B

The following visualizations represent the relationships between a predicted Application Type (Amazon) connection versus all other features in the dataset using Model B.

6. What level of significance do you think the features in the plot below, impact Model B's ability in predicting the Protocol connection to Amazon? *



Mark only one oval.

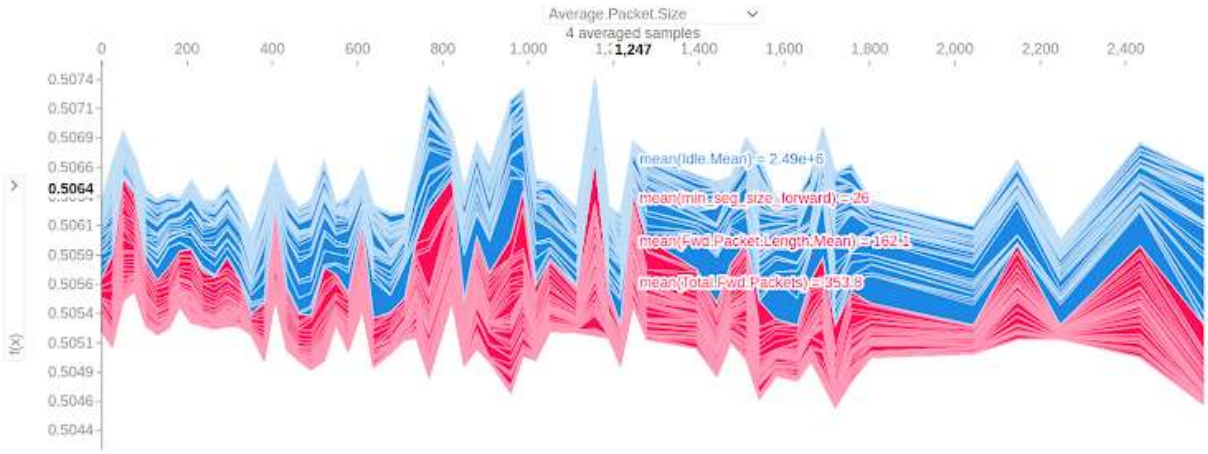
- ◯ Most features impact significantly
- ◯ Little features impact significantly
- ◯ No features impact significantly

7. What is your overall sentiment towards the simplicity or complexity of the previous visualization above? *

Mark only one oval.

	1	2	3	4	5	
Simple (easy to understand)	◯	◯	◯	◯	◯	Complex (not a good representation)

8. How much does the Average Packet Size feature impact the presented features in the predicted Protocol to Amazon? *



Mark only one oval.

◯ The average packet size greatly impacts the features presented

◯ The average packet size impacts half of the observed connections with the features presented

◯ The average packet size has little impact to the observed connections with the features presented

9. What is your overall sentiment towards the simplicity or complexity of the previous visualization above? *

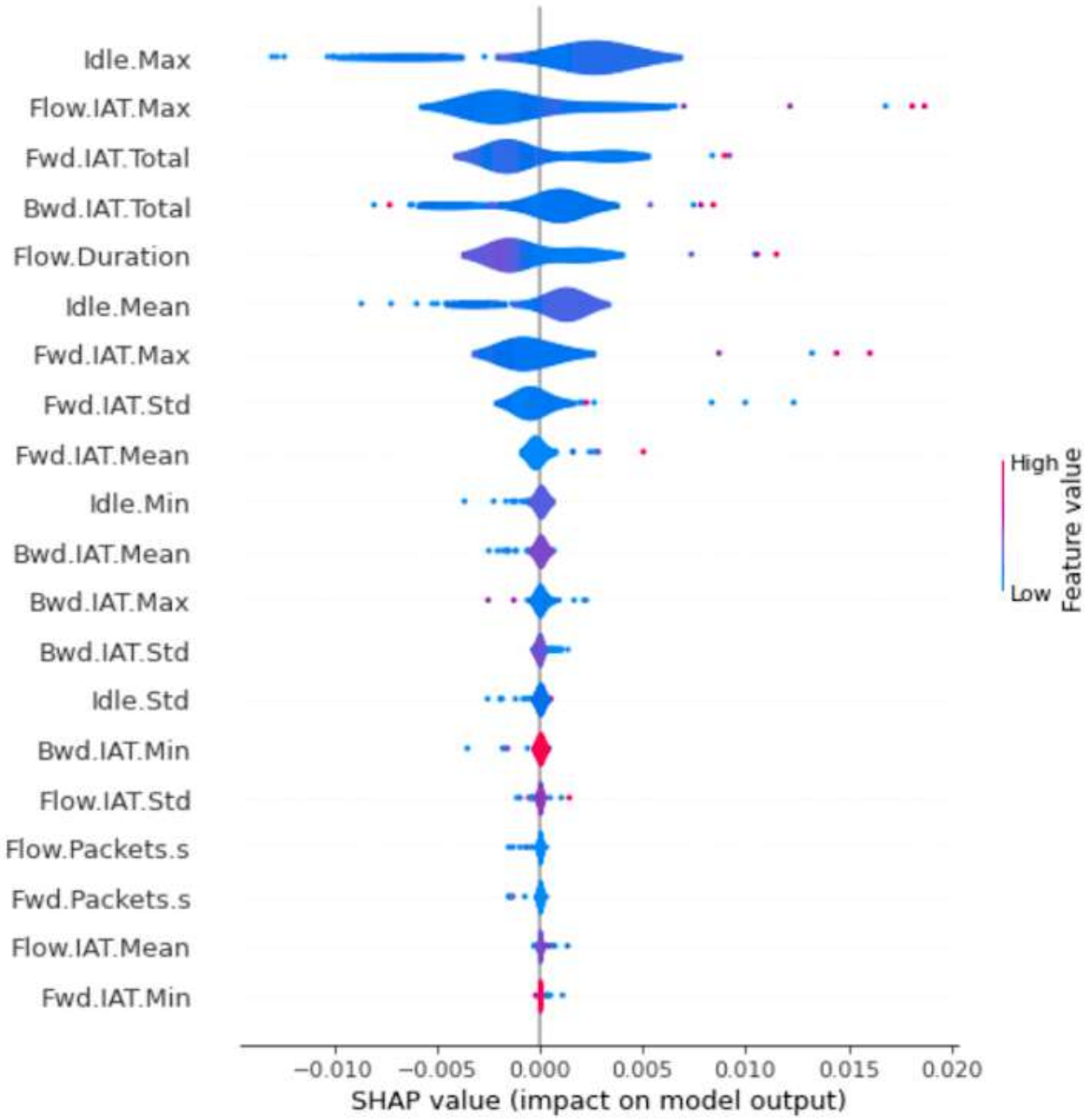
Mark only one oval.

	1	2	3	4	5	
Simple (easy to understand)	◯	◯	◯	◯	◯	Complex (not a good representation)

Model C

The following visualizations represent the relationships between a predicted Application Type (Amazon) connection versus all other features in the dataset using Model C.

10. What level of significance do you think the features in the plot below, impact Model C's ability in predicting the Protocol connection to Amazon? *



Mark only one oval.

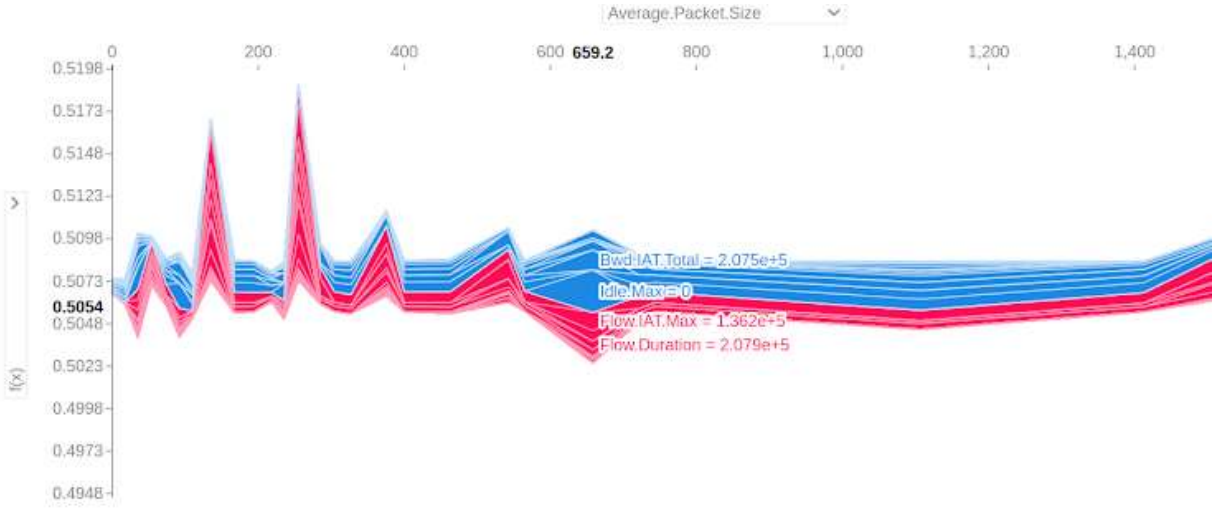
- ◯ Most features impact significantly
- ◯ Little features impact significantly
- ◯ No features impact significantly

11. What is your overall sentiment towards the simplicity or complexity of the previous visualization above? *

Mark only one oval.

	1	2	3	4	5	
Simple (easy to understand)	◯	◯	◯	◯	◯	Complex (not a good representation)

12. How much does the Average Packet Size feature impact the Flow Duration and packet inter-arrival times (IAT) features in the predicted Protocol to Amazon? *



Mark only one oval.

◯ The average packet size greatly impacts Flow Duration and IAT features

◯ The average packet size impacts half of the observed connections with Flow Duration and IAT features

◯ The average packet size has little impact to the observed connections with Flow Duration and IAT features

13. What is your overall sentiment towards the simplicity or complexity of the previous visualization above? *

Mark only one oval.

	1	2	3	4	5	
Simple (easy to understand)	◯	◯	◯	◯	◯	Complex (not a good representation)

Final Selection

14. Based on all the visualizations observed, which Model do you think was the most accurate in predicting the Amazon Protocol?

Mark only one oval.

- ◯ Model A
- ◯ Model B
- ◯ Model C