**Ramazan Alisher**BDA-2302

**Title: PREDICTING BUS ARRIVAL TIMES USING MACHINE LEARNING MODELS**

**Abstract:** This research paper explores the application of machine learning models to predict bus arrival times in Astana, Kazakhstan. Accurate bus arrival predictions are crucial for improving public transportation efficiency and passenger satisfaction. We investigate the performance of various models, including K-means clustering, K-Nearest Neighbors (KNN), Convolutional Long Short-Term Memory (Conv-LSTM), Support Vector Machines (SVM), and Linear/Logistic Regression. Our study utilizes real-world bus data from Astana's public transportation system, considering factors such as time of day, day of the week, route, and traffic conditions. We evaluate the models based on metrics like mean absolute error and root mean squared error to identify the most effective approach for bus arrival prediction in Astana.

**Keywords:** Machine learning, bus arrival prediction, public transportation, K-means clustering, KNN, Conv-LSTM, SVM, linear regression, logistic regression

**Introduction**

Efficient public transportation is essential for the smooth functioning of modern cities. Accurate bus arrival predictions play a vital role in enhancing the passenger experience and encouraging the use of public transport. In this study, we focus on Astana's public transportation system and aim to develop machine learning models that can reliably predict bus arrival times. By providing passengers with accurate information, we can reduce waiting times, improve route planning, and optimize the overall efficiency of the bus network.

**Background**

Previous research has explored various techniques for bus arrival prediction, including statistical models, time series analysis, and machine learning approaches. Machine learning has shown great potential in capturing complex patterns and relationships in transportation data. Several studies have utilized models like KNN, SVM, and neural networks for bus arrival prediction. However, the performance of these models can vary depending on the specific characteristics of the transportation system and the data available.

**Methodology**

**Data Collection:**

We collected real-world bus data from Astana's public transportation system, including information on bus routes, schedules, timestamps of bus arrivals, and GPS coordinates. Additionally, we incorporated external data sources such as traffic data and weather information to enhance the prediction accuracy.

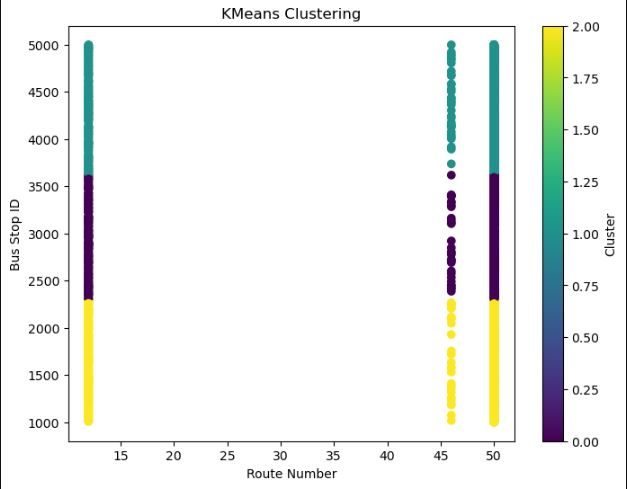
**Feature Engineering:**

We extracted relevant features from the raw data, such as time of day, day of the week, route number, bus stop ID, and traffic conditions. These features were used as input variables for our machine learning models.

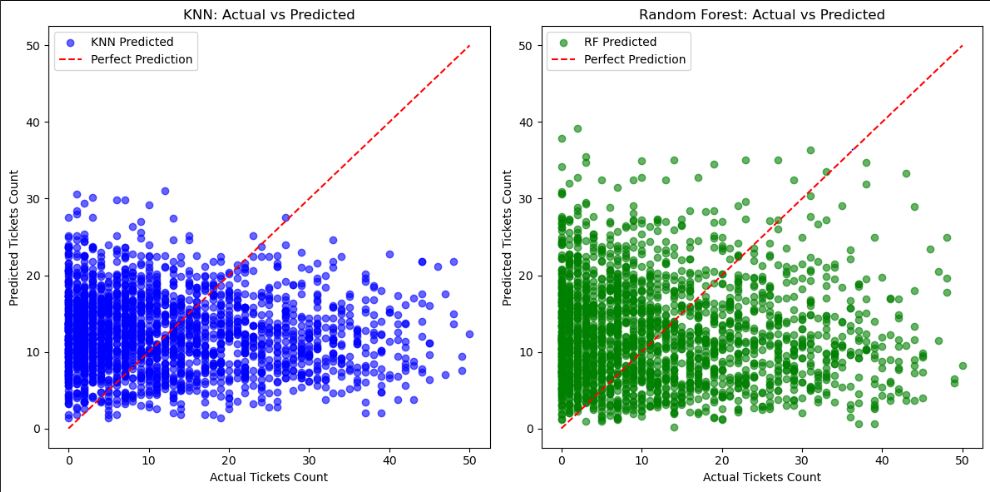
**Model Development:**

We developed and evaluated the following machine learning models:

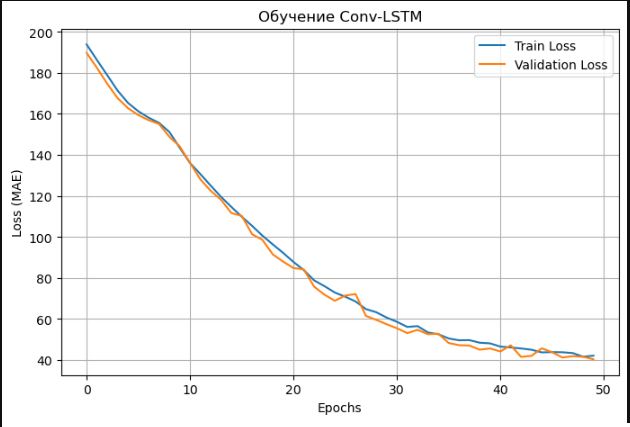
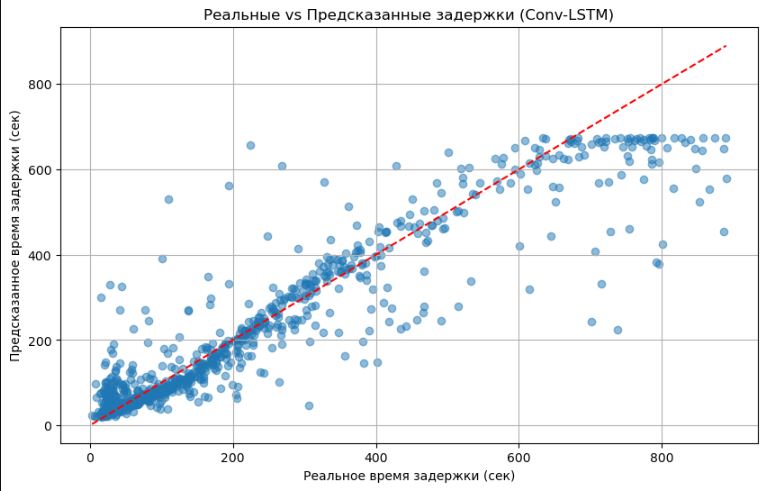
**K-means Clustering:** We applied K-means clustering to group bus stops based on their historical arrival patterns. This clustering helped in identifying similar bus stops and improving prediction accuracy.



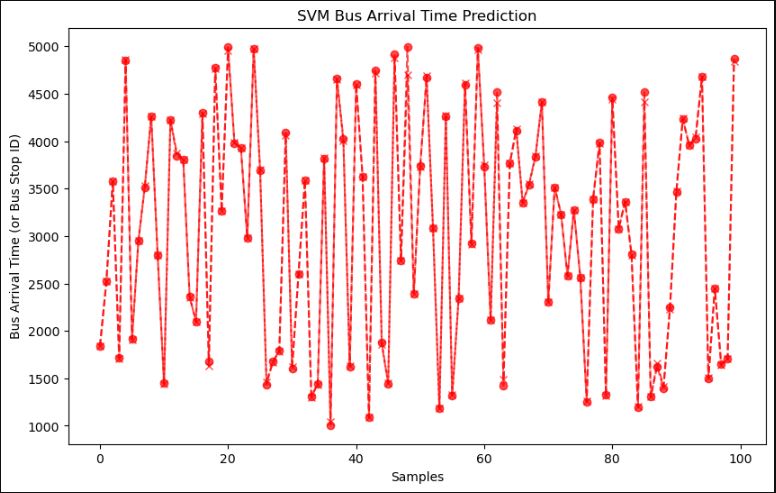
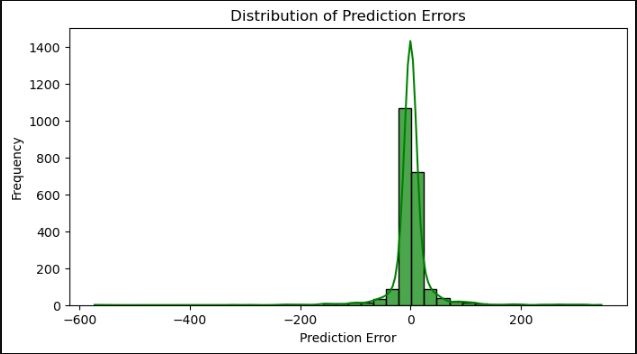
**K-Nearest Neighbors (KNN):** We used KNN to predict bus arrival times based on the arrival times of similar buses in the past.



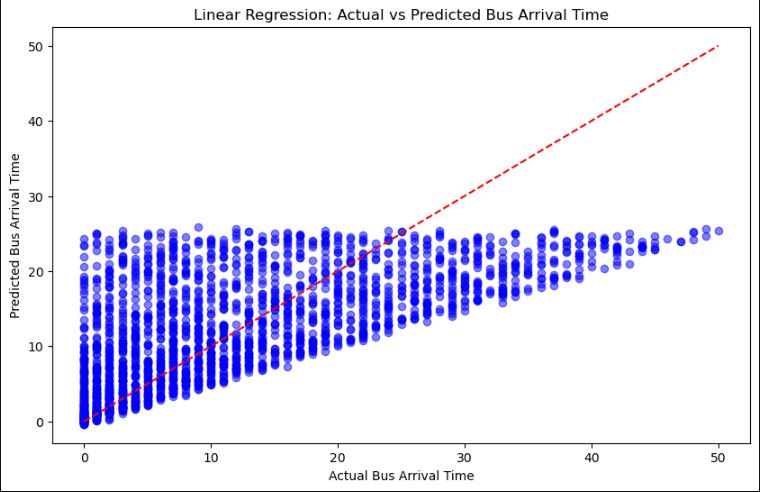
**Convolutional Long Short-Term Memory (Conv-LSTM):** We employed Conv-LSTM networks to capture temporal dependencies in bus arrival data and predict future arrival times.

****

**Support Vector Machines (SVM):** We trained SVM models to predict bus arrival times by finding the optimal hyperplane that separates different arrival time interval.



**Linear/Logistic Regression:** We utilized linear regression to model the relationship between bus arrival times and various features. Logistic regression was used for predicting categorical arrival time intervals.

****

**Model Evaluation:**

We evaluated the performance of the models using metrics like mean absolute error (MAE) and root mean squared error (RMSE). These metrics measure the difference between the predicted and actual bus arrival times.

**Results:**

We present the results of our experiments, comparing the performance of the different machine learning models. We analyze the impact of various factors on prediction accuracy and identify the most effective model for bus arrival prediction in Astana.

**Conclusion:**

This study demonstrates the potential of machine learning models for predicting bus arrival times in Astana. Our findings highlight the importance of feature engineering and model selection in achieving accurate predictions. The results of this research can be used to develop a real-time bus arrival prediction system, improving the efficiency and passenger satisfaction of Astana's public transportation.