

DYNAMIC PRICING PREDICTION FOR CABS USING IBM WATSON

TEAM MEMBERS

- THRIVENI PEGYAPURAM
- NAGAVELLI MANISH KUMAR
- N. ADITHYA VARDHAN

INTRODUCTION:

A car price prediction has been a high interest research area, as it requires noticeable effort and knowledge of the field expert. Considerable number of distinct attributes are examined for the reliable and accurate prediction. To build a model for predicting the price of used cars in Bosnia and Herzegovina, we applied three machine learning techniques (artificial neural network and Random Forest). However, the mentioned techniques were applied to work as an ensemble. The data used for the prediction was collected from the web portal autopijaca.ba using web scraper that was written in PHP programming language. Respective performances of different algorithms were then compared to find one that best suits the available data set. The final prediction model was integrated into Java application. Furthermore, the model was evaluated using test data and the accuracy of 87.38% was obtained. a cab rental start-up company has successfully run the pilot project and now wants to launch their cab service across the country. After collecting the historical data from the pilot project, the certain variables are obtained which leads to the fare amount. The project is based on the Forecasting/Prediction type of data science problem statement. The variables that are obtained after the pilot project are as follows:

- Pickup date and time indicating when the cab rides started.
- Pickup longitude indicating the longitude coordinates of where the cab rides started.
- Pickup latitude indicating the latitude coordinates of where the cab rides started.
- Drop-off longitude indicating the longitude coordinates of where the cab rides ended.
- Drop-off latitude indicating the latitude coordinates of where the cab rides

ended. • Passenger count indicating the number of passengers in the car rides.

LITERATURE SURVEY:

In this chapter, we discuss various applications and methods which inspired us to build our project. We did a background survey regarding the basic ideas of our project and used those ideas for the collection of information like the technological stack, algorithms, and shortcomings of our project which led us to build a better project. Predicting price of a used cars has been studied extensively in various researches discussed, in her paper written for Master thesis [2], that regression model that was built using Random Forest Regression can predict the price of a car that has been leased with better precision than multivariate regression or some simple multiple regression. This is on the grounds that Random Forest Regression is better in dealing with datasets with more dimensions and it is less prone to overfitting and underfitting. The weakness of this research is that a change of simple regression with more advanced Random Forest regression was not shown in basic indicators like mean, variance or standard deviation.

HARDWARE AND SOFTWARE REQUIREMENTS:

Hardware Requirements:

- Processor: Minimum 1 GHz; Recommended 2GHz or more.
- Ethernet connection (LAN) OR a wireless adapter (Wi-Fi)
- Hard Drive: Minimum 32 GB; Recommended 64 GB or more.
- Memory (RAM): Minimum 1 GB; Recommended 4 GB or above.

Software Requirements:

1. Anaconda Navigator

2. Spyder

3. Jupiter Notebook

OBJECTIVES:

Write what are all the technical aspects that students would get if they complete this project.

- Knowledge on Machine Learning Algorithms.
- Knowledge on Python Language with Machine Learning
- Knowledge on Statistics and Graphs and their relations
- You'll be able to understand the problem to classify if it is a regression or a classification kind of problem.
- You will be able to know how to pre-process / clean the data using different data pre-processing techniques.
- Applying different algorithms according to the dataset and based on visualization.
- Real Time Analysis of Project
- Building an ease of User Interface (UI)
- Navigation of ideas towards other projects(creativity)
- Knowledge on building ML Model.
- You will be able to know how to find the accuracy of the model.
- How to Build web applications using the Flask framework.
- To find correlations between each variable and the target variable.

- To build a linear regression model.
- To calculate efficiency of the model and errors in predictions

EXPERIMENTAL INVESTIGATIONS:

```
#Importing Libraries

import numpy as np #linear algebra
import pandas as pd #data processing
import matplotlib.pyplot as plt
import seaborn as sns #used for data visualization
import pickle
from collections import Counter as c # return counts
from sklearn.preprocessing import LabelEncoder #importing the LabelEncoding from sklearn
from sklearn.model_selection import train_test_split #split data in train and test array
from sklearn.ensemble import RandomForestRegressor #regression ML algorithm
```

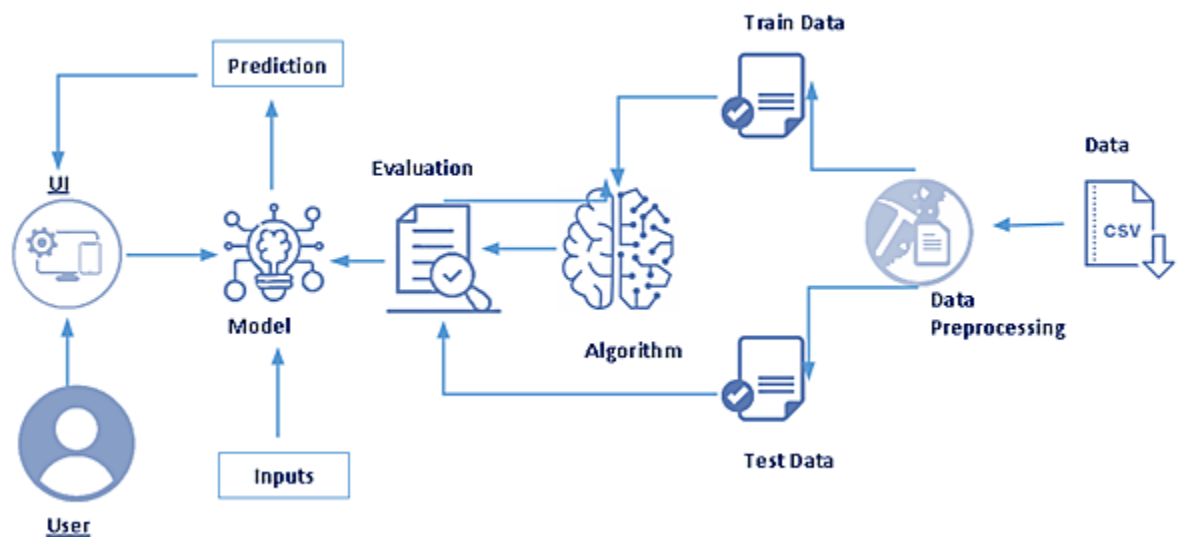
```
#Cab rides represent the for the type of cab, location, name etc.
rides_df.head()
```

	distance	cab_type	time_stamp	destination	source	price	surge_multiplier	id	product_id	name
0	0.44	Lyft	1544952607890	North Station	Haymarket Square	5.0	1.0	424553bb-7174-41ea-aeb4-fe06d4f4b9d7	lyft_line	Shared
1	0.44	Lyft	1543284023677	North Station	Haymarket Square	11.0	1.0	4bd23055-6827-41c6-b23b-3c491f24e74d	lyft_premier	Lux
2	0.44	Lyft	1543366822198	North Station	Haymarket Square	7.0	1.0	981a3613-77af-4620-a42a-0c0866077d1e	lyft	Lyft
3	0.44	Lyft	1543553582749	North Station	Haymarket Square	26.0	1.0	c2d88af2-d278-4bfd-a8d0-29ca77cc5512	lyft_luxsuv	Lux Black XL
4	0.44	Lyft	1543463360223	North Station	Haymarket Square	9.0	1.0	e0126e1f-8ca9-4f2e-82b3-50505a09db9a	lyft_plus	Lyft XL

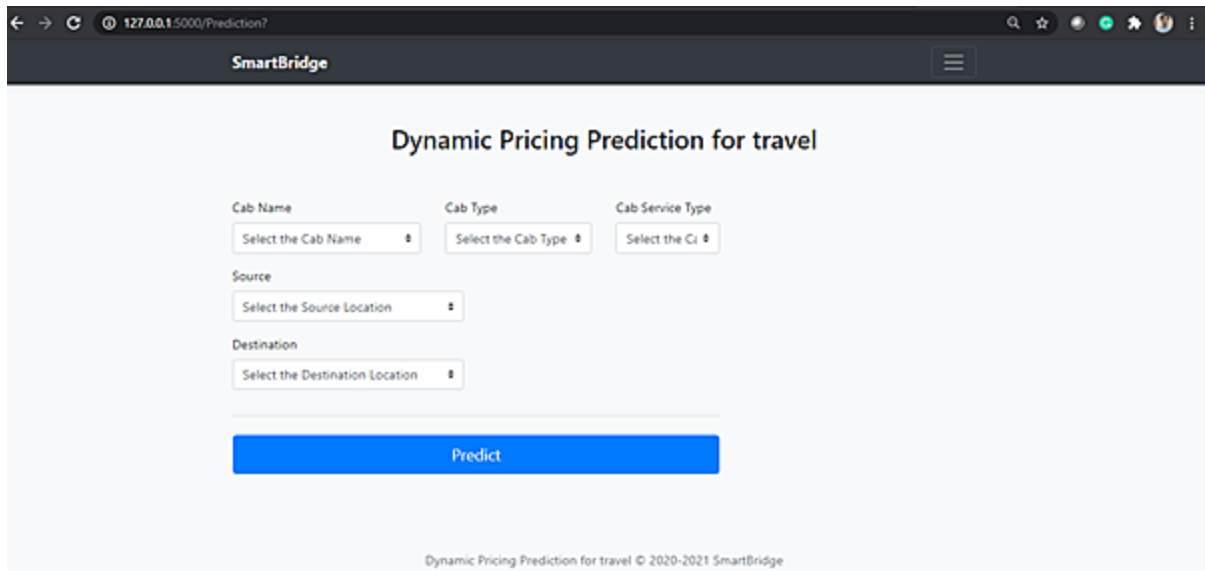
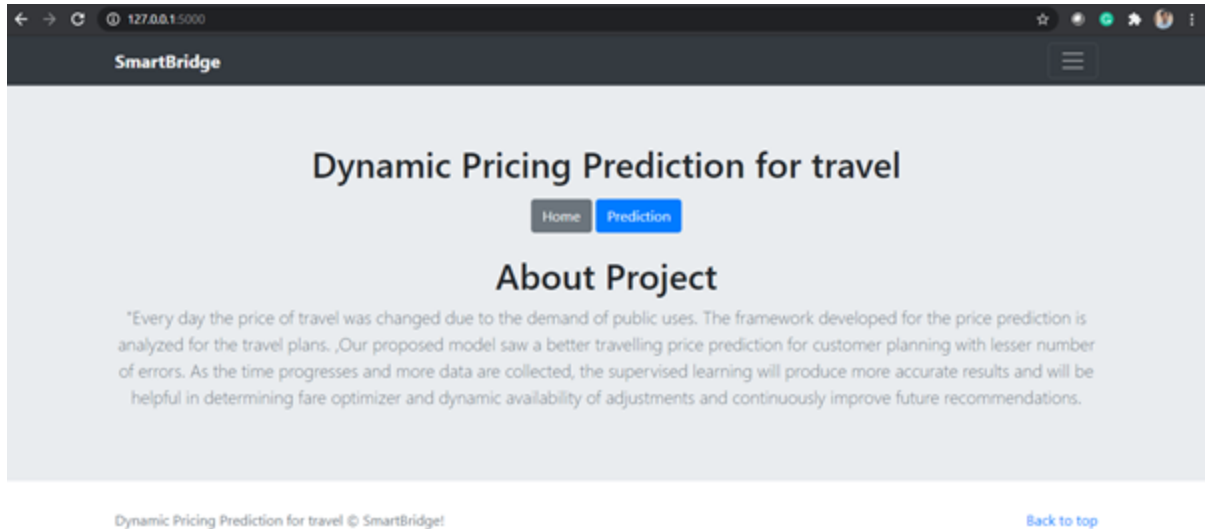
```
#Weather data represents the temperature ,humidity etc.
weather_df.head()
```

	temp	location	clouds	pressure	rain	time_stamp	humidity	wind
0	42.42	Back Bay	1.0	1012.14	0.1228	1545003901	0.77	11.25
1	42.43	Beacon Hill	1.0	1012.15	0.1846	1545003901	0.76	11.32
2	42.50	Boston University	1.0	1012.15	0.1089	1545003901	0.76	11.07
3	42.11	Fenway	1.0	1012.13	0.0969	1545003901	0.77	11.09
4	43.13	Financial District	1.0	1012.14	0.1786	1545003901	0.75	11.49

ARCHITECTURE:



RESULT:





ADVANTAGES & DISADVANTAGES:

Advantages:

1. They need to work more on their network.
2. There is a lack of communications with the drivers.
3. Cabs brings a whole lot of exclusive perks for corporate employees.
4. Insurance
5. Your safety
6. Pricing automation.

Disadvantages:

1. There are payment issues in this cab service.
2. You have a big responsibility for your passengers.

3. Cab drivers don't make much money.
4. You don't have regular schedule.
5. Cab drivers have to work at nighttime
6. Some cab drivers even have to work on weekends and holidays.
7. Can be problematic for your family life.
8. Drivers sometimes try to game the system for bigger profits

CONCLUSION:

Every day the price of travel was changed due to the demand for public uses. The framework developed for the price prediction is analyzed for the travel plans. For the same travel plan offered at a fixed price for a particular group of customers, our proposed model saw a final fare with a lesser number of errors in predicting customer planning. As time progresses and more data are collected, the supervised learning will produce more accurate results and will be helpful in determining fare optimizer and dynamic availability of adjustments and continuously improve future recommendations.

FUTURE SCOPE:

- The dependent algorithm variable turned out to be 'ride fare,' while 'travel distance' and 'travel time' are the independent variables.
- The linear regression model for prediction of dynamic price of trips is providing an efficiency of 93.40%.
- It is better suited for the prediction of target variable which is trip fare, and it performs very well.

- Further this work can be carried out using different machine learning algorithms and techniques in order to get higher efficiency and lower errors.

BIBLIOGRAPHY:

<https://www.youtube.com/watch?v=5mDYijMfSzs>