IFT 520:Project Progress Report

Topic: Network Anomalies detection

Group members:

Saloni Mourya ASU Id: (1228595123)

Hetvi Patel ASU Id: (1230385431)

Mithul Sudharsan Ravikumar ASU Id: (1229035836)

Venkata Sai Chandra Sekhar Gudivada ASU Id: (1230548360)

Information Technology, Arizona State University

Dr. Jim Helm

October 29, 2023

Executive Summary:

The Network Anomaly Detection System project, which aims to leverage machine learning for enhanced network security, has reached the halfway point. This report provides a comprehensive overview of the achievements, challenges faced, and the roadmap for the remaining project timeline.

What have we done till now:

A. Data Collection and Preprocessing:

Our team has accomplished a significant milestone by procuring historical network traffic data from a wide array of sources, which encompass logs, flow data, and packet captures. To ensure the data is in prime condition for our project, we've meticulously undertaken data preprocessing tasks. This includes handling missing data values, normalizing features, and systematically labeling instances as either demonstrating normal or anomalous behavior. These critical preparatory steps lay the solid foundation upon which our machine learning models will be built. This phase is paramount as it sets the stage for training highly accurate machine learning models

B. Model Development and Training:

Our machine learning model development has advanced significantly. In particular, we have chosen to combine neural networks, isolation forests, and support vector machines (SVM). We have rigorously trained these models with our carefully labeled dataset. We've adjusted the model's hyperparameters, a process called as hyperparameter tuning, to make sure they function as best they can. Our preliminary evaluations of these models have shown encouraging results in terms of their capacity to discriminate patterns linked to typical network behavior.This crucial phase takes us one step closer to achieving our project's goal.

C. Real-Time Data Processing and Alert Generation:

We are actively working on the real-time data processing pipeline. This pipeline is designed to continuously monitor network flows and packets in real-time. Additionally, our alert generation mechanism is already operational. It plays a vital role in sending timely notifications as soon as anomalies are detected in the network. In the weeks ahead, we have user testing planned to evaluate how well our system performs in terms of responsiveness and alert accuracy. This user feedback will be invaluable in further refining and optimizing our real-time data processing and alert generation processes.

What is our plan for the future:

A. System Improvement Proposals:

As we continuously evaluate the performance of our system, our team is actively formulating proposals for system enhancements. The objective is to further elevate the system's accuracy and overall effectiveness. One of the key areas of focus involves fine-tuning the alert generation mechanism, making it even more responsive and precise in identifying anomalies. Additionally, we're delving into the exploration of additional machine learning models to bolster the system's robustness. This phase underscores our commitment to refining and optimizing the system as it evolves.

B. Continued Development:

The project will continue with a focus on enhancing the system's accuracy and effectiveness based on feedback and ongoing evaluations.

Challenges and issues encountered:

A. Data Quality Challenges:

In the journey of working with historical network data, our team encountered several data quality challenges. These hurdles prompted us to invest substantial effort in cleaning and preprocessing the data meticulously. We had to devise strategies for addressing missing values and identifying outliers to enhance the reliability of our dataset. Tackling these data quality challenges became a vital step in ensuring the overall success of our project.

B. Model Training Complexity:

The complexity associated with training our machine learning models presented a set of formidable challenges. This complexity was most pronounced when it came to optimizing hyperparameters for each algorithm. We found ourselves in a process of continuous refinement and experimentation, which was imperative to attain satisfactory model performance. Overcoming the complexities of model training has been an ongoing effort and a key aspect of our project's evolution.

C. Real-Time Processing Optimization:

Reducing latency and maximizing efficiency in the real-time data processing pipeline has proven to be challenging. In order to ensure that the system operates as responsively as feasible in practical scenarios, we're putting a lot of effort into streamlining and optimizing this pipeline. Optimizing the system's efficiency is about making sure it can process real-time data without any lag. This ongoing optimization work is critical to our project's success in real-world applications.

Final time table:

A. Remaining Milestones:

1. System Improvement Proposals (Next 2 Week):

- Collaboratively develop proposals for system improvements based on ongoing evaluations.

2. Testing and Feedback Collection (Next 1 Week):

- Conduct user testing to gather feedback on the system's responsiveness and alert accuracy.

3. Final Evaluation and Report Compilation (Next 1 Week):

- Conduct a final evaluation of the system's performance and compile a comprehensive project report summarizing the findings.

B. Project Completion (Dec/2023):

The project is slated for completion by the end of dec/2023, with the final deliverables including a fully functional Network Anomaly Detection System, documentation, and a detailed project report.