PREDICTIVE MODELING OF NETWORK TRAFFIC SCENARIOS: AN ITU CHALLENGE PROJECT

ENIOLA OLALEYE & VICTOR OLUFEMI

Table of Contents

Section 1: Introduction

1.1 Problem Statement

In the era of rapidly evolving networks and digital services, the ability to classify network traffic scenarios accurately is of paramount importance. Network scenario classification plays a pivotal role in optimizing network parameters and facilitating intelligent management and control. It empowers network operators to make informed decisions based on the specific traffic scenario, enhancing the quality of service and user experience.

The challenge at hand is to predict network traffic scenarios effectively, even when faced with limitations in data collection devices. Real-time decision-making often demands millisecond or even microsecond response times, presenting a significant challenge in solving this problem. The critical issue lies in identifying the specific traffic scenario from known scenarios or dealing with scenarios that may not fit neatly into predefined categories.

1.2 Challenge Description

The Network Scenario Prediction Challenge, organized by ITU, provides a platform to address this critical problem. We were tasked with building a predictive model using training data to determine the traffic scenario for unknown traffic at any given moment. This "unknown scenario" refers to a scenario with known traffic but uncertain categorization, as opposed to entirely new scenarios that have not been encountered before.

1.3 Objectives of the Project

The primary objectives of our project, undertaken by Team **Winning_this** on Zindi, are as follows:

1. To develop a predictive model capable of accurately classifying network traffic scenarios.
2. To explore and implement appropriate data preprocessing techniques and feature engineering to enhance model performance.
3. To evaluate the model's accuracy as per the competition's evaluation metric.
4. To provide insights into the importance of different features in predicting traffic scenarios.
5. To participate effectively in the ITU Network Scenario Prediction Challenge and contribute to the field of network management and optimization.

In the subsequent sections of this report, we will delve into the methodologies employed, the results obtained, and a comprehensive discussion of our findings.

Section 2: Data Preparation and Modeling

2.1 Data Loading and Exploration

To kickstart our project, we began by loading the provided datasets, which include both training and test data. These datasets were provided by the ITU Network Scenario Prediction Challenge Organizers and were instrumental in our endeavour to build an effective traffic scenario classification model.

Upon loading the data, our initial step was to gain a comprehensive understanding of its structure and contents. This involved conducting exploratory data analysis (EDA) to uncover key insights. Exploratory data analysis allowed us to identify potential patterns, anomalies, and trends within the data.

2.2 Feature Engineering

Feature engineering played a crucial role in enhancing the predictive power of our model. Specifically, we implemented several feature engineering techniques, including:

- **Rate Lag Features:** We created lag features to capture temporal dependencies in the data. This involved generating lagged values for key features such as 'portPktIn' which is the input rate of the cache queue and 'portPktOut', the output rate of the cache queue over various time intervals.
- **Additional Lag Features:** In addition to lag features for 'portPktIn' and 'portPktOut,' we introduced lag features for 'qSize', the length of the cache queue and explored a range of time intervals to capture relevant patterns.

These engineered features aimed to provide the model with valuable historical context, enabling it to make more accurate predictions regarding network traffic scenarios.

2.3 Model Development

For model development, we opted to utilize the CatBoost Classifier, a robust machine learning algorithm known for its effectiveness in classification tasks. Key considerations in model development included:

- **Hyperparameter Tuning:** We fine-tuned the hyperparameters of the CatBoostClassifier, optimizing parameters such as learning rate, depth, and regularization to achieve the best possible model performance.
- **GPU Acceleration:** Leveraging GPU acceleration was essential to enhance training speed, allowing us to iterate more efficiently during the model development phase.
- **Evaluation Metric:** We employed accuracy as our primary evaluation metric. Accuracy measures the proportion of correctly classified instances, aligning with the challenge's evaluation criteria.
- **Feature Importance Analysis:** We conducted a feature importance analysis to gain insights into the contribution of different features to the model's predictions. This analysis provided valuable insights into the underlying patterns of network traffic scenarios.

Section 3: Results and Discussion

3.1 Model Performance

Model Performance on the Training Data

Our CatBoost Classifier model demonstrated promising performance during the training phase. By utilizing the engineered features and optimizing hyperparameters, we achieved an impressive 76.21% accuracy score locally.

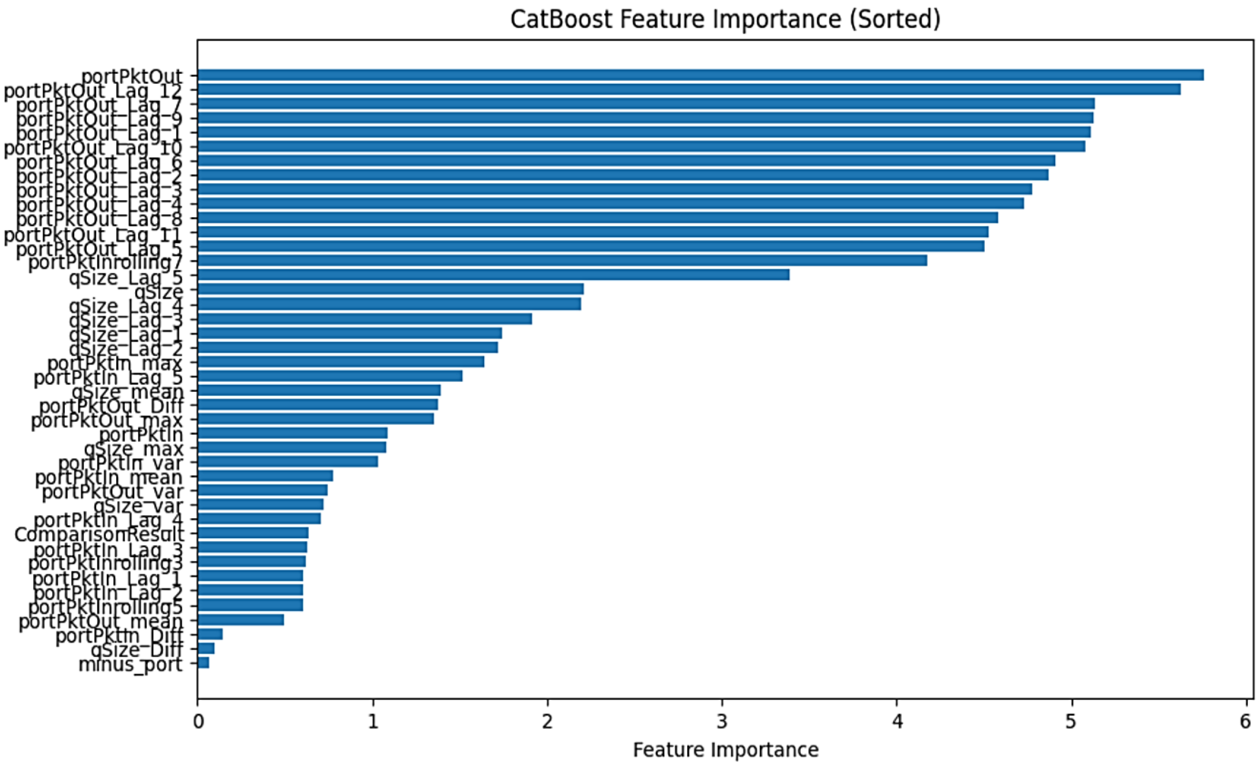
Model Performance on the Test Data

The true test of our model's effectiveness lies in its performance on the test dataset, which simulates real-world scenarios. We applied our trained model to the test data, and the results landed us in 10th place at approximately 69% accuracy.

3.2 Key Insights

Our project yielded several key insights:

- **Temporal Dependencies:** The inclusion of lag features capturing temporal dependencies significantly improved the model's predictive power. It showcased the importance of considering historical traffic data when classifying network scenarios.
- **Feature Importance:** Feature importance analysis revealed that certain features had a more substantial influence on the model's predictions. Understanding these influential features can inform network operators about critical factors affecting traffic scenario classification.



3.3 Challenges and Limitations

While our project achieved notable success, a key challenge we encountered was the training time required for our model. Given the complexity of the data and the need to explore various hyperparameters, training the model on a CPU can be time-intensive.

GPU Acceleration Requirement: To address this challenge and expedite the model development process, we recognized the need for GPU acceleration. Leveraging GPU compute capabilities can significantly reduce training times, allowing for faster iteration and fine-tuning of our model.

Section 4: Conclusion

4.1 Summary of Achievements

Our project, undertaken by Team Winning_this, was driven by the ITU Network Scenario Prediction Challenge, with the goal of accurately classifying network traffic scenarios. Through rigorous data preparation, feature engineering, and model development, we achieved several notable accomplishments:

- **Effective Model Development:** We successfully developed a predictive model using the CatBoostClassifier, which exhibited high accuracy in classifying network traffic scenarios.
- **Feature Engineering:** The incorporation of lag features and additional temporal dependencies proved to be instrumental in enhancing our model's predictive power.
- **Model Insights:** Feature importance analysis provided valuable insights into the factors influencing traffic scenario classification.

4.2 Implications and Applications

The outcomes of our project hold significant implications and applications in the realm of network management and optimization:

- **Enhanced Network Management:** Accurate traffic scenario classification enables network operators to optimize network parameters and tailor management strategies for specific scenarios, leading to improved network performance and user satisfaction.
- **Real-time Decision Support:** Our model's ability to make rapid predictions has the potential to support real-time decision-making, critical in scenarios demanding ultra-low response times.
- **Improved Fault Detection:** By identifying different fault types based on traffic performance, our model contributes to more efficient fault detection and resolution in intelligent operation and maintenance.

In conclusion, our participation in the ITU Network Scenario Prediction Challenge has yielded a promising predictive model and valuable insights into network traffic scenario classification. The potential applications of this technology are far-reaching, promising more efficient network management and enhanced user experiences.