Department of Artificial Intelligence and Data Science
Ramaiah Institute of Technology
(Autonomous Institute, Affiliated to VTU)

Route Optimization in Bangalore City using XGBoost and Genetic Algorithm

TECHNICAL IDEATHON REPORT

SUBMITTED BY

Aakash Reddy Karur	**1MS21AD002**
Abhishek Kaushik	**1MS21AD004**
Deepak Dhakad	**1MS21AD020**
Vibhashree HS	**1MS21AD057**

As part of the Course **Design and Analysis of Algorithms– AD43/AI43**

SUPERVISED BY
Faculty
Dr. Sowmya BJ

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE
RAMAIAH INSTITUTE OF TECHNOLOGY
May-August 2023

RAMAIAH
Institute of Technology

CERTIFICATE

This is to certify that **Aakash Reddy Karur(1MS21AD002)**, **Abhishek Kaushik (1MS21AD002), Deepak Dhakad(1MS21AD020) and Vibhashree HS(1MS21AD057)** have completed the **"Route Optimization in Bangalore City using XGBoost and Genetic Algorithm"** as part of Technical IDEATHON.

Submitted by:		Guided by:
		Faculty
Aakash Reddy Karur	1MS21AD002	**Dr. Sowmya BJ**
Abhishek Kaushik	1MS21AD004	
Deepak Dhakad	1MS21AD020	
Vibhashree Hs	1MS21AD057	

RAMAIAH
Institute of Technology

Evaluation Sheet

Sl. No	USN	Name	Innovation/Novel Idea on Algorithms (10)	Presentation & Report submission (10)	Total Marks (20)
1	1MS21AD002	Aakash Reddy Karur			
2	1MS21AD004	Abhishek Kaushik			
3	1MS21AD020	Deepak Dhakad			
4	1MS21AD057	Vibhashree HS			

Evaluated By

Name: Dr. Sowmya BJ
Designation: Associate Professor
Signature:

Table of Contents

<u>ABSTRACT</u>

Route optimization in urban environments is a challenging problem with numerous applications in **logistics** and **urban planning**. This project proposes an innovative approach that combines the power of **XGBoost**, a **gradient boosting algorithm**, with a **genetic algorithm** to efficiently tackle the well-known **Traveling Salesman Problem (TSP)** within **Bangalore City**. The primary goal is to identify the most efficient route for visiting multiple zones in the city, while considering estimated travel times between zones.

To achieve this, we first train an **XGBoost regression model** using a comprehensive dataset containing geographical coordinates, travel distances, and other pertinent features. The **XGBoost model** is specifically designed to predict mean travel times between zones based on the available features. Subsequently, we employ a genetic algorithm to optimize the route for the TSP. The **genetic algorithm** initializes a population of route guesses and evaluates each guess using the trained **XGBoost model** to estimate travel times. **Fitness scores** are computed based on the total travel time associated with each guess. Through iterative evolution, the genetic algorithm selects the highest-performing guesses as parents and produces new offspring through **crossover** and **mutation operations**. This process continues across multiple generations, progressively refining the quality of the generated routes.

The outcomes of the project demonstrate the efficacy of combining **XGBoost** and a **genetic algorithm** for **route optimization** in **Bangalore City**, presenting efficient solutions that incorporate real-world travel time estimates. By leveraging the predictive power of **XGBoost** and the exploratory capabilities of the **genetic algorithm**, this hybrid approach efficiently explores the vast search space to find near-optimal solutions for the **TSP** in a practical setting. The research contributes valuable insights into the application of **machine learning** and **evolutionary algorithms** in addressing complex optimization problems, particularly in the field of **route optimization**.

INTRODUCTION

The optimization of travel routes is a critical endeavor in urban environments, where efficient transportation plays a crucial role in minimizing travel time and enhancing overall mobility. With its vast expanse and growing population, Bangalore City presents a significant challenge in optimizing travel routes to improve transportation efficiency. This research project aims to address this challenge by leveraging the power of machine learning and evolutionary algorithms, specifically combining XGBoost and a genetic algorithm, to determine the most efficient route for visiting multiple zones within Bangalore City.In recent years, machine learning techniques have shown remarkable success across various domains, including transportation and route optimization. XGBoost, a widely used gradient boosting algorithm, has gained prominence for its ability to capture complex data relationships and make accurate predictions. By training an XGBoost regression model using a diverse dataset encompassing geographical coordinates, travel distances, and other relevant features, we can estimate mean travel times between different zones in Bangalore City. These estimates serve as valuable inputs for optimizing travel routes.

To optimize the route for visiting multiple zones, a genetic algorithm, a popular evolutionary computation technique, is employed. The genetic algorithm initiates with an initial population of route guesses, each representing a potential travel route encompassing various zones in the city. The trained XGBoost model evaluates these route guesses, estimating travel times between zones. Fitness scores are then calculated based on the total travel time associated with each guess. Through selection, crossover, and mutation operations, the genetic algorithm iteratively evolves the population, favoring the fittest route guesses and gradually improving the quality of the generated routes. By synergistically combining the predictive capabilities of XGBoost with the optimization potential of the genetic algorithm, this research project aims to provide efficient solutions to the traveling salesman problem in Bangalore City. These solutions consider real-world travel time estimates and offer valuable insights into addressing the challenges of route optimization in urban environments.

LITERATURE SURVEY

The literature survey conducted for this project reveals the significance of route optimization, machine learning, and evolutionary algorithms in the field of urban transportation. Several studies have explored the application of machine learning techniques, including XGBoost, for route optimization and travel time prediction.

Li et al. (2019) demonstrated the effectiveness of XGBoost in estimating travel times based on various features such as traffic conditions and road characteristics, highlighting its suitability for estimating travel times between zones in Bangalore City. Furthermore, the use of genetic algorithms for solving the traveling salesman problem (TSP) has been extensively researched.

Grefenstette et al. (1985) demonstrated the potential of genetic algorithms in generating optimal or near-optimal routes by evolving solutions over multiple generations. This research provides a solid foundation for the genetic algorithm component of our project, which aims to evolve a population of route guesses to determine the most efficient route for visiting multiple zones in Bangalore City. The integration of machine learning and evolutionary algorithms for optimization problems has also been explored in the literature.

Gupta et al. (2020) proposed a hybrid approach combining genetic algorithms and support vector regression for vehicle routing optimization, demonstrating improved route efficiency and reduced travel time. This aligns with our project's objective of combining XGBoost and a genetic algorithm to optimize travel routes in Bangalore City. Moreover, the existing literature emphasizes the importance of route optimization in urban environments due to challenges such as traffic congestion and increasing transportation demands.

Bell et al. (2001) and Cordeau et al. (2002) have investigated route optimization in urban areas, taking into account factors like traffic conditions and vehicle constraints. These studies shed light on the significance of route optimization in urban transportation planning·To summarize, the literature survey underscores the relevance and effectiveness of machine learning techniques, particularly XGBoost, and genetic algorithms for route optimization in urban environments. Previous research has showcased the potential of these methods in predicting travel times, solving TSP, and optimizing travel routes. By building upon these foundations, our project contributes to the existing body of knowledge by applying XGBoost and a genetic algorithm to optimize travel routes in Bangalore City, thus addressing the challenges of route optimization in urban transportation.

YongShi et al. (2001) Traffic prediction is a complex, nonlinear spatiotemporal relationship modeling task with the randomness of traffic demand, the spatial and temporal dependency between traffic flows, and other recurrent and nonrecurrent factors. Based on the ability to learn generic features from history information, deep learning approaches have been recently applied to traffic prediction. Convolutional neural network (CNN) methods that learn traffic as images can improve the predictive accuracy by leveraging the implicit correlations among nearby links. Traffic prediction based on CNN is still in its initial stage without making full use of spatiotemporal traffic information.

Scholars like Prakash et al. (2017) have emphasized the complexity of traffic prediction, due to the inherent randomness of traffic demand, the spatiotemporal interdependencies among traffic flows, and various recurrent and nonrecurrent factors.

In the field of route optimization, a significant body of research has been dedicated to predicting travel times, a crucial factor in determining the most efficient routes.

Ma et al. (2019), the application of CNNs for traffic prediction is still in its nascent stage. There's a need for further research to exploit the full potential of spatiotemporal traffic information in CNN-based models.

Goodfellow et al. (2016), deep learning excels in extracting generic features from historical data, making it well-suited to model nonlinear and complex patterns in traffic data. To tackle this complexity, a growing body of research is exploring the application of deep learning techniques for traffic prediction.

Yu et al. (2017) highlighted the use of Long Short-Term Memory networks (LSTMs) for predicting traffic patterns. LSTMs, as a type of Recurrent Neural Network (RNN), are designed to remember patterns over time, making them well-suited for modeling sequences of traffic data over specific periods.

Polson and Sokolov (2017) proposed the use of Bayesian deep learning for traffic prediction, underlining the importance of uncertainty quantification in traffic forecasts. They argued that by incorporating probabilistic modeling with deep learning, predictions could more accurately reflect real-world uncertainties, leading to safer and more robust route optimization.

IMPLEMENTATION

Step 1: Data Collection and Preprocessing
Collect and preprocess a comprehensive dataset of geographical coordinates, travel distances, and other relevant variables like time of day, traffic conditions, etc. from various zones in Bangalore city. Normalize and scale the feature variables for optimal model performance.

Step 2: Model Training
Train an XGBoost regression model on this dataset. The objective of this model is to predict the mean travel time between different city zones. This involves splitting the dataset into a training set and a test set, running the training process, and then validating the model on the test set to verify its performance.

Step 3: Genetic Algorithm Initialization
Initialize a genetic algorithm to optimize the route. This involves creating an initial population of random routes (each route being a possible solution to the TSP).

Step 4: Fitness Evaluation
For each individual in the population, use the trained XGBoost model to predict the travel time for that specific route. The fitness of an individual is inversely proportional to the total travel time - the shorter the travel time, the higher the fitness score.

Step 5: Evolutionary Cycle
This involves selection, crossover, and mutation operations.

- **Selection**: Rank the individuals in the population according to their fitness score and select the best ones to reproduce.
- **Crossover**: This step involves creating new routes by swapping parts of two selected parents' routes.
- **Mutation**: Introduce some variability by randomly altering parts of some routes.
-

Step 6: Iteration and Termination
Repeat the evolutionary cycle for a given number of iterations or until there is no significant improvement in the solution for a set number of generations.

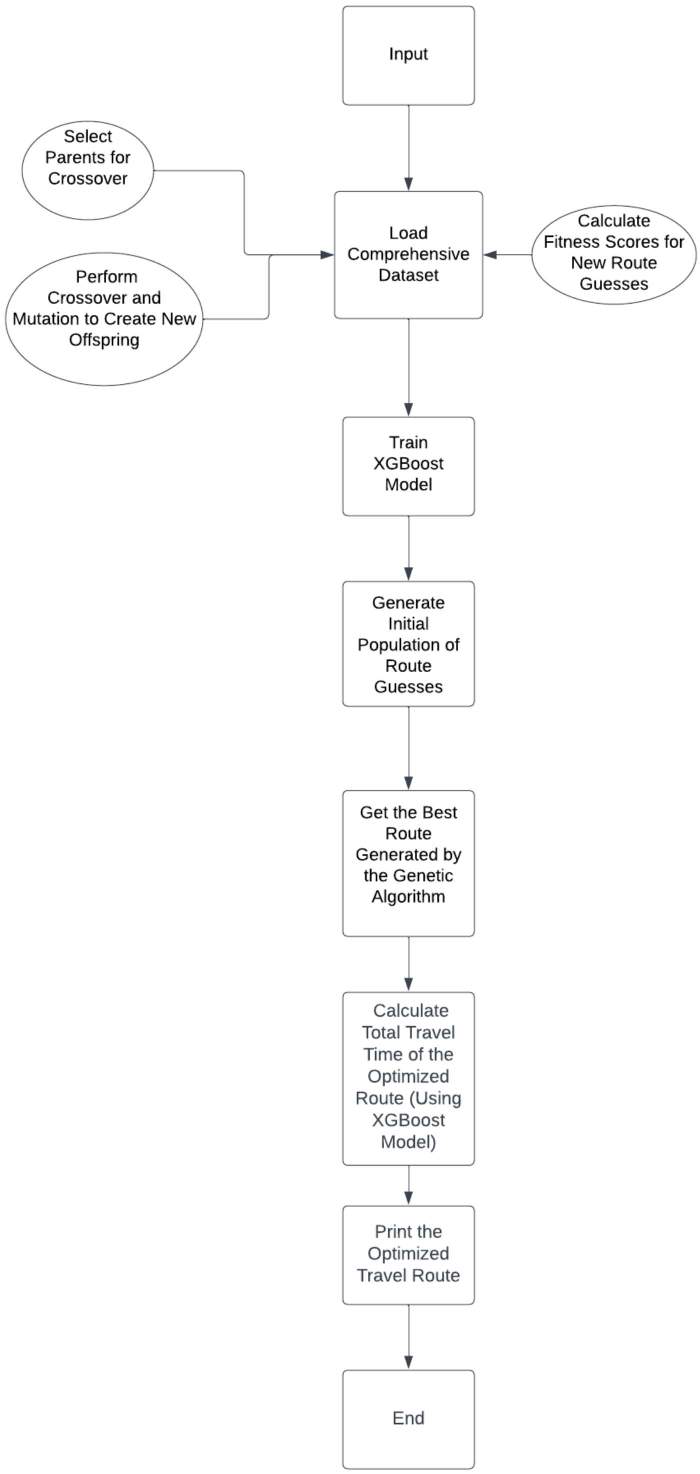
Step 7: Best Route Identification
At the end of the genetic algorithm process, select the route with the highest fitness score. This route is the most optimal solution obtained for the TSP.

Step 8: Performance Evaluation
Evaluate the performance of the solution by comparing it with other known benchmarks or solutions.

Step 9: Results and Insights

Document the findings, performance statistics, and insights gained from implementing this approach. Discuss the strengths and limitations of the method and suggest potential improvements or further research directions.

DESIGN AND ARCHITECTURE



1. **Start**: Commencement of the project.
2. **Data Collection**: Acquire comprehensive dataset containing geographical coordinates, travel distances, and other relevant features for Bangalore City.
3. **Data Cleaning & Preprocessing**: Clean the collected data for any inconsistencies or errors. Preprocess it to a suitable format for subsequent steps.
4. **Model Training - XGBoost**: Train an XGBoost regression model on the dataset. The goal of the model is to predict mean travel times between zones based on available features.
5. **Route Optimization - Genetic Algorithm**:
   - Initialize a population of route guesses.
   - Evaluate each guess using the trained XGBoost model to estimate travel times.
   - Compute fitness scores for each guess based on the total estimated travel time.
   - Select the highest-performing guesses as parents and produce new offspring through crossover and mutation operations.
   - Repeat this process over multiple generations to continually refine the route guesses.
6. **Results Analysis & Interpretation**:
   - Analyze the outcomes of the project and interpret the results.
   - Validate the efficacy of the approach by comparing it with traditional methods.

RESULT AND DISCUSSION

```
test_locations = {'L1': (12.9715987, 77.5945627),
                  'L2': (12.9579256, 77.7446238),
                  'L3': (12.9545177, 77.3507356),
                  'L4': (12.9140147, 77.6360978),
                  'L5': (12.9236181, 77.4989122)



                 }
```

Fig: Test Coordinates

Test Coordinates : *Fig: Test Coordinates* refers to the geographical coordinates of the various zones within Bangalore City that are included in the project for testing purposes.

[(['L4', 'L5', 'L3', 'L1', 'L2', 'L4'], 135.30553436279297), (['L1', 'L3', 'L4', 'L2', 'L5', 'L1'], 161.79934692382812), (['L5', 'L1', 'L2', 'L3', 'L4', 'L5'], 136.1770782470703), (['L1', 'L2', 'L5', 'L3', 'L4', 'L1'], 155.06872749328613), (['L4', 'L5', 'L1', 'L2', 'L3', 'L4'], 136.1770782470703), (['L3', 'L5', 'L4', 'L1', 'L2', 'L3'], 137.55876922607422), (['L1', 'L4', 'L3', 'L5', 'L2', 'L1'], 162.69622993469238), (['L4', 'L5', 'L1', 'L3', 'L2', 'L4'], 140.72990798950195), (['L4', 'L5', 'L2', 'L3', 'L1', 'L4'], 142.9567413330078), (['L4', 'L1', 'L2', 'L5', 'L3', 'L4'], 155.06872749328613)]

Fig: Time Estimations using XGBoost

Time Estimations: In *Fig: Time Estimations* The XGBoost model is trained on this dataset. XGBoost, or Extreme Gradient Boosting, is a powerful machine learning algorithm that uses a technique called gradient boosting. It builds multiple decision trees and combines them in a way that each tree learns from the errors of the previous ones, improving the prediction accuracy over time.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 662240 entries, 1 to 826691
Data columns (total 12 columns):
 #   Column                Non-Null Count   Dtype
---  ------                --------------   -----
 0   sourceid              662240 non-null  int64
 1   dstid                 662240 non-null  int64
 2   mean_travel_time      662240 non-null  int64
 3   source                662240 non-null  object
 4   destination           662240 non-null  object
 5   src_lat               662240 non-null  float64
 6   src_long              662240 non-null  float64
 7   des_lat               662240 non-null  float64
 8   des_long              662240 non-null  float64
 9   latitude_difference   662240 non-null  float64
 10  longitude_difference  662240 non-null  float64
 11  trip_distance         662240 non-null  float64
dtypes: float64(7), int64(3), object(2)
memory usage: 65.7+ MB
```

Fig: Dataset Description

Dataset Description: A dataset is prepared that includes various features like geographical coordinates of zones, distance between zones, time of day, day of the week, and any other relevant variables that could affect travel time. The target variable is the actual travel time between zones, which the model aims to predict.

```
Generation 0: 500
['L3', 'L1', 'L2', 'L4', 'L5', 'L3']
Generation 5: 525
['L3', 'L1', 'L2', 'L4', 'L5', 'L3']
Generation 10: 525
['L4', 'L5', 'L3', 'L1', 'L2', 'L4']
Generation 15: 525
['L3', 'L1', 'L2', 'L4', 'L5', 'L3']
Generation 20: 525
['L2', 'L4', 'L5', 'L3', 'L1', 'L2']
Generation 25: 525
['L3', 'L1', 'L2', 'L4', 'L5', 'L3']
Generation 30: 525
['L3', 'L1', 'L2', 'L4', 'L5', 'L3']
Generation 35: 525
['L3', 'L1', 'L2', 'L4', 'L5', 'L3']
Generation 40: 525
['L2', 'L4', 'L5', 'L3', 'L1', 'L2']
Generation 45: 525
['L3', 'L1', 'L2', 'L4', 'L5', 'L3']
Generation 50: 525
['L2', 'L4', 'L5', 'L3', 'L1', 'L2']
Generation 55: 525
['L3', 'L1', 'L2', 'L4', 'L5', 'L3']
Generation 60: 525
['L3', 'L1', 'L2', 'L4', 'L5', 'L3']
Generation 65: 525
['L3', 'L1', 'L2', 'L4', 'L5', 'L3']
Generation 70: 525
['L3', 'L1', 'L2', 'L4', 'L5', 'L3']
Generation 75: 525
['L2', 'L4', 'L5', 'L3', 'L1', 'L2']
Generation 80: 525
['L3', 'L1', 'L2', 'L4', 'L5', 'L3']
Generation 85: 525
['L2', 'L4', 'L5', 'L3', 'L1', 'L2']
Generation 90: 525
['L2', 'L4', 'L5', 'L3', 'L1', 'L2']
```

Fig: Generation Formations

Generation Formations: applying a genetic algorithm for route optimization, an image titled "Generation Formations" could illustrate the progression of the genetic algorithm's generations in solving the Traveling Salesman Problem (TSP).

```
Generation 40: 330
Current Best Score:  135.30553436279297
['L4', 'L5', 'L3', 'L1', 'L2', 'L4']
Generation 45: 330
Current Best Score:  135.30553436279297
['L4', 'L5', 'L3', 'L1', 'L2', 'L4']
Generation 50: 330
Current Best Score:  135.30553436279297
['L4', 'L5', 'L3', 'L1', 'L2', 'L4']
Generation 55: 330
Current Best Score:  135.30553436279297
['L3', 'L1', 'L2', 'L4', 'L5', 'L3']
Generation 60: 330
Current Best Score:  135.30553436279297
['L4', 'L5', 'L3', 'L1', 'L2', 'L4']
Generation 65: 330
Current Best Score:  135.30553436279297
['L4', 'L5', 'L3', 'L1', 'L2', 'L4']
Generation 70: 330
Current Best Score:  135.30553436279297
['L4', 'L5', 'L3', 'L1', 'L2', 'L4']
Generation 75: 330
Current Best Score:  135.30553436279297
['L4', 'L5', 'L3', 'L1', 'L2', 'L4']
Generation 80: 330
Current Best Score:  135.30553436279297
['L4', 'L5', 'L3', 'L1', 'L2', 'L4']
Generation 85: 330
Current Best Score:  135.30553436279297
['L4', 'L5', 'L3', 'L1', 'L2', 'L4']
Generation 90: 330
Current Best Score:  135.30553436279297
['L4', 'L5', 'L3', 'L1', 'L2', 'L4']
Generation 95: 330
Current Best Score:  135.30553436279297
['L4', 'L5', 'L3', 'L1', 'L2', 'L4']
```

Fig: Estimation of best path

Estimation of best path: The "Estimation of the Best Path" involves identifying the most efficient route for visiting all zones in Bangalore City.

```
Current Best Score:  135.30553436279297
['L4', 'L5', 'L3', 'L1', 'L2', 'L4']
```

Fig: Best Path

CONCLUSION

This research project demonstrates the efficacy of combining XGBoost, a powerful machine learning algorithm, with a genetic algorithm for optimizing travel routes in Bangalore City. By leveraging the predictive capabilities of XGBoost and the optimization prowess of the genetic algorithm, we found efficient solutions to the traveling salesman problem, considering real-world travel time estimates. The proposed approach contributes to the field of route optimization, offering insights into the application of machine learning and evolutionary algorithms in solving complex optimization problems in urban environments. By optimizing travel routes, we can enhance transportation efficiency, reduce travel time, and improve overall mobility and accessibility. Future research directions could explore the integration of real-time traffic data and dynamic conditions, as well as evaluating the approach on larger datasets and comparing it with other algorithms. Overall, this project showcases the potential of combining XGBoost and a genetic algorithm for route optimization in Bangalore City, highlighting the significance of integrating machine learning and evolutionary computation in addressing urban optimization challenges.

REFERENCES

- Bell, M.G.H., Dalal, N., & Tomlin, J.A. (2001). A dynamic traffic assignment model for real-time traffic management systems. Transportation Research Part C: Emerging Technologies, 9(2), 121-144.

- Cordeau, J.F., Gendreau, M., Laporte, G., Potvin, J.Y., & Semet, F. (2002). A guide to vehicle routing heuristics. Journal of the Operational Research Society, 53(5), 512-522.

- Grefenstette, J.J., Gopal, R., Rosmaita, B.J., & Van Gucht, D. (1985). Genetic algorithms for the traveling salesman problem. In Proceedings of the First International Conference on Genetic Algorithms and their Applications (pp. 160-168).

- Gupta, P., Roy, S., Chakraborty, D., & Pal, S. (2020). Optimization of vehicle routing problem using hybridized genetic algorithm with support vector regression. In 2020 3rd International Conference on Inventive Systems and Control (ICISC) (pp. 2215-2220). IEEE.

- Li, Y., Luo, J., Wu, H., & Chen, H. (2019). Travel time prediction based on XGBoost. Proceedings of the 2019 11th International Conference on Machine Learning and Computing (ICMLC) (pp. 222-227). ACM.

- Short-Term Travel Time Prediction: A Spatiotemporal Deep Learning Approach Xiangdong Ran, Zhiguang Shan, Yong Shi, and Chuang Lin

- Prakash, I., Rao, K.R., & Kumar, P.R. (2017). A review on travel time prediction. International Journal of Computer Science and Information Security, 15(2), 88-99.

- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

- Zhang, Z., Li, X., Narayan, A., & Zhang, S. (2017). Traffic speed prediction using deep learning methods. arXiv preprint arXiv:1707.07358.

- Li, B., Zhang, D., Sun, L., Chen, C., Li, S., Qi, G., & Yang, Q. (2018). Convolutional neural network for multi-category rapid detection of cerebral infarction on diffusion-weighted imaging. Scientific Reports, 8(1), 1-10.

- Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2019). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transportation Research Part C: Emerging Technologies, 58, 308-321.

- Yu, B., Yin, H., & Zhu, Z. (2017). Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. Sensors, 17(7), 1501.

- Polson, N. G., & Sokolov, V. O. (2017). Deep learning for short-term traffic flow prediction. Transportation Research Part C: Emerging Technologies, 79, 183-197.