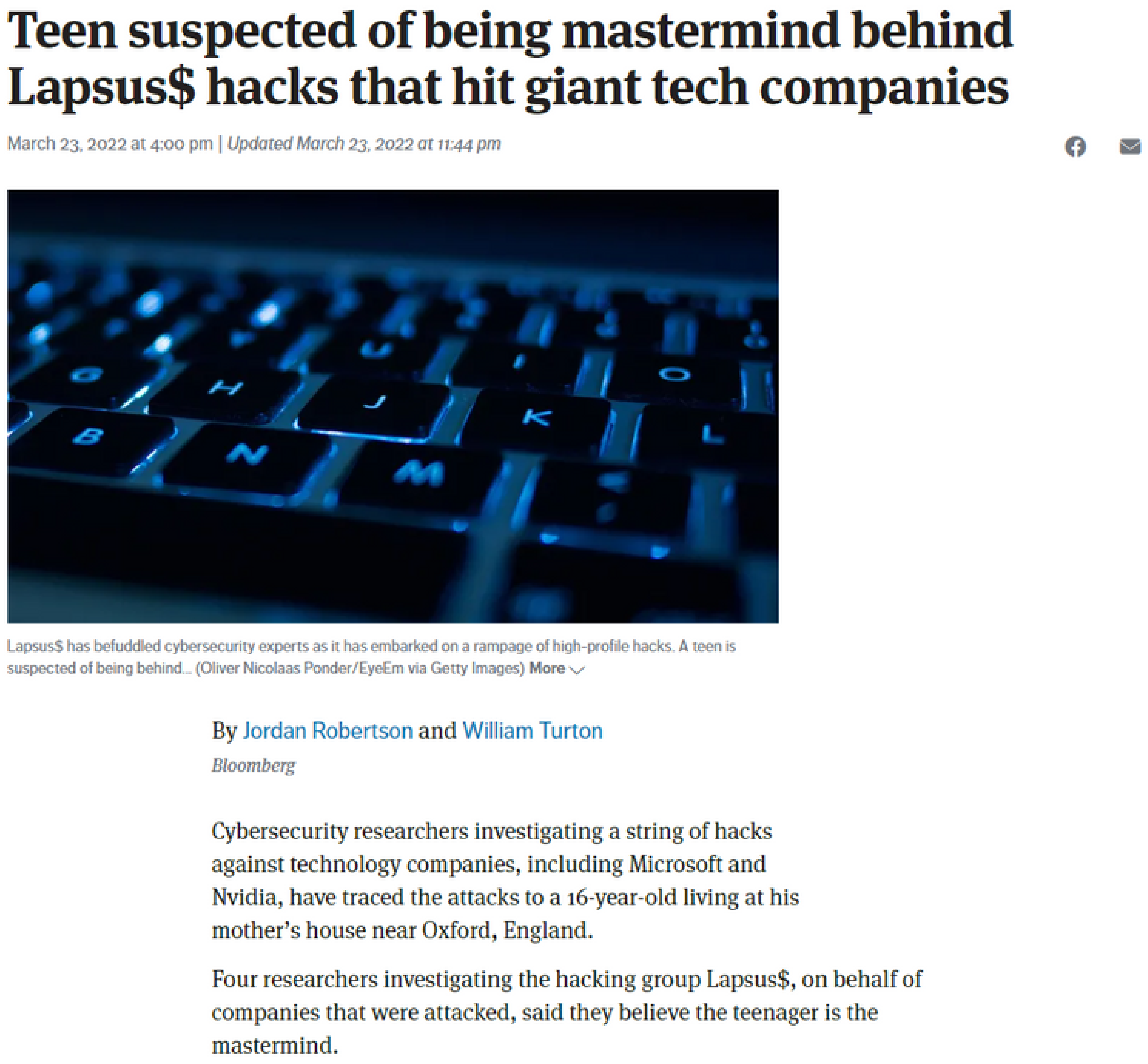
Using ML/AI to Identify

Cybersecurity Threats

Joshua Lin

Problem Statement

Cybersecurity attacks can result in credit card information, customer addresses, customer emails, logins, and other proprietary documents being accessed by unauthorized attackers. This is an increasingly growing problem that, when unchecked, can potentially result in entire companies collapsing. This project hopes to use digital packet and payload behavior to create an effective network intrusion detection system (IDS) to identify when an unauthorized attacker is probing a server.



Teen suspected of being mastermind behind Lapsus$ hacks that hit giant tech companies

March 23, 2022 at 4:00 pm | *Updated March 23, 2022 at 11:44 pm*

Lapsus$ has befuddled cybersecurity experts as it has embarked on a rampage of high-profile hacks. A teen is suspected of being behind... (Oliver Nicolaas Ponder/EyeEm via Getty Images) More

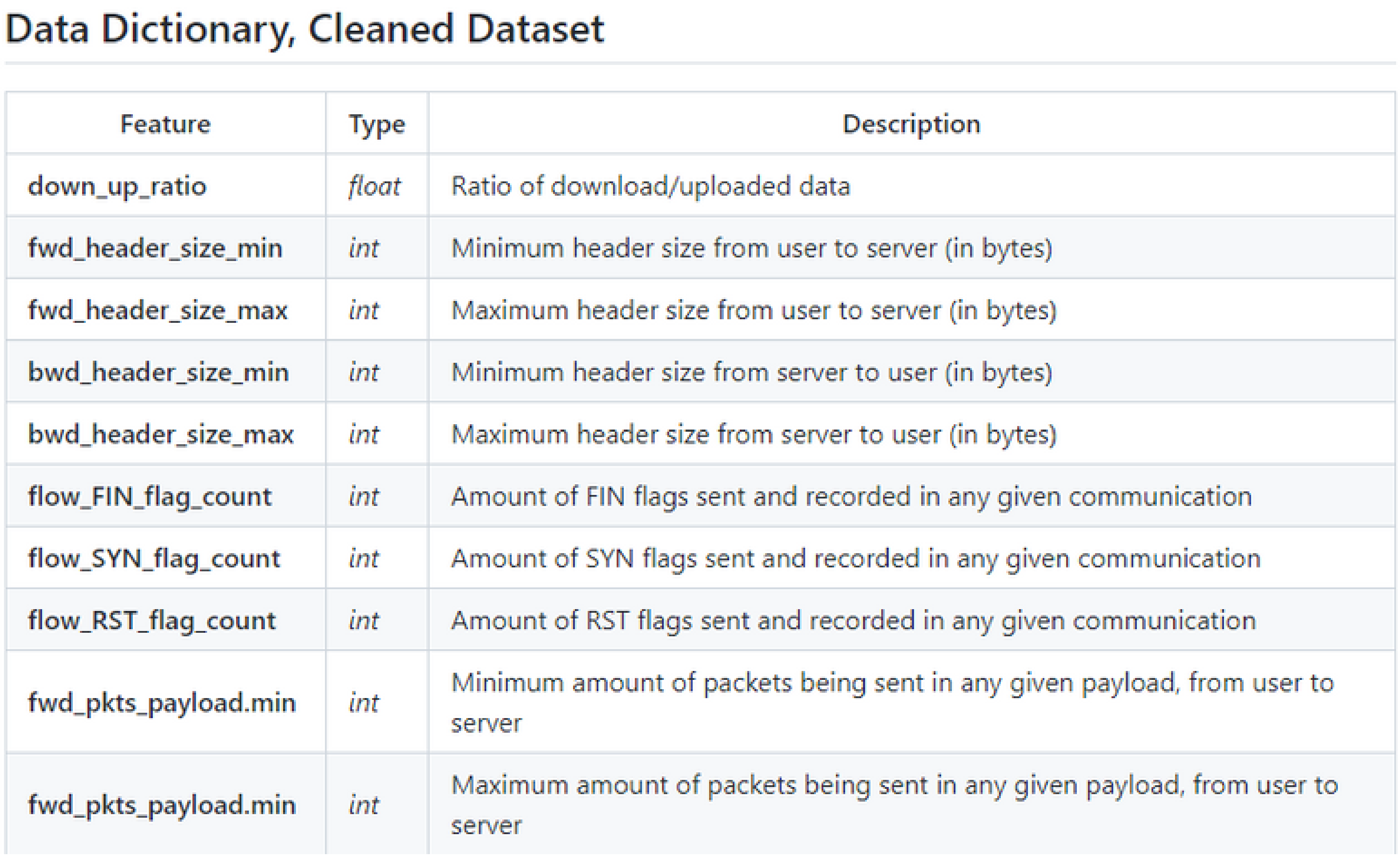
By Jordan Robertson and William Turton

Bloomberg

Cybersecurity researchers investigating a string of hacks against technology companies, including Microsoft and Nvidia, have traced the attacks to a 16-year-old living at his mother's house near Oxford, England.

Four researchers investigating the hacking group Lapsus$, on behalf of companies that were attacked, said they believe the teenager is the mastermind.

Elements of the Dataset



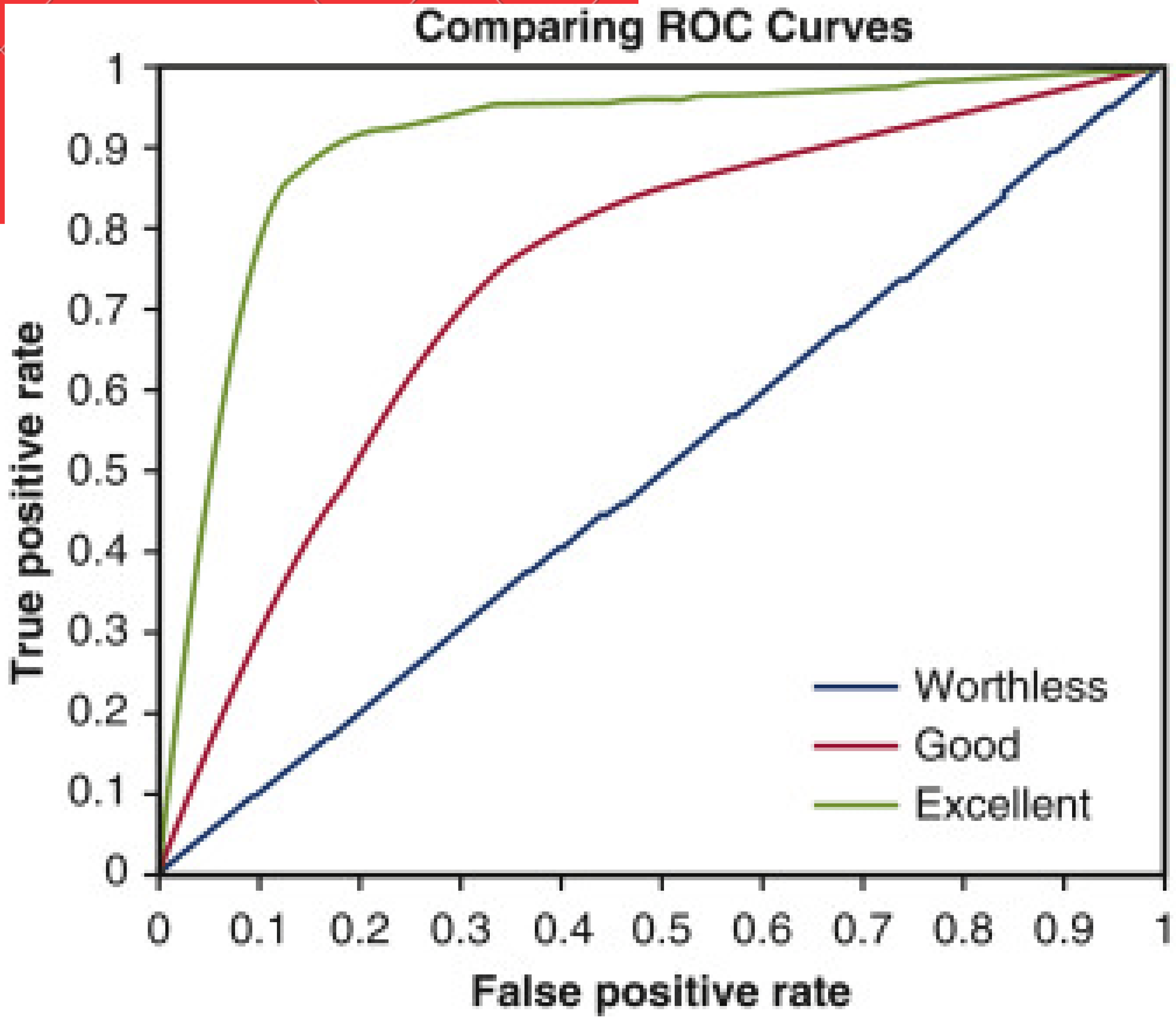
Data Dictionary, Cleaned Dataset

Feature	Type	Description
down_up_ratio	*float*	Ratio of download/uploaded data
fwd_header_size_min	*int*	Minimum header size from user to server (in bytes)
fwd_header_size_max	*int*	Maximum header size from user to server (in bytes)
bwd_header_size_min	*int*	Minimum header size from server to user (in bytes)
bwd_header_size_max	*int*	Maximum header size from server to user (in bytes)
flow_FIN_flag_count	*int*	Amount of FIN flags sent and recorded in any given communication
flow_SYN_flag_count	*int*	Amount of SYN flags sent and recorded in any given communication
flow_RST_flag_count	*int*	Amount of RST flags sent and recorded in any given communication
fwd_pkts_payload.min	*int*	Minimum amount of packets being sent in any given payload, from user to server
fwd_pkts_payload.min	*int*	Maximum amount of packets being sent in any given payload, from user to server

Size of packet/payloads sent

Rate at which data is sent back and forth

TSP Flags



Classification Problem

Overall goal: To create a model that can successfully identify the difference between an authorized user and an attacker probing a server.

Definition of success for this problem: To have a model with an overall accuracy rate of 99.7% or higher.

Approach

Methodology

Feature Engineering

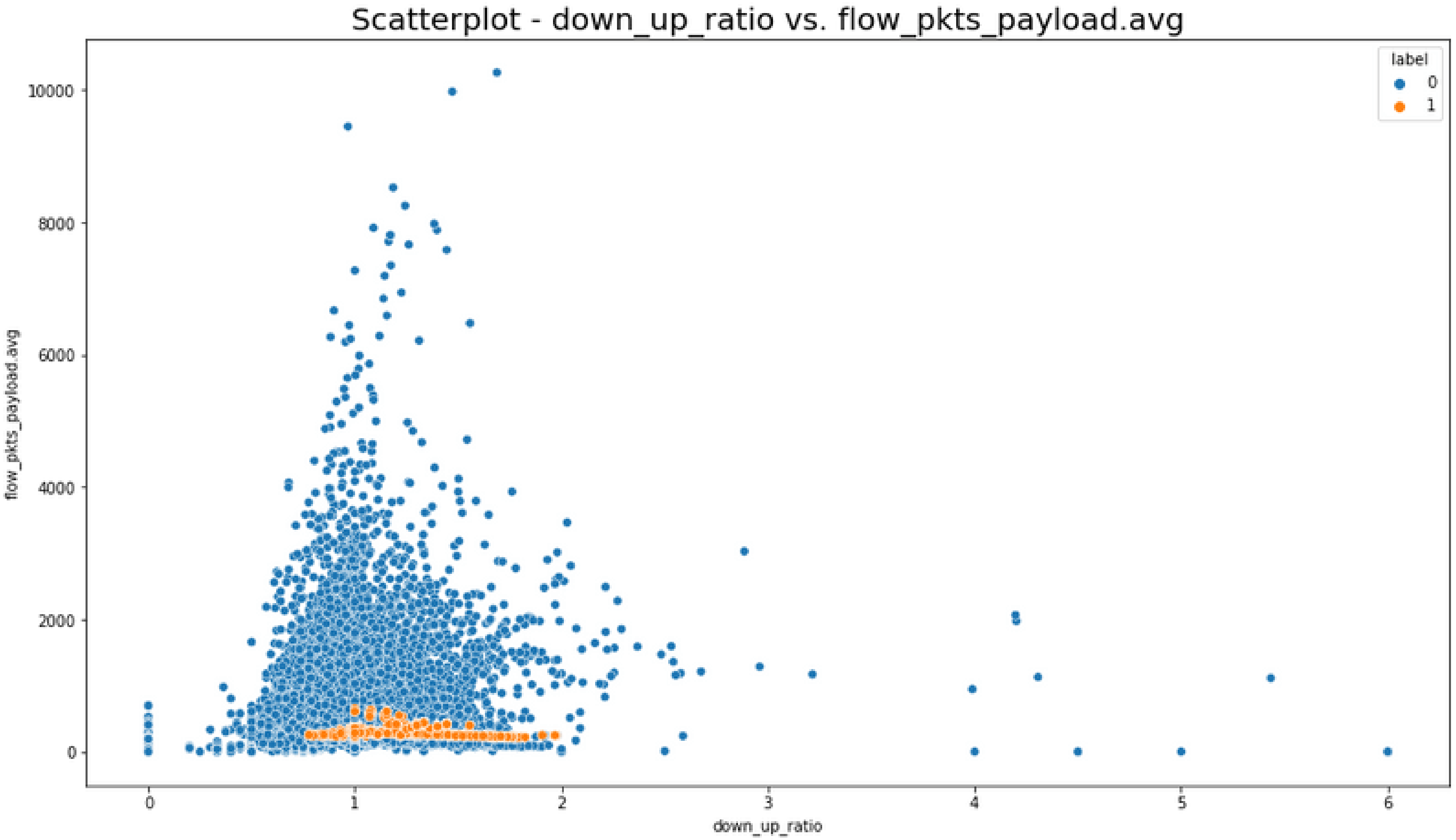
Quantitative, automated feature engineering

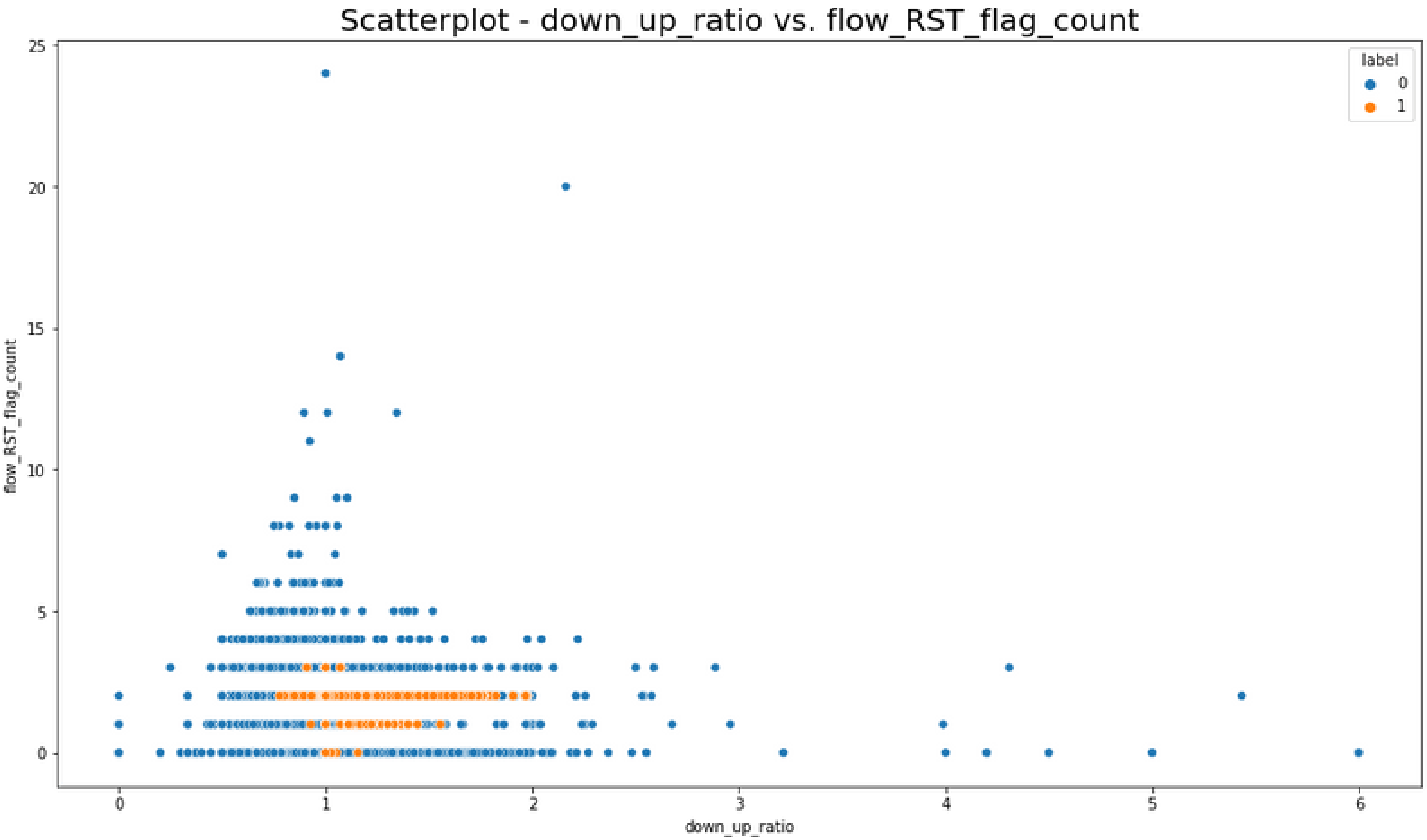
Exploratory Data Analysis

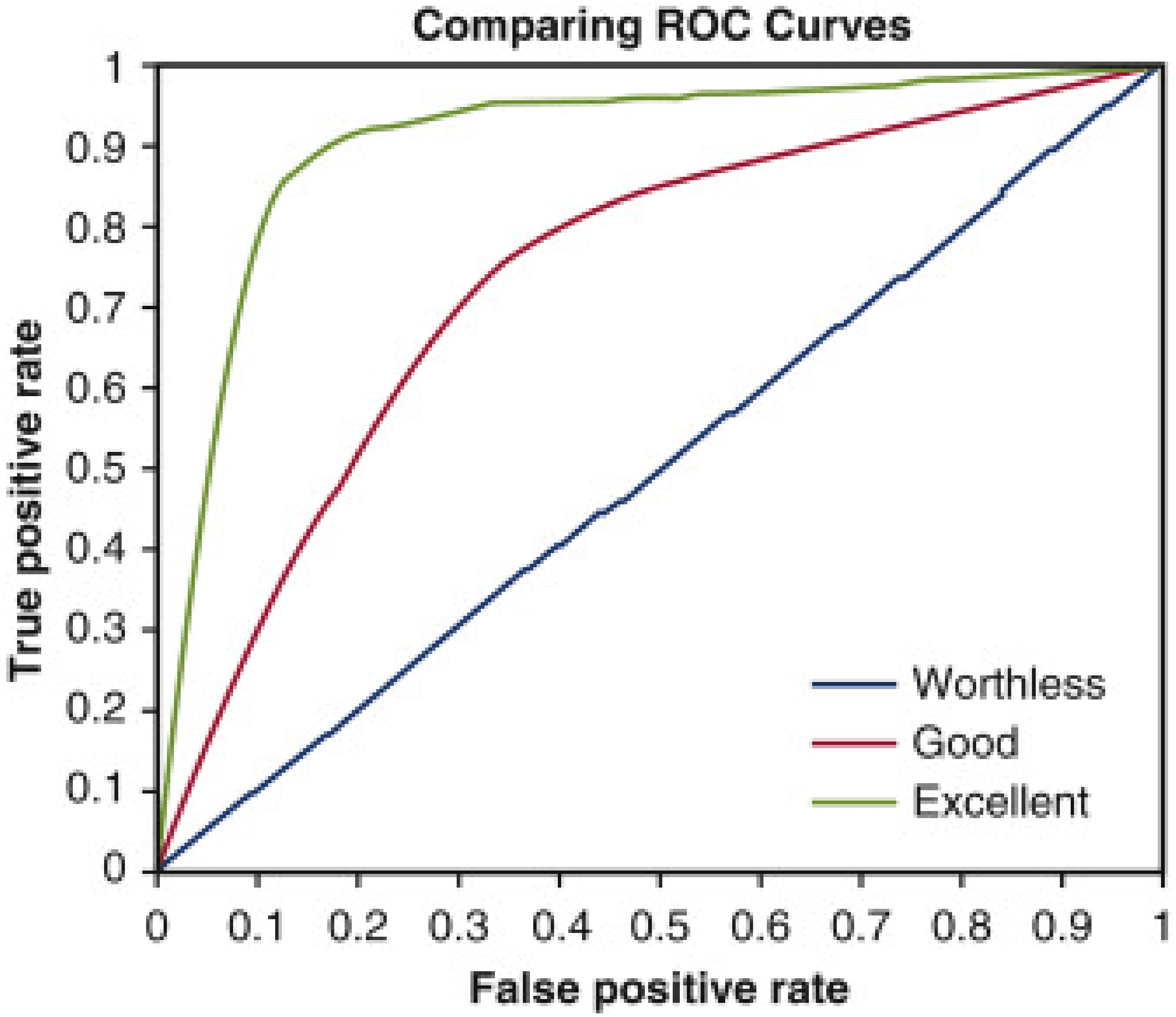
Primary finding: attackers try to mimic the digital behavior of authorized users while retrieving as much information as possible, and leaving behind the least amount of data as possible

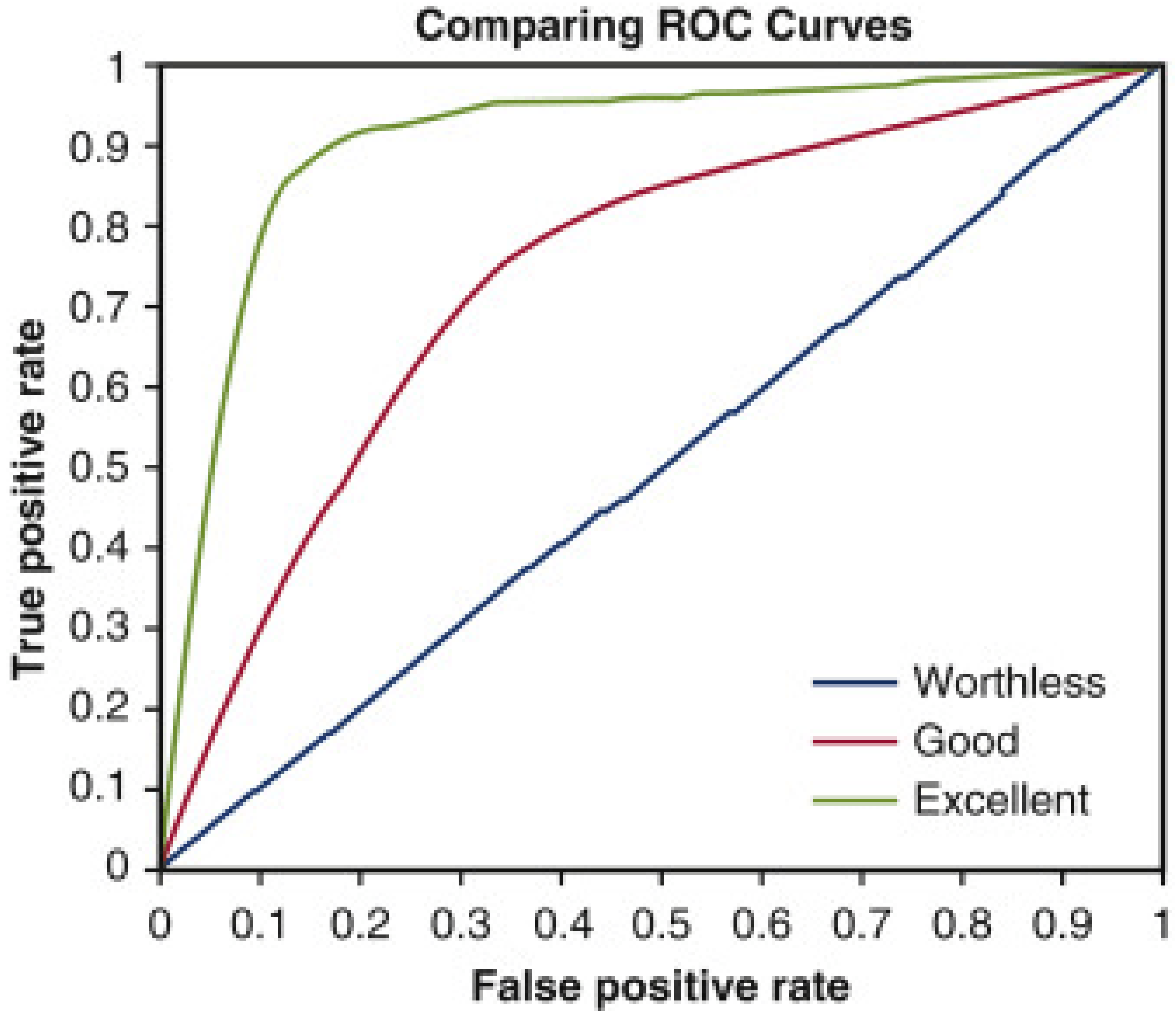
Modeling

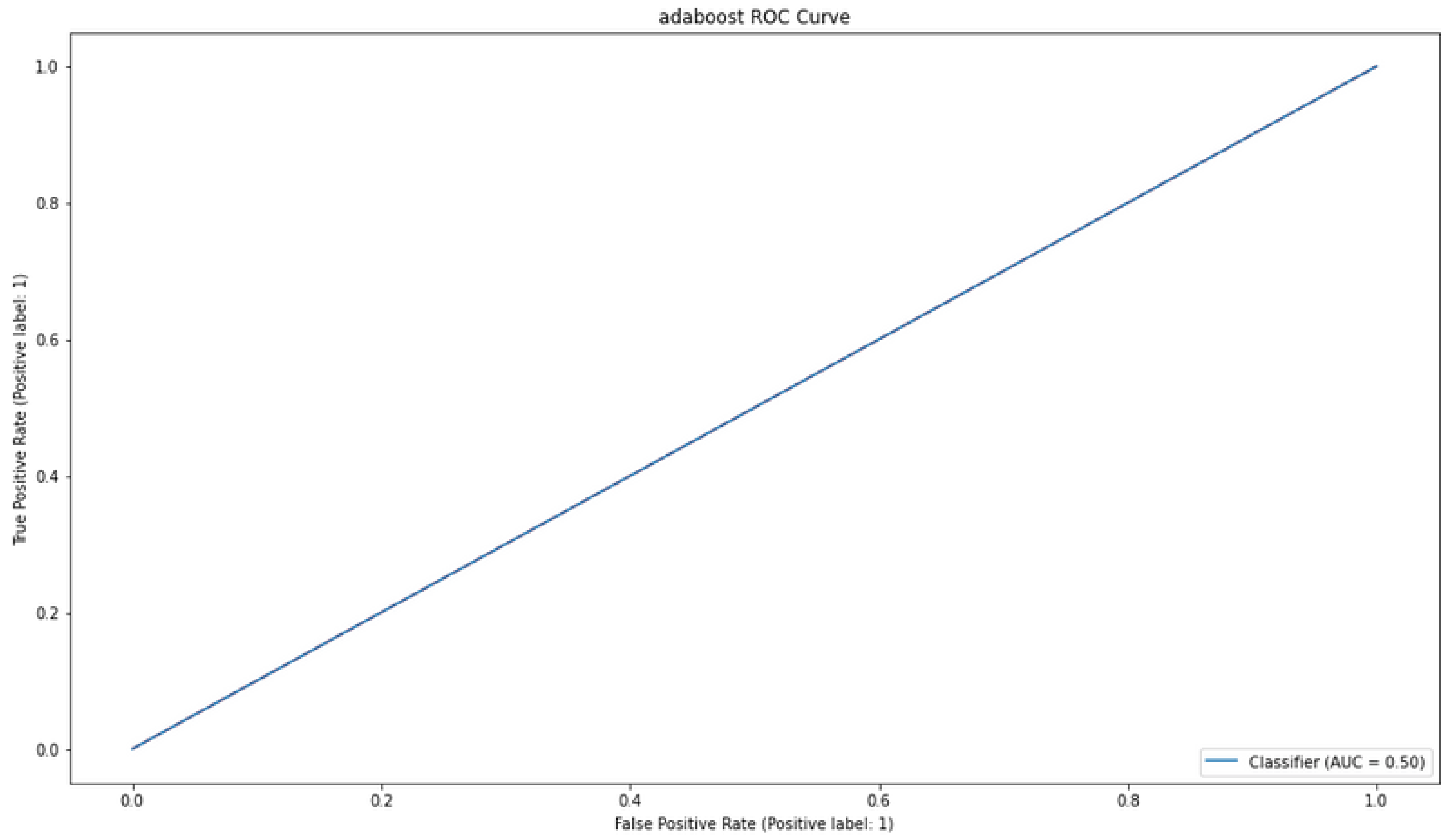
Neural network deployment

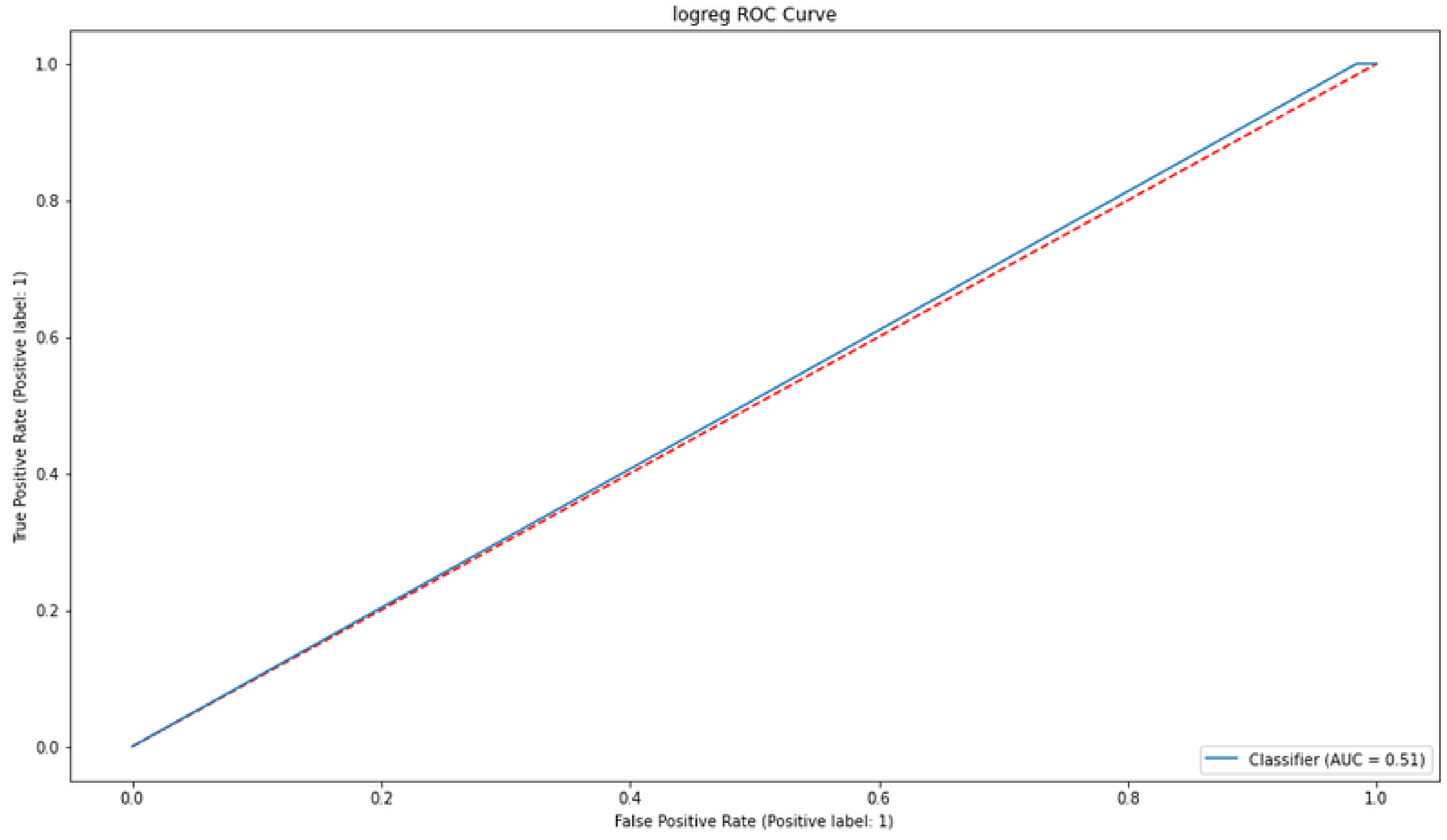


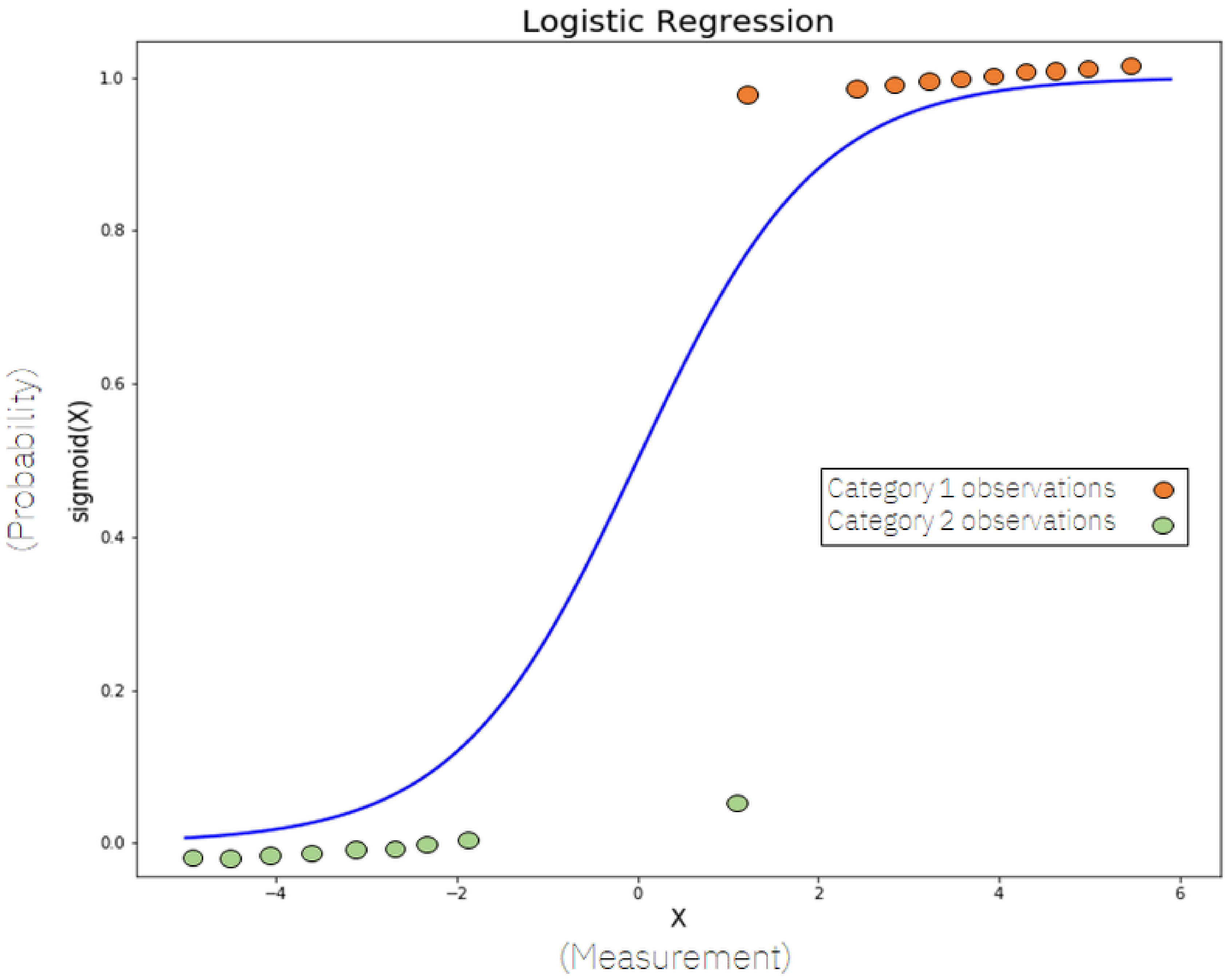

Scatterplot - down_up_ratio vs. flow_RST_flag_count
label
0
1
flow_RST_flag_count
0
5
10
15
20
25
down_up_ratio
0
1
2
3
4
5
6


Comparing ROC Curves
True positive rate
False positive rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Worthless
Good
Excellent

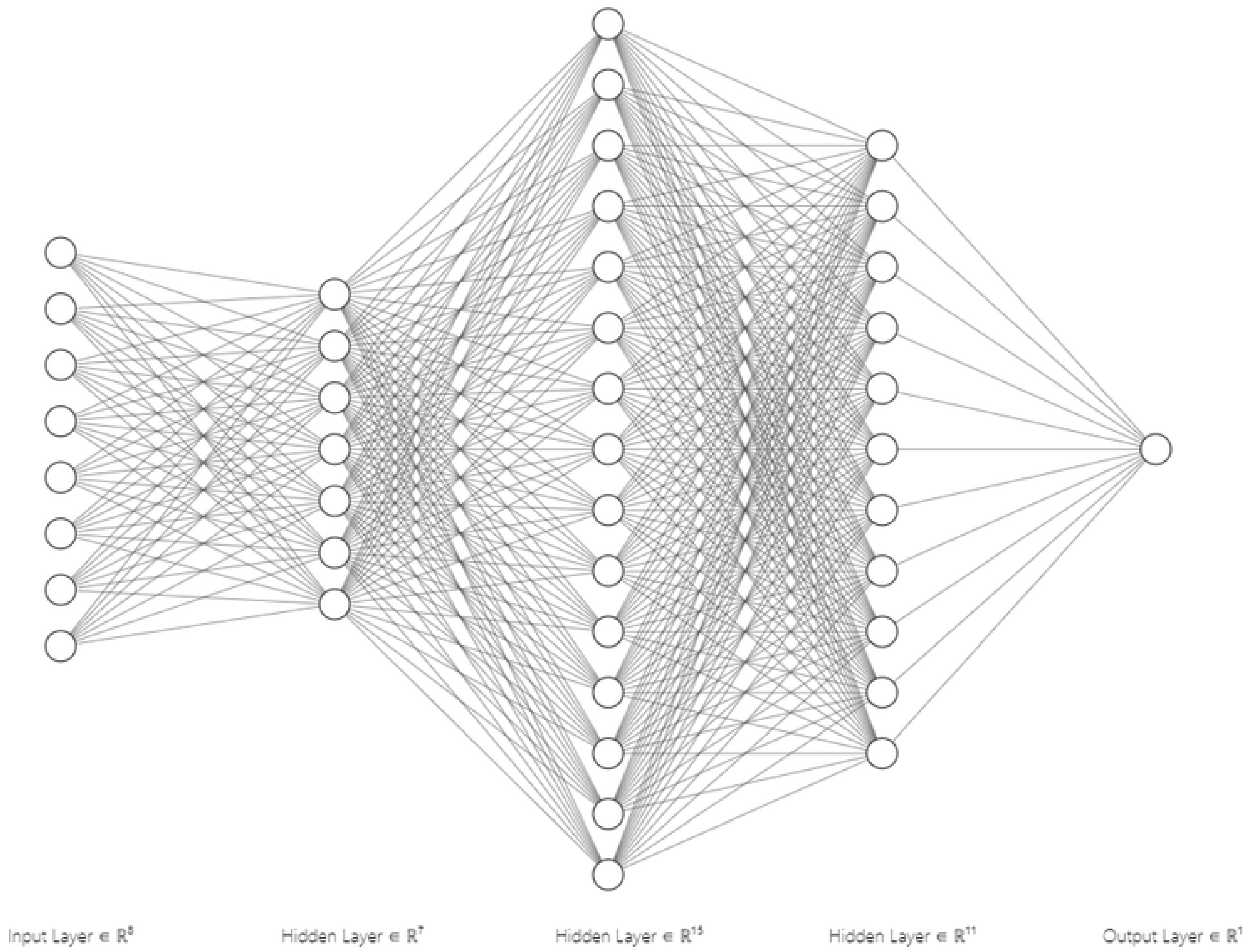

Comparing ROC Curves
True positive rate
False positive rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Worthless
Good
Excellent

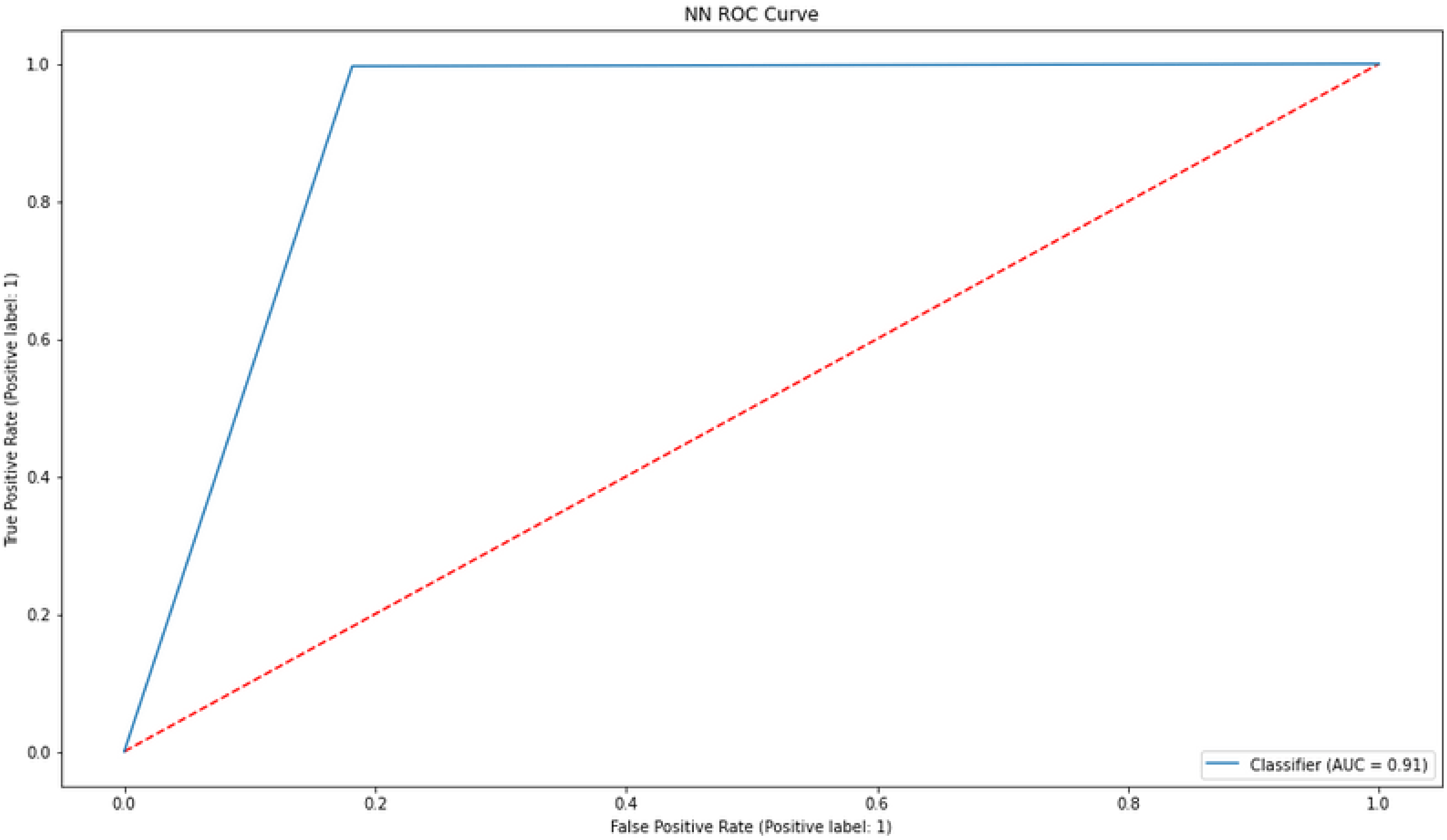

adaboost ROC Curve
True Positive Rate (Positive label: 1)
False Positive Rate (Positive label: 1)
0.0
0.2
0.4
0.6
0.8
1.0
Classifier (AUC = 0.50)

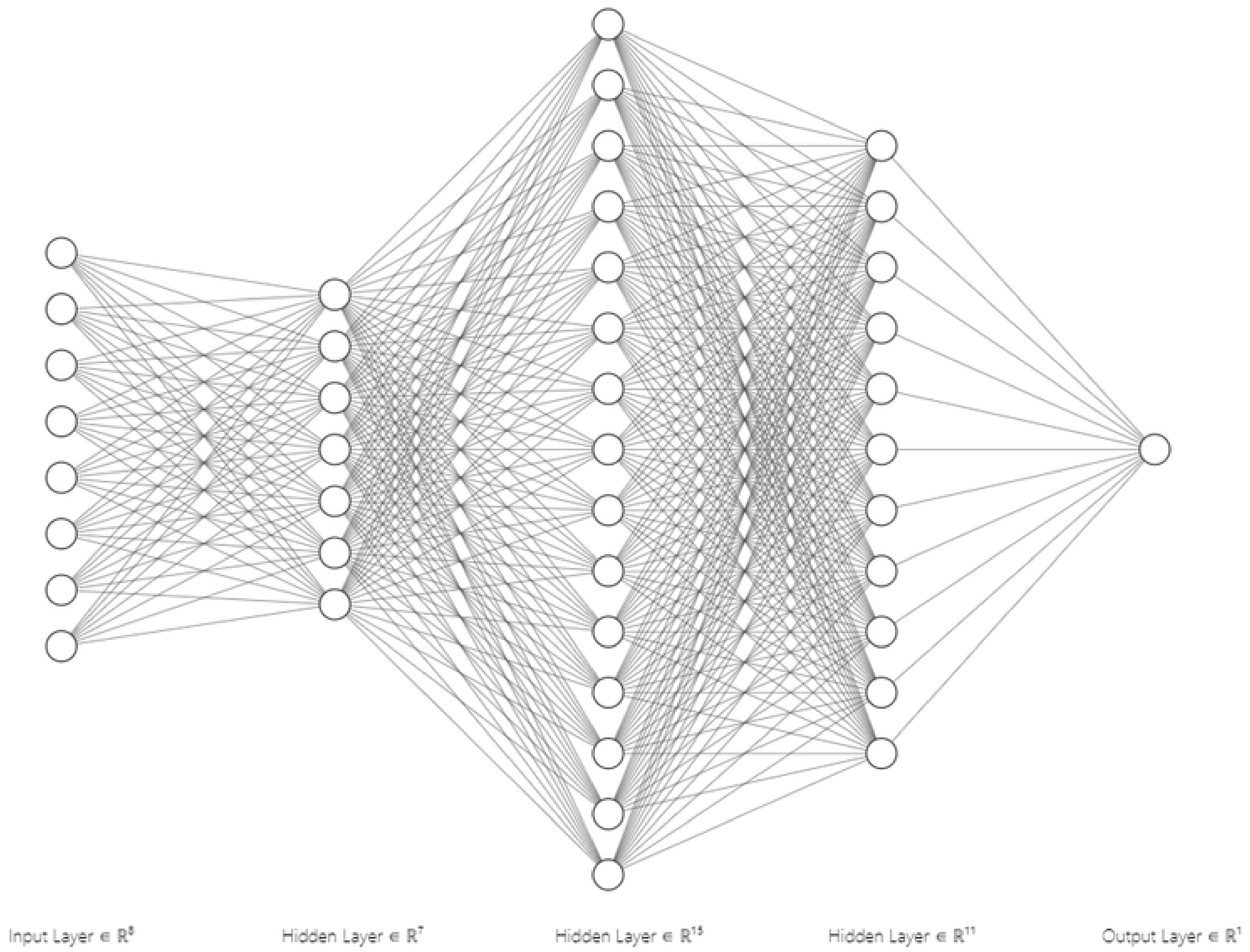

logreg ROC Curve
True Positive Rate (Positive label: 1)
False Positive Rate (Positive label: 1)
0.0
0.2
0.4
0.6
0.8
1.0
Classifier (AUC = 0.51)



Neural Network


Input Layer ∈ $\mathbb{R}^8$
Hidden Layer ∈ $\mathbb{R}^7$
Hidden Layer ∈ $\mathbb{R}^{15}$
Hidden Layer ∈ $\mathbb{R}^{11}$
Output Layer ∈ $\mathbb{R}^1$


NN ROC Curve
1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate (Positive label: 1)
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate (Positive label: 1)
Classifier (AUC = 0.91)



Input Layer ∈ $\mathbb{R}^8$
Hidden Layer ∈ $\mathbb{R}^7$
Hidden Layer ∈ $\mathbb{R}^{15}$
Hidden Layer ∈ $\mathbb{R}^{11}$
Output Layer ∈ $\mathbb{R}^1$

Next Step: Improvement

Model almost perfectly identifies cybersecurity threats but classifies authorized users as threats 18% of the time

Model is not production-ready, but still has the potential to be if we were to integrate more models into the process, but that takes time

Next steps would be integrating other models on top of the current working model in order to compensate for false positives

Thank You