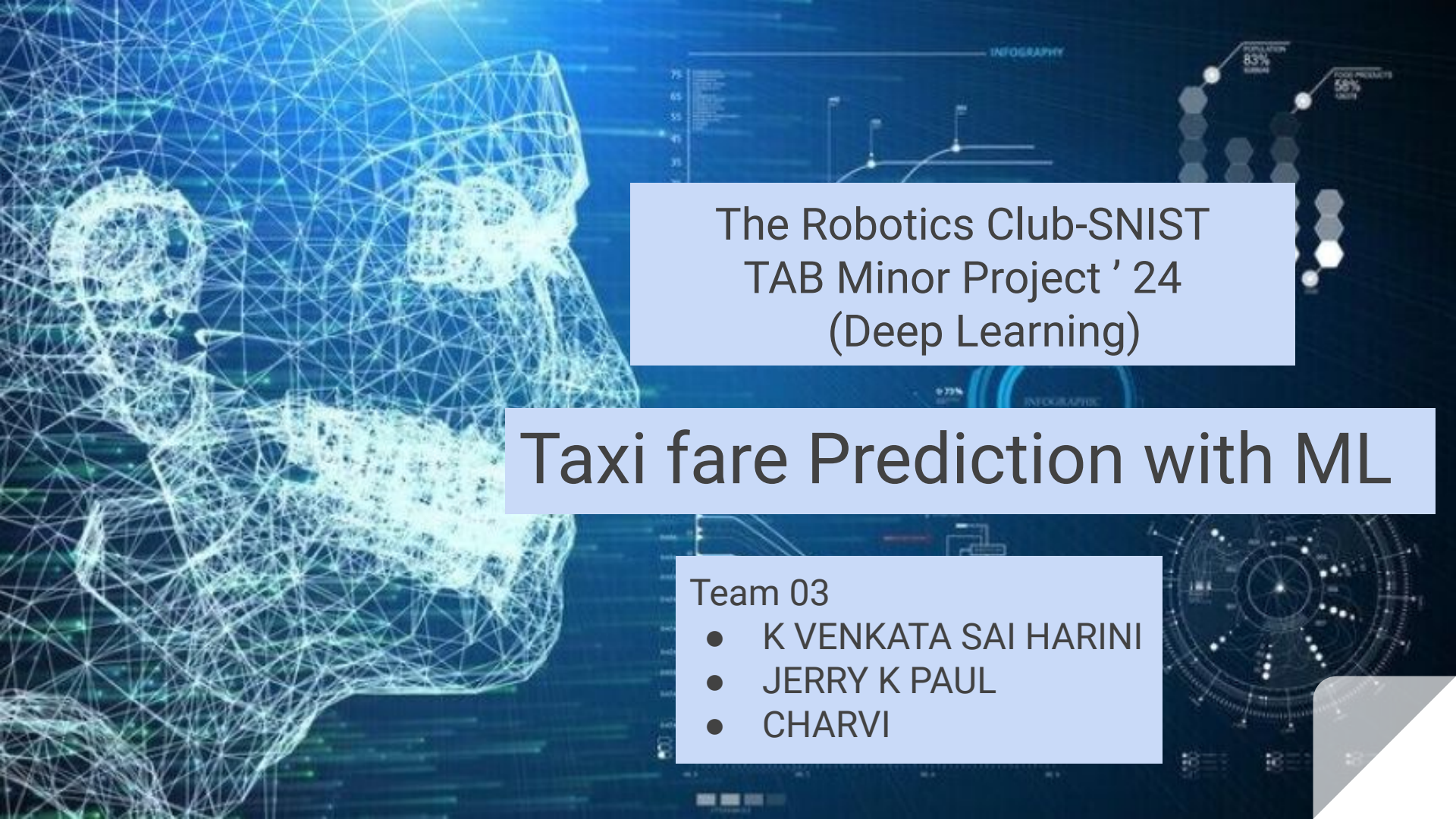


The Robotics Club-SNIST
TAB Minor Project ’ 24
(Deep Learning)

Taxi fare Prediction with ML

Team 03

- K VENKATA SAI HARINI
- JERRY K PAUL
- CHARVI

Introduction

Objective: To develop a machine learning model for predicting taxi fares based on input features such as trip duration, distance traveled, number of passengers, base fare, tips, miscellaneous fees, and surge pricing.

Our goal was to create a regression model capable of estimating taxi fares accurately. After data preprocessing, we used the Scikit-Learn library to train and evaluate the model. Key evaluation metrics included Mean Absolute Error (MAE), Mean Squared Error (MSE), and the R-squared score.

Abstract- Problem statement

To predict the fare of the taxi ride using Random Forest given pickup and drop-off locations, the pickup timestamp, and the passenger count.

Significant factors of Taxi Fare: Distance travelled, Duration, No. of passengers, fare, tips, misc. Fee, surge applied

Abstract- Approach

1.**Data Collection:** Gather historical taxi ride data including fare amounts and relevant features such as pickup/drop-off locations, distance travelled, time of day and also additional factors like weather conditions.

2.**Data Preprocessing**: Clean the data by handling missing values, removing outliers, and encoding categorical variables. Additionally, feature engineering may be performed to create new features or extract valuable information from existing ones.

3.**Feature Selection**: Select the most relevant features that have a significant impact on the fare amount. This can be done through techniques like correlation analysis or feature importance ranking from machine learning models.

Abstract- Approach

4.**Model Selection**: Choose an appropriate regression model for prediction. Common choices include linear regression, decision trees, **random forests**, gradient boosting machines, or neural networks. Experiment with different models to find the one that performs best on the dataset.

5.**Model Training**: Split the data into training and testing sets. Train the selected model on the training data using appropriate techniques such as cross-validation or regularization to prevent overfitting.

6.**Model Evaluation**: Evaluate the trained model's performance on the testing data using metrics like mean squared error, mean absolute error, or R-squared to assess how well the model generalizes to unseen data.

7.**Model Tuning**:Fine-tune the model hyperparameters to optimize its performance further. Techniques like grid search or random search can be employed to find the best combination of hyperparameters.

Tools and Technology used

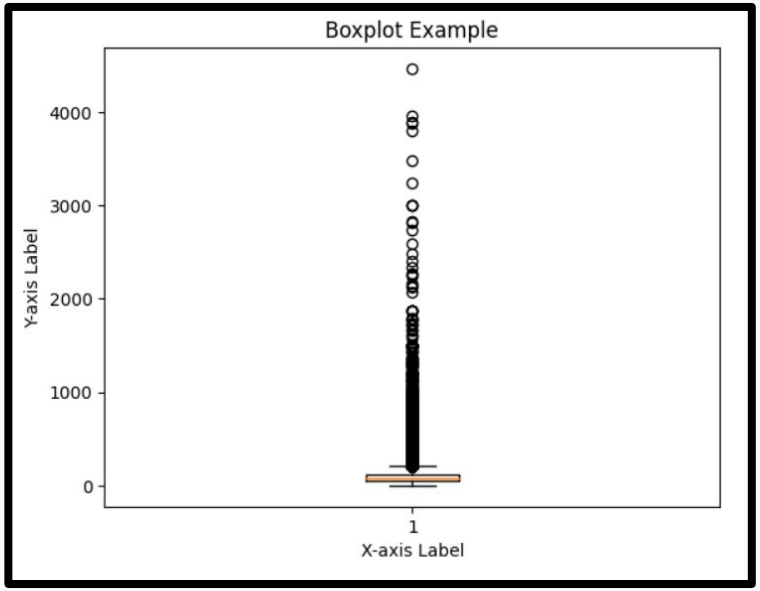
Hardware: Computer System

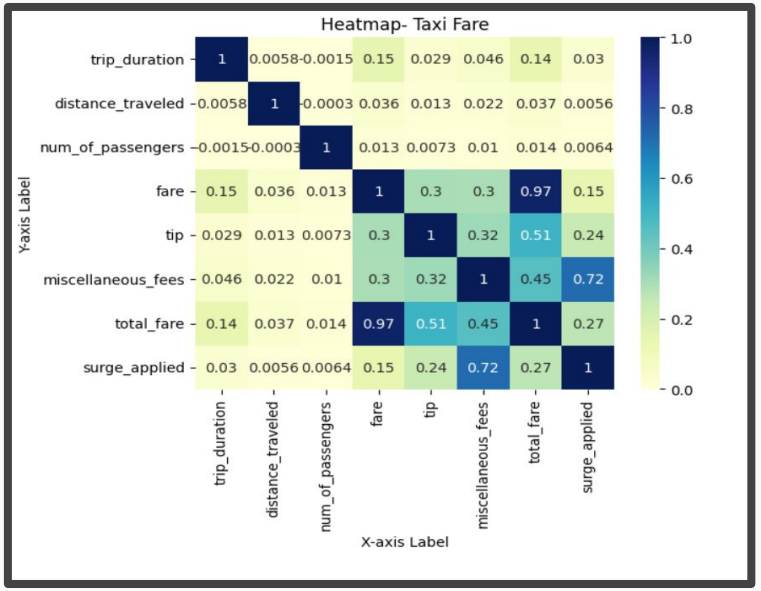
Software: Google Colab

Language: Python

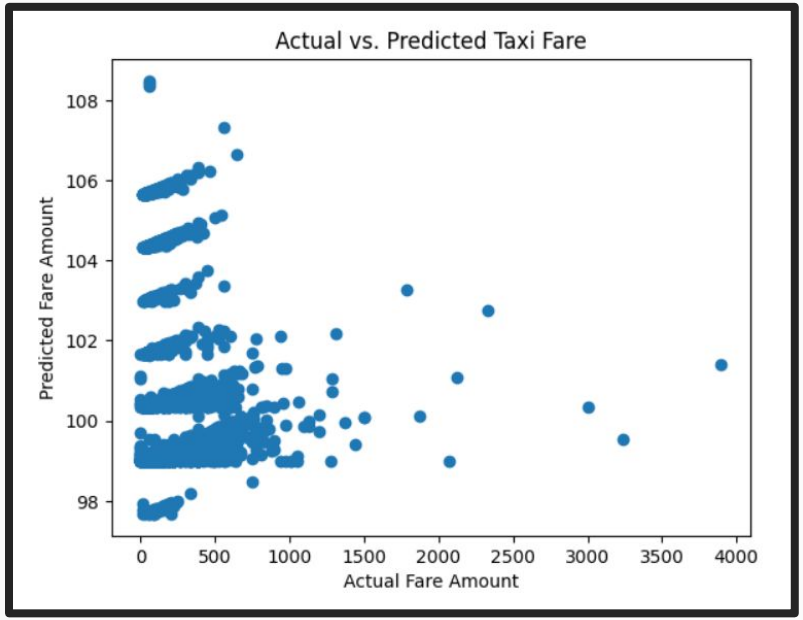
Libraries: Pandas, Numpy, Matplotlib, Seaborn, Sklearn

Project-


Boxplot Example
Y-axis Label
4000
3000
2000
1000
0
1
X-axis Label


Heatmap- Taxi Fare
trip_duration 1 0.0058 -0.0015 0.15 0.029 0.046 0.14 0.03
distance_traveled 0.0058 1 -0.0003 0.036 0.013 0.022 0.037 0.0056
num_of_passengers -0.0015 -0.0003 1 0.013 0.0073 0.01 0.014 0.0064
fare 0.15 0.036 0.013 1 0.3 0.3 0.97 0.15
tip 0.029 0.013 0.0073 0.3 1 0.32 0.51 0.24
miscellaneous_fees 0.046 0.022 0.01 0.3 0.32 1 0.45 0.72
total_fare 0.14 0.037 0.014 0.97 0.51 0.45 1 0.27
surge_applied 0.03 0.0056 0.0064 0.15 0.24 0.72 0.27 1
Y-axis Label
X-axis Label
trip_duration
distance_traveled
num_of_passengers
fare
tip
miscellaneous_fees
total_fare
surge_applied
1.0
0.8
0.6
0.4
0.2
0.0

Project-



NOTE

Below Rs.800, the model prediction is very accurate.
But as the actual fare increases, prediction becomes less accurate.

Conclusion

Using ML techniques to predict the taxi fare enhances the efficiency and accuracy in the fare prediction.

This taxi fare prediction project demonstrates the power of machine learning in solving real-world problems. By leveraging data analysis and model development, we've created a tool that can provide accurate fare estimates for taxi rides, benefiting both passengers and service providers.

THANK YOU


100%
EVOLUTION
78%
STATUS
0110
0100
1001
Artificial Intelligence vs Machine Learning vs Data Analytics