International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Taxi Fare Prediction

Om Prakash Barapatre[1]*, Priyanshu Sagar*[2]*, Aakash Verma*[3]*, Harshil Rathod*[4]*, Pranjal Pandey*[5]

[1]*Assistant Professor, Bhilai Institute of Technology, Raipur, Chhattisgarh, India*
[2,3,4,5] *Student, Bhilai Institute of Technology, Raipur, Chhattisgarh, India*

ABSTRACT

Predictive analytics uses archived data to predict long-term events. Understand important trends using mathematical models from historical data. The model then uses existing data to predict long-term predictions or derive actions to achieve visual results. Predictive analysis has recently been highly praised due to the development of assistive technology in the field of machine learning, which deals with large amounts of data. Many industries use predictive analytics to create accurate forecasts, such as: B. Showing fares for city trips. This resource planning is enabled by forecasting, so you can more accurately predict taxi rides, for example. A taxi startup company takes many factors into consideration. This project attempts to understand patterns and predict freight rates using various methods. This project is designed to predict taxi fares in a particular city. The project includes various steps, including training and testing with different variables such as pick-up and drop-off locations to predict taxi fares. The important thing here is that the project is convenient and easy to use for the customer. We use Geopy to make our projects more reliable for our customers. While other projects require you to explicitly specify the coordinates to find an address, Geopy allows you to access the name of a place directly.

Keywords: Geopy, Machine Learning, Predictive analytics

1. INTRODUCTION

Introduction Machine learning is widely used throughout the work in prediction using systems. There are different types of app machines that use machine learning for prediction. Among them are also supervised learning and unsupervised learning. For problems related to business needs, machine learning can also be called predictive. Machine learning has many different origins. Supervised learning is one of the most commonly used learning methods in machine learning. To train the data here, we need some things, such as a dataset, and we need to use an algorithm that predicts the output of the program we run. If you don't have a dataset to train your model on, the resulting output will be unreliable and the rate will be inaccurate. Unsupervised learning is like learning without the help of datasets or external factors. An appropriate algorithm is automatically selected and attempts to predict the output with near accuracy. It is mainly used to make accurate predictions using various methods and algorithms provided to people. Therefore, machine learning concepts are widely used in many companies as they are used every day in everyday life and will continue to be used in the future as the scope is very wide**.**

2. SOFTWARE REQUIREMENTS

1.	Operating system	:	Windows 10
2.	Languages used	:	Python
3.	Python version	:	3.5 or 3.6
4.	Notebooks	:	Jupyter Notebook
5.	Emulators	:	No emulators used
6.	Software Libraries	:	Pandas, Matplotlib, Numpy, Seaborn, Math, Sklearn.

3.WORKING:

According to industry standards, the process of data analysis mainly includes six key steps, and this process is abbreviated as CRISP DM process, a cross-industry process for data mining. is. The six main steps of the CRISP DM methodology for developing a model are:

- Business Understanding
- Data Understanding
- Data Preparation
- Data Preprocessing
- Modeling
- Evaluation

3. ALGORITHMS:

3.1 RANDOM FOREST

Random Forest Random Forest is a popular supervised learning method. Machine learning algorithm. It can be used for both classification and regression problems in ML. It is based on the concept of ensemble learning, which is the process of combining multiple classifiers to solve complex problem and improve model performance

3.2 Linear Regression

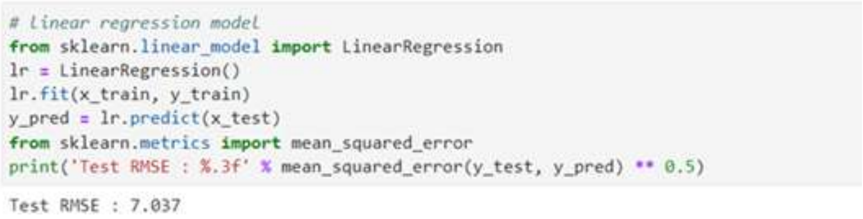
Linear Regression A supervised machine learning algorithm. This is primarily used after the correlation step. If you want to predict the value of the y variable using the value of another variable, you can use the Iinear algorithm. Linear regression is one of the simplest and most popular machine learning algorithms. A statistical method used for predictive analysis. Linear regression predicts continuous/real or numeric variables such as sales, salary, age, product price, etc. A linear regression algorithm is called linear regression because it shows a linear relationship between a dependent variable (y) and one or more independent variables (y).

Linear regression shows a linear relationship, so it determines how the value of a dependent variable changes depending on the value of an independent variable3.

4. RESULT:

We follow several criteria to determine these according to industry standards. Some of these also calculate error rates and accuracy. Our project uses MAE and MAPE. MAE (Mean Absolute Error) is one of the error measures used to calculate the predictive performance of a model

Method	Mae Error (in Percentage)
Random Forest	20.2135
Linear Regression	28.1744



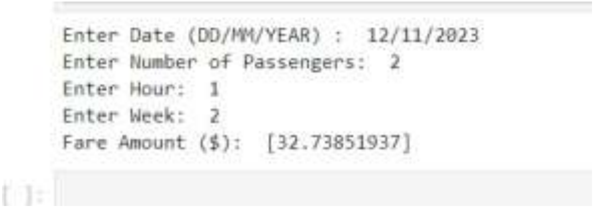
```
# linear regression model
from sklearn.linear_model import LinearRegression
lr = LinearRegression()
lr.fit(x_train, y_train)
y_pred = lr.predict(x_test)
from sklearn.metrics import mean_squared_error
print('Test RMSE : %.3f' % mean_squared_error(y_test, y_pred) ** 0.5)

Test RMSE : 7.037
```

Fig 5.1 Mean absolute error

The second matrix for identifying or comparing better models is precision. This is the ratio of the number of correct predictions to the total number of predictions made. Accuracy = number of correct predictions / total number of predictions made.

It can also be calculated from MAE as Accuracy = 1-MAPE.

Method	Accuracy (in Percentage)
Random Forest	79.7864
Linear Regression	71.8255



```
Enter Date (DD/MM/YEAR) :  12/11/2023
Enter Number of Passengers:  2
Enter Hour:  1
Enter Week:  2
Fare Amount ($):  [32.73851937]
```

Fig 5.2 Accuracy

CONCLUSION

Taxi fare prediction is one of the essential applications in the automotive industry. The purpose of the research is to predict taxi fares using random forests and linear regression algorithms. We achieved our project goal by predicting taxi fares (Random Forest accuracy was 79.78%). The accuracy of linear regression is 71.82D

The quality of a regression model is determined by whether the predictions match the actual values. In regression problems, the dependent variable is continuous. In classification problems, the dependent variable is a categorical variable. Random forests can be used to solve both regression and classification problems. Decision trees are nonlinear. Unlike linear regression, there is no equation that describes the relationship between the independent and dependent variables. Of the remaining three models, Random Forest is the best model. This is because it has the lowest RMSE score and the highest R-squared score. This will tell you where the variability is highest and how well the model fits this data.

REFERENCES:

[1] References Gunjan Panda, Supriya p.panda "Machine learning using exploratory analysis to predict taxi trips". International Journal for Research in Applied Science & Engineering Technology (IJRASET), August 2019.

[2] Kelareva, Elena. "Predicting the future with Google Maps API" Web her blog post. Geo Developers Blog, https://maps-apis.googleblog.com/2015/11/predicting-future-with-google-maps-apis.html Accessed December 15, 2016.

[3] Wu, Chun-Hsin, Jiang Ming Ho and Da Tsai Li. “Travel Time Prediction Using Support Vector Regression,” IEEE Transactions on Intelligent Transportation Systems 5.4 (2004): 276-281.

[4] Van Lindt, J. W.C., S.P. Hoogendoorn, and Henk J. Van Zuylen. “Accurate prediction of highway travel times using state-space neural networks under missing data” Transportation Research Part C: Emerging Technologies 13.5 (2005): 347-369.

[5] Vanajakshi, L., S.C. Subramanian, and R. Sivanandan. “Travel time prediction under heterogeneous traffic conditions using Global Positioning System data from buses” IET Intelligent Transport Systems 3.1 (2009): 1-9.

[6] Weijie Wang and Yanmin Lu, Analysis of mean absolute error (MAE) and root mean square error (RMSE) in rounding model evaluation, ICMEMSCE, IOP Publishing, 324 (2018), doi:10.1088/1757 - 899X/324/1/012049

[7] X. Qian, S. V. Ukusuri.: Time-of-day pricing in the taxi market. IEEE Transactions on Intelligent Transportation Systems, Bd. June 18, 2017.

[8] Yildirimoglu, Mehmet, and Nicolas Geroliminis. “Experienced Travel Time Prediction for Congested Highways” Transportation Research Part B: Methodology 53 (2013): 45-63.