Model to Predict Passenger Transport Demand Based on RNN

Rohit Rawate, Shreyas Patole, Sakshi Gaikwad, Dheeraj Mugdiya, Prof. M. P. Desai
Department of Information Technology
Smt. Kashibai Navle College of Engineering Pune, Maharashtra, India

Abstract: *Particularly in urban areas, taxis are an essential part of mobility. Predicting future demand for taxis in a specific area will be very beneficial to internet-based transportation companies like Ola, Uber, and others so that we can significantly reduce customer/passenger waiting times and help taxi drivers to move to specific areas where demand is high, ultimately making passengers, drivers, and companies happy. We like to forecast the demand for taxis in a certain place for the proposed project.*

Keywords: Taxi Demand Prediction, Recurrent Neural Network

I. INTRODUCTION

Taxis typically cruise around without a defined destination when there is no passenger on board; this is known as "cruising." Such a taxi may find clients by cruising the streets or by being close to those who have requested a taxi through taxi aggregation services like Uber, Ola, etc. In these situations, the taxi must be at the right place at the right time in order to enhance revenue by cutting down on cruising time.Nowadays, all city dwellers must commute in order to get to their destinations from where they are. One of the key forms of transportation in cities is the taxi. As a result, several internet-based businesses like Uber and Ola are now involved in it on a significant basis. But, these businesses and the cab drivers are dealing with some serious issues. One of the biggest hurdles for all cab drivers is finding a passenger.

Fuel usage increases and fewer passengers are transported if the taxi driver takes longer to get to each new passenger. Since we are rookie taxi drivers, we frequently are unsure of the best location to pick up a new passenger since we lack accurate knowledge of the location's and time's taxi demand.Both rookie and seasoned taxi drivers can utilise this information about anticipated taxi demand to find their way to the city's busiest locations more quickly. As a result, the availability of taxi services in urban areas helps to fulfil demand.Demand forecasting is difficult since it depends on so many different factors. Due to rain, there may be unexpected increases in demand because of nearby activities like cricket matches, music festivals, or religious gatherings. These incidents also cause a sharp rise in the area's need for taxis. Most of the time, we rely on manual labor, yet it is insufficient. So, we seek improved deep learning and machine learning algorithms that are based on regression.

II. LITERATURE SURVEY

Fei Miao et al [1] proposed a modern robust transportation system that senses data collected from transportation systems that help in analyzing the passenger demands. Prediction Methods on taxi passenger demand were travel time and travelling speed according to traffic monitoring data have been developed. In their proposed model Robotic mobility on-demand systems that minimize the number of rebalancing trips and best parking systems that allocate resource based on a driver's payments. Estimations show that under the robust dispatch framework we design, the average demand-supply ratio imbalance is reduced by 31.7%, and the average total idle driving distance is reduced by 10.130% or about 20 million miles in total in one year.

Mohammad Saiedur Rahaman et al [2] defined the neighbourhood identification problem in the presence of a large number of heterogeneous contextual modules. It codifies research as a problem of less wait time prediction for taxi drivers at airports and investigate heterogeneous elements related to time, weather, flight arrivals and taxi trips. The queue managers continuously monitor the concurrent queues related to taxis and passengers and instruct taxi drivers to join the passengers at the terminal when there is higher demand. To ensure the seamless operation of this process, the queue manager estimates the demand for taxis in future. Airport satisfaction ratings depend on the proper management of both passenger queues and taxi. Aiming to maintain demand-supply symmetry of taxis the airport transport managers employ an approach where it requires extended human intervention. This paper shows Pearson's correlation scores of 0.393 (with mean) and 0.484 (with median). They argue that the quality of the neighbourhood they identified is significantly improved by the consideration of relevant heterogeneous contextual factors, thus the performance is boosted (i.e. mean prediction error is less than 0.09 and the median prediction error is less than 0.06).

Desheng Zhang et al [3] states that the existing system of data collection is offline and collected by manual investigations and it may result in inaccurate data for real-time analysis. By implementing this, they can infer arriving passenger moments by investigating the logical information. They used 450GB dataset of 14,000 taxicabs for a half year and it achieves 83% accuracy and outperforms the statistical model by 42%. Passenger demand prediction may be halted by bad weather, special events or accidents. The use of Dmodel yields 83% accuracy and outperforms statistical model by 42%. They used a Hidden Markov Chain for implementation. The Dmodel based dispatching outperforms basic and SDD based by 11% on an average. This is due to the accurate inference by Dmodel.

Luis Damas et al [4] proposed a novel methodology for predicting the spatial distribution of taxi–passengers for a short time horizon using streaming input data. The results elaborated so that the framework proposed can provide effective learning into the spatio-temporal distribution of taxi–passenger demand for a horizon of 30 minutes. This paper focuses on the real-time choice problem of which is the best cab service stand to go to after a passenger drop-off (i.e., the stand where another passenger can be picked up within a short span of time). An intelligent approach regarding this flaw will improve network reliability for both companies and customers; an intelligent distribution of vehicles throughout stands will minimize the average waiting time to pick up a passenger, while the distance travelled will be more profitable.

Biao Leng et al [5] elaborated the battle between two taxi companies in China namely Didi and Kauidadi that occurred in 2014. The two companies are backed up by internet giants like Tencent and Alipay. These companies promoted the taxi drivers by giving them incentives for each ride and also allowed the users also to use their application by giving frequent discounts and offers and also promoted payment through the mobile phone. In this paper, they collected a 37-day trip data and use 9000 entries in Beijing. For the first 18 days there was no battle and for the next 19 days, the battle was competitive. The spatial temporal data are studied and based on the comprehensive analysis, benefits and drawbacks are discussed.

2.1 Objectives of System

- To collect the data which is suitable for our problem statement.
- To pre-process the data collected.
- To select appropriate machine learning and deep learning algorithms.
- To prepare a predictive regression model by training it with the collected data set.
- By comparing the different models, we will select the best model to predict the taxi demand which helps the taxi drivers.

III. IMPLEMENTATION DETAILS OF MODULE

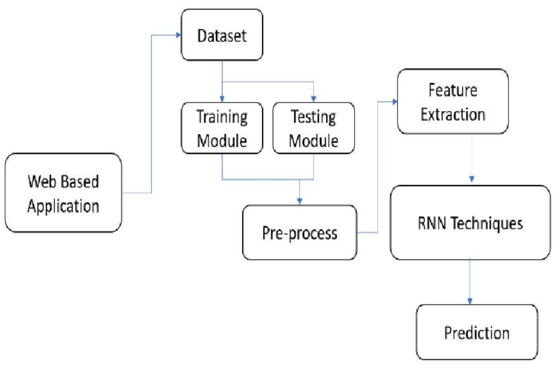

Figure a: Block Diagram

The proposed system undergoes some modules such as :-

- **Data pre-processing:** It is a data mining approach that entails putting raw data into an intelligible format. Data cleansing techniques include filling in missing numbers, smearing noisy data, and resolving discrepancies in the data. The dataset is cleaned, and decimal values are changed into appropriate float values because it has some missing values.
- **Data splitting:** A training set and a testing set are created from the newly created dataset. The division is carried out on an 80-20 split. To train the model, 80% of the dataset is selected as the Training Set. The model is tested and its accuracy is examined using the remaining 20%, which is known as the Test Set.
- **Feature Selection:** The properties of the data used to train machine learning models have a significant impact on the model's performance. Model performance may be adversely affected by irrelevant or only partially relevant features.
- **Classification:** By adjusting the training set to the classifier model, the model is trained. After testing, the classifier model assigns the air quality a good or bad rating. The categorizations closely match the testing set.

3.1 Working

The data set for taxi rides will be obtained via Kaggle. The pre-processing of the obtained data is a data mining approach that entails converting the raw data into a comprehensible format. Data is frequently inaccurate, inconsistent, lacking in specific patterns of activity, and/or incomplete. Data cleaning, data integration, data transformation, data reduction, and other tasks are among the main duties of the data preparation. The missing data are located, and the mean values are used in their place. The cleaned dataset is provided to the recurrent neural network model, which is trained using historical date, time, pickup location, weather, and other data. As a result, the model can forecast future demand for the taxi service. Because they are the only ones having an internal memory, recurrent neural networks (RNN) are a strong and resilient subclass of neural networks and one of the most promising algorithms available right now.

IV. CONCLUSION

For anticipating the demand for taxis in various parts of the city, a sequential learning model using recurrent neural network is developed. The field of taxi services can be improved so that it is more user-friendly and provides more precise forecasts. The dataset was recently gathered from multiple sources,

Like Kaggle and pre-processing and data cleaning were carried out. The Outliers were removed to get accurate results and we concentrated on specific area for anticipating the ride hailing users and the time of the day when the demand is high.

REFERENCES

[1]. Fei Miao, Shuo Han, Shan Lin, Qian Wang, John A. Stankovic, Abdeltawab Hendawi, Desheng Zhang, Tain he and George J. Pappas, "Data-Driven Robust Taxi Dispatch Under Demand Uncertainties", IEEE Transactions on Control Systems Technology, , Jan.2019.

[2]. Jun Xu , Rouhollah Rahmatizadeh, Ladislau Bölöni, and Damla Turgut, “Real-Time Prediction of Taxi Demand Using Recurrent Neural Networks”, IEEE Transaction on Intelligent transport system, vol. 19, no. 8, pp. 2572-2581,Aug.2018.

[3]. Mohammad Saiedur Rahaman, Yongli Ren, Margaret Hamilton and Flora D. Salim, "Wait Time Prediction for Airport Taxis Using Weighted Nearest Neighbor Regression", IEEE Access, vol. 6, pp. 74660 - 74672, Nov.2018.

[4]. Desheng Zhang, Tian He, Shan Lin, Sirajum Munir and John A. Stankovic, "Taxi-Passenger-Demand Modeling Based on Big Data from a Roving Sensor Network", IEEE Transactions on Big Data, vol. 3, no. 3, pp. 362 - 374, Sept.2017.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

IJARSCT

Impact Factor: 7.301

Volume 3, Issue 3, April 2023

[5]. Biao Leng, Heng Du, Jianyuan Wang, Li Li and Zhang Xiong, "Analysis of Taxi Drivers' Behaviors Within a Battle Between Two Taxi Apps", IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 1, pp. 296 - 300, jan.2016.

[6]. Fei Miao, Student Member, Shuo Han, Member, Shan Lin, John A. Stankovic, Desheng Zhang, Sirajum Munir, Hua Huang, Tian He, and George J. Pappas, "Taxi Dispatch With Real-Time Sensing Data in Metropolitan Areas: A Receding Horizon Control Approach", IEEE Transactions on Automation Science and Engineering, vol. 13, no. 2, pp. 463 - 478, Mar.2016.