Department of Computer Science & Engineering

United International University

CSE6011: Data Mining

Forecasting SMS Traffic and Balance Availability with Machine Learning

Author:

Mohammad Saifur Rahman

ID : 0122420002

Nirupom Das Dipto

ID : 0122230023

Md. Raqibur Rahman

ID : 0122310001

Md Mizanur Rahman

ID : 0122420016

Supervisor:

Dr. Mohammad Nurul Huda

Professor, Head of CSE Dept.

Department of CSE, UIU

October 2024

Contents

Chapter 1

Introduction

This report presents a comprehensive analysis of SMS traffic and balance availability for a specific set of clients, with a particular focus on predicting potential service interruptions during holidays. By utilizing advanced machine learning techniques, such as **polynomial regression**, we aim to accurately forecast future SMS traffic and balance levels, enabling clients to optimize their usage and avoid unexpected service disruptions.

By using advanced data analytics, this study provides a comprehensive forecast of SMS traffic and balance levels, enabling organizations to proactively manage their client communication strategies and resource allocation. This predictive model accurately anticipates SMS traffic during peak periods, such as holidays, and offers hourly updates on balance levels for the next two days, ensuring a seamless and uninterrupted service experience for clients.

Ultimately, this research contributes to the field of machine learning applications and provides practical solutions for organizations seeking to effectively manage their SMS balance allocation to clients, particularly during peak demand periods like holidays.

1.1 Background

At Wintel Limited, there was a need for SMS balance forecasting analysis to predict SMS traffic and balance levels in advance. By anticipating these factors, the company can avoid SMS shortages during peak times, such as holidays or unexpected events, ensuring continuous service availability.

1.1.1 The Problem

In today's fast-paced world, SMS communication has become an integral part of our daily lives. For organizations with a large customer base, managing SMS traffic and ensuring adequate balance availability can be a complex challenge, especially during peak periods such as holidays. Service interruptions due to insufficient balance or overloaded networks can lead to customer dissatisfaction and financial losses.

1.1.2 The Solution

To address these challenges, this report presents a comprehensive analysis of SMS traffic and balance prediction & automation for a specific set of clients. By employing advanced machine learning techniques, such as **polynomial regression**, we aim to develop a predictive model that can accurately forecast future SMS traffic and balance levels. This information will empower organizations to make informed decisions regarding their communication strategies and resource allocation, ensuring a seamless and uninterrupted service experience for their clients.

1.1.3 The Benefits

- **Improved Service Quality:** Accurate predictions of SMS traffic and balance levels will enable organizations to proactively address potential issues, ensuring a consistent and reliable service experience for their clients.
- **Optimized Resource Allocation:** By understanding future demand, organizations can allocate resources more efficiently, reducing employees hours of struggle during the holidays and minimizing service disruptions.
- **Enhanced Customer Satisfaction:** A reliable and uninterrupted SMS service can significantly improve customer satisfaction and loyalty.
- **Data-Driven Decision Making:** The insights gained from this analysis will provide organizations with a data-driven foundation for making informed decisions about their SMS communication strategies.

This report aims to contribute to the field of machine learning applications and provide practical solutions for organizations seeking to optimize their SMS communication infrastructure and deliver exceptional customer service.

1.2 Literature Review

1.3 Research Gap

1.4 Objectives

- The objectives of this paper is to develop a predictive model using advanced machine learning techniques to accurately forecast masking and non-masking SMS traffic and balance levels for a specific set of clients. More specifically the purpose of this project is generate next 2 days predictions.
- To identify potential service interruptions during peak periods, such as holidays, and provide early warnings to organizations and it's clients.
- To enable organizations to optimize their SMS communication strategies and resource allocation by providing timely information on SMS traffic and balance levels.
- To contribute to the field of machine learning applications by demonstrating the effectiveness of polynomial regression in predicting SMS traffic and balance.
- To provide practical solutions for organizations seeking to improve their SMS communication infrastructure and deliver exceptional customer service.

Chapter 2

Methodology

2.1 Corpus Collection

2.1.1 Description of the data sources used

For this project, data from Wintel Limited is used. To achieve better results and improve model building, a real-life, real-time dataset will be used, ensuring more accurate and practical outcomes.

2.1.2 Data Collection Process

In wintel each 4 hours a cron/scheduler will be called where this model will work and generate the predictions. For Both masking and non-masking sms traffic and balance seperate models will be called. In each day approximately 6 times the this process will work. We will pick last 180 days dataset from wintext portal mssql database and use it to the model.

2.1.3 Data preprocessing techniques

- Timestamps were converted into a consistent datetime format to facilitate accurate time-series analysis.
- SMS traffic counts and balance levels were ensured to be in numeric formats for proper calculation and analysis.

- Duplicate records in SMS traffic logs were identified by checking for identical timestamps, client IDs, and traffic counts.

- Special care was taken to adjust for traffic and balance patterns during holidays, which often exhibit different behaviors from regular periods. Historical holiday data was cross-referenced with the current dataset to ensure accurate classification and analysis.

- SMS Traffic per Operator & Client: The SMS traffic per operator and client was calculated to analyze usage patterns.

- Balance per Client: The balance per client was calculated to assess balance allocation.

- Balance per Operator: The balance per operator was calculated to assess balance allocation to opertor wise.

2.1.4 Training Data

We collected the daily operator balance data spanning over the last 6 months (180 days) from the `ds_dly_operator_balance` table, which includes the fields: operator ID, available SMS balance (masking and non-masking), date, and day number. Before training, missing balance values were filled with the median balance for each operator, and features were standardized to ensure that each operator's data contributed equally to the model.

2.1.5 Test Data

For testing, we reserved the most recent data (from the last 7 days) to measure the accuracy of the SMS balance forecast. The test dataset contains unseen operator balances, and we used RMSE to evaluate the prediction accuracy.

2.1.6 Cross-Validation

To avoid overfitting and ensure model generalization, we applied 5-fold cross-validation. The dataset was split into 5 equal parts, with each part used as a test set while the remaining

data served as the training set. This process was repeated 5 times, and the average RMSE across all folds was used to evaluate model performance.

Chapter 3

Experimental Results and Analysis

Chapter 4

Conclusion and Future Works

4.1 Conclusion

This report presents a comprehensive study on forecasting SMS traffic and balance availability using machine learning techniques. By leveraging historical data from Wintel Limited, we developed a predictive model capable of providing accurate forecasts for SMS traffic and balance levels over the next two days. This model has demonstrated its ability to anticipate potential service interruptions, particularly during holidays and other peak periods, ensuring clients can optimize their SMS usage and avoid disruptions.
Key benefits of this work include:

- Improved service reliability during peak traffic periods.
- Enhanced resource allocation based on accurate demand forecasts.
- Increased customer satisfaction through proactive management of SMS traffic and balances.

Overall, this project contributes to the practical application of machine learning in managing communication resources, especially in time-sensitive environments where uninterrupted service is critical.

4.2 Future Works

While the current model demonstrates strong predictive capabilities, several areas for future research and improvement have been identified:

- **Model Optimization:** Future iterations of the model can explore the use of advanced machine learning techniques, such as deep learning or ensemble models, to improve prediction accuracy, especially for more complex traffic patterns. Also we don't have currently all 180 days real dataset to actually verify original outcomes. After getting 180 days dataset we can verify the model more accurately.

- **Incorporation of Additional Features:** Additional factors, such as external events, marketing campaigns, or changes in client behavior, can be incorporated into the model to refine its forecasts further.

- **Real-Time Prediction:** While this report focused on hourly predictions, the model can be extended to provide real-time updates, allowing for even more dynamic resource management.

- **Scalability:** Further research is needed to test the scalability of the model across a broader set of clients and operators, particularly for large-scale applications where SMS traffic volume is significantly higher.

- **Exploring Different Algorithms:** Polynomial regression was used for this project, but future work could evaluate the performance of alternative algorithms, such as decision trees, random forests, or gradient boosting methods, to find the best fit for the problem domain.

- **Integration with Automated Systems:** Finally, future work can focus on integrating the predictive model into automated systems that can automatically allocate resources or send alerts to clients when balances are predicted to drop below a critical threshold.

By addressing these areas, we can further enhance the reliability and robustness of SMS traffic and balance prediction systems, ultimately improving client experiences and operational efficiency.