

Project Proposal

Taxi Ride Demand Prediction using Machine Learning

We are 5 in our team, and we collaboratively planned 'Taxi Ride Demand Prediction using Machine Learning' as our project because - accurately predicting taxi ride demand, taxi companies can optimize their resources and reduce operational costs. This is particularly important during peak hours or events when there is a surge in demand. The prediction model can help the taxi companies to better allocate their resources, reduce wait times for passengers, and increase the number of rides per day and moreover; this project offers an opportunity to gain hands-on experience in data analysis and machine learning, which are important skills in today's job market. The project will involve working with large datasets, implementing machine learning algorithms, and evaluating the model's performance. These skills can be applied to a wide range of industries and projects beyond the taxi industry.

Participant Details:

- 1) Dheeraj Podishetty (dheerajpodishetty@my.unt.edu)
- 2) Sai Krishna Kakollu (Saikrishnakakollu@my.unt.edu)
- 3) Sai Karthik Gamineedi (saikarthikgamineedi@my.unt.edu)
- 4) Amulya Bodempudi (amulyabodempudi@my.unt.edu)
- 5) Rahul Varma Muppalla (RahulVarmaMuppalla@my.Unt.edu)

Roles & responsibilities of team members:

In this project, my role is to lead my team members and guide them on the various modules of the project, and responsible for developing the python code for designed functions and modules. Karthik and Krishna support us in preparing documentation of the project. Rahul and Karthik assist us on data analysis and testing the data and predicted results. Amulya helps in developing and manipulating the code to get accurate outcome along with me.

Workflow:

We will have meetings twice a week outside of class based on our convenient time, one is offline, which is after our class at 12:00 PM CST/ CDT at Discovery Park to discuss the doubts and clarifications on the project tasks, and the other is through zoom video conferencing/ screen sharing based on everyone's feasibility. We will use OneDrive folder to collaborate on documents and data. We will also use a UNT email and Teams to communicate for updates and questions.

Shared OneDrive Url: https://myunt-my.sharepoint.com/personal/dheerajpodishetty_my_unt_edu/Documents/5502%20-%20Principles%20of%20DS/Taxi%20Ride%20Demand%20Prediction%20-%20Shared

Project Abstract:

Taxi Ride Demand Prediction using Machine Learning is a project that aims to develop accurate and reliable machine learning models to predict the demand for taxi rides in each area. The project involves collecting and pre-processing large amounts of data on taxi rides, weather patterns, and other relevant factors, and using this data to train and fine-tune machine learning algorithms. The project has several potential applications, such as improving the efficiency and reliability of taxi services. The goal of the taxi demand prediction problem is to accurately forecast the volume of taxi requests for a given area during a given time using data from prior taxi requests. Our objective is to increase the taxi demand prediction's accuracy by employing various prediction techniques. In this project, machine learning is used to investigate the demand prediction issue for taxi rides. Predicting how many taxi trips will be needed in a specific area and during a specific time frame can help taxi businesses manage their fleet more effectively and satisfy customers.

The predicted accuracy of several machine learning methods, including regression models and time series analysis, is compared, and evaluated. The data comes from a real-world taxi service and contains a range of variables, including the weekdays and week off statistics that is using categorical feature engineering, which is a type of supervised learning method and other events that may influence the demand for taxi rides. The results reveal that machine learning models can accurately forecast taxi ride demand, with some models outperforming others based on the specific context and data provided. For taxi businesses looking to enhance their operations and customer experience through data-driven decision-making, the findings have real-world applications.

Overall, Taxi Ride Demand Prediction using Machine Learning is an exciting and challenging project that has the potential to make a significant impact on transportation and urban planning.



Fig: Workflow of the project

Our results might look similar to the graph below where the comparison between the actual and the predicted values of the count of cabs booked per day.

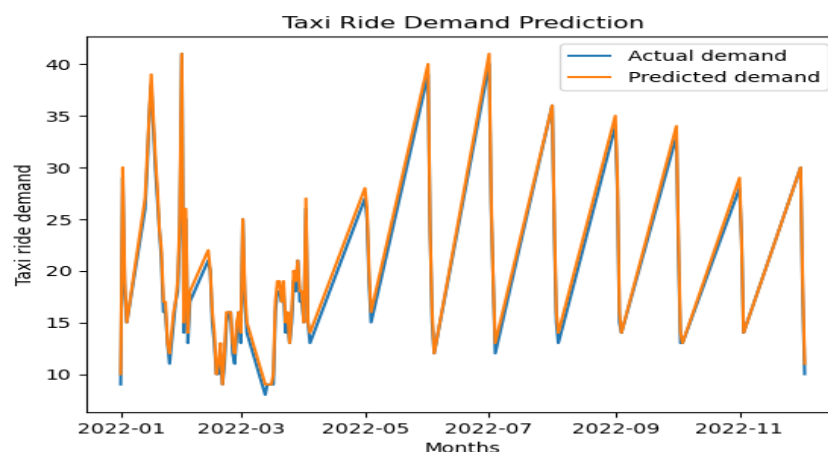


Fig: Relation between the actual demand vs predicted demand

Project Design:

Technologies and Algorithms:

- Python programming language
- Scikit-learn machine learning library.
- Pandas, NumPy, Statistics and Polars data manipulation libraries
- Aggregation Algorithm – To find the relation between the dependent(outcomes) and independent variables(predictions)

Milestones:

- Data collection and pre-processing (by Week 2)
- Exploratory data analysis and visualization (by Week 3)
- Dataset overview and selecting the algorithm (by Week 4)
- Model selection and training the machine with algorithm (by Week 6)
- Model evaluation and optimization (by Week 7)
- Final testing and prediction (by Week 8)

Resources and Related Projects:

Resources:

- Scikit-learn documentation (<https://scikit-learn.org/stable/documentation.html>)
- Pandas documentation (<https://pandas.pydata.org/docs/>)
- NumPy documentation (<https://numpy.org/doc/>)
- Statistics (<https://docs.python.org/3/library/statistics.html>)
- Polars(<https://www.pola.rs/>)

Contrasting resources:

- Uber Engineering's article on demand prediction using neural networks (<https://eng.uber.com/neural-networks/>)

This article describes how Uber uses neural networks to predict demand for its ride-hailing services. While our project focuses on predicting demand for traditional taxi services, this resource provides a good comparison to our approach and insights into neural network models.

- Kaggle's NYC Taxi and Limousine Commission Trip Record Data (<https://www.kaggle.com/c/nyc-taxi-trip-duration>)

This dataset contains historical data on taxi trips in New York City, which is like the data we will be using in our project. We can use this dataset to compare our model's performance to other models and approaches on Kaggle.