Uber Data Analysis with Python

GROUP MEMBERS:-

Contents

1 Introduction to Python

2 What is Machine Learning?

3 Types Of Machine Learning

4 What is Supervised Learning?

5 What is Linear Regression?

6 ABSTRACT :

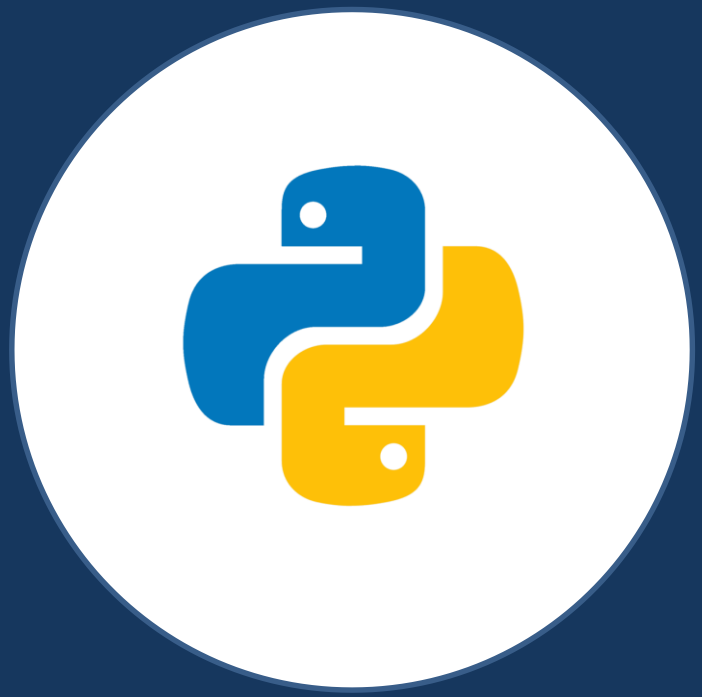
7 Introduction

8 REQUIREMENTS

9 System used and conclusion

Introduction to Python

Python is a popular high level, object oriented and interpreted language



High level

Interpreted

Object oriented

What is Machine Learning?

Machine learning is a sub-set of artificial intelligence (AI) that allows the system to automatically learn and improve from experience without being explicitly programmed

	Time	V1	V2	V3	V4	V5
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193

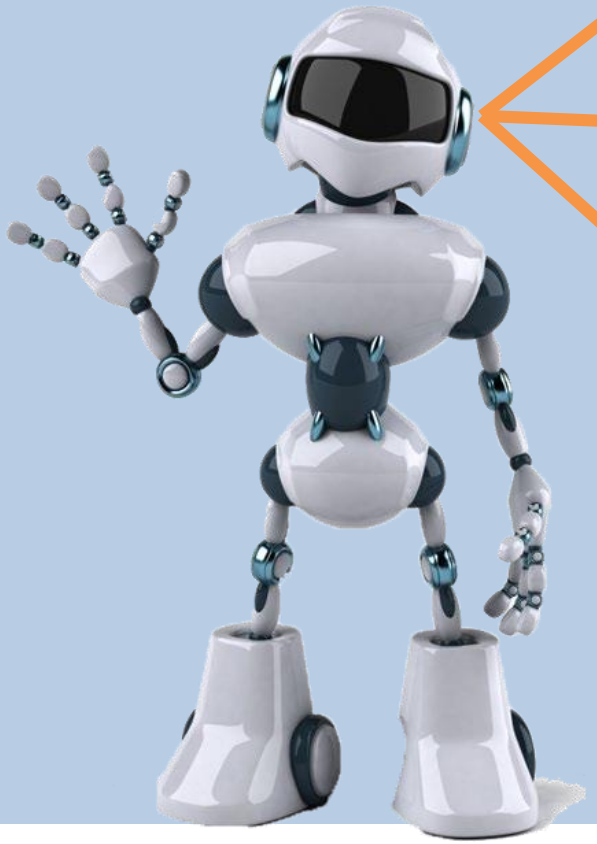
Training Data

Model Building

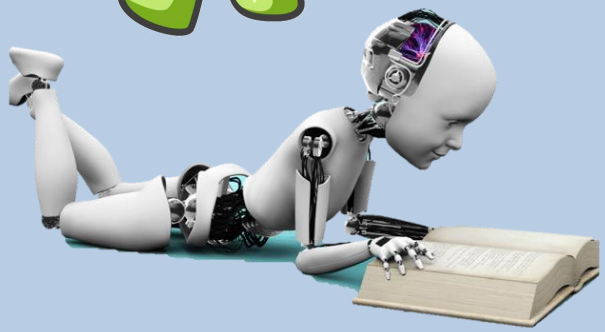
	Time	V1	V2	V3	V4
284802	172786.0	-11.881118	10.071785	-9.834783	-2.066656
284803	172787.0	-0.732789	-0.055080	2.035030	-0.738589
284804	172788.0	1.919565	-0.301254	-3.249640	-0.557828
284805	172788.0	-0.240440	0.530483	0.702510	0.689799
284806	172792.0	-0.533413	-0.189733	0.703337	-0.506271

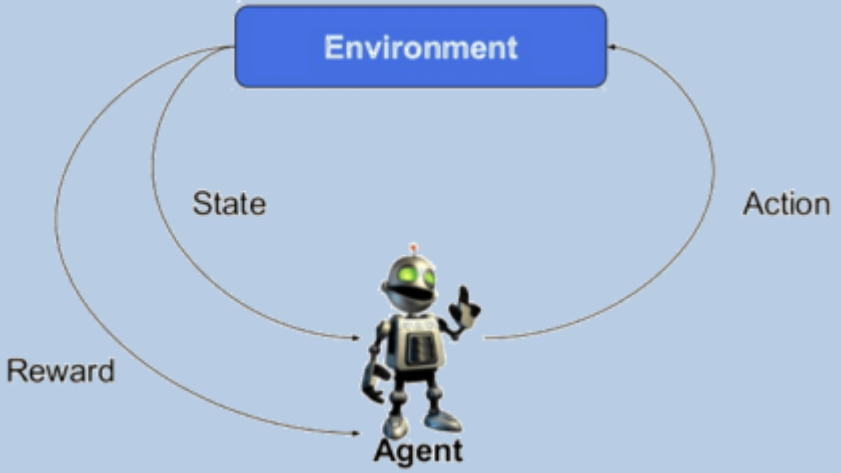
Testing Data

Types Of Machine Learning



- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning





What is Supervised Learning?

Supervised learning works as a supervisor or teacher. Basically, In supervised learning, we teach or train the machine with labeled data (that means data is already tagged with some predefined class). Then we test our model with some unknown new set of data and predict the level for them

- Learning from the labelled data and applying the knowledge to predict the label of the new data(test data), is known as ***Supervised Learning***
- ***Types of Supervised Learning:***
  - **Linear Regression**
  - **Logistic regression**
  - **Decision Tree**
  - **Random Forest**
  - **Naïve Bayes Classifier**

What is Linear Regression?

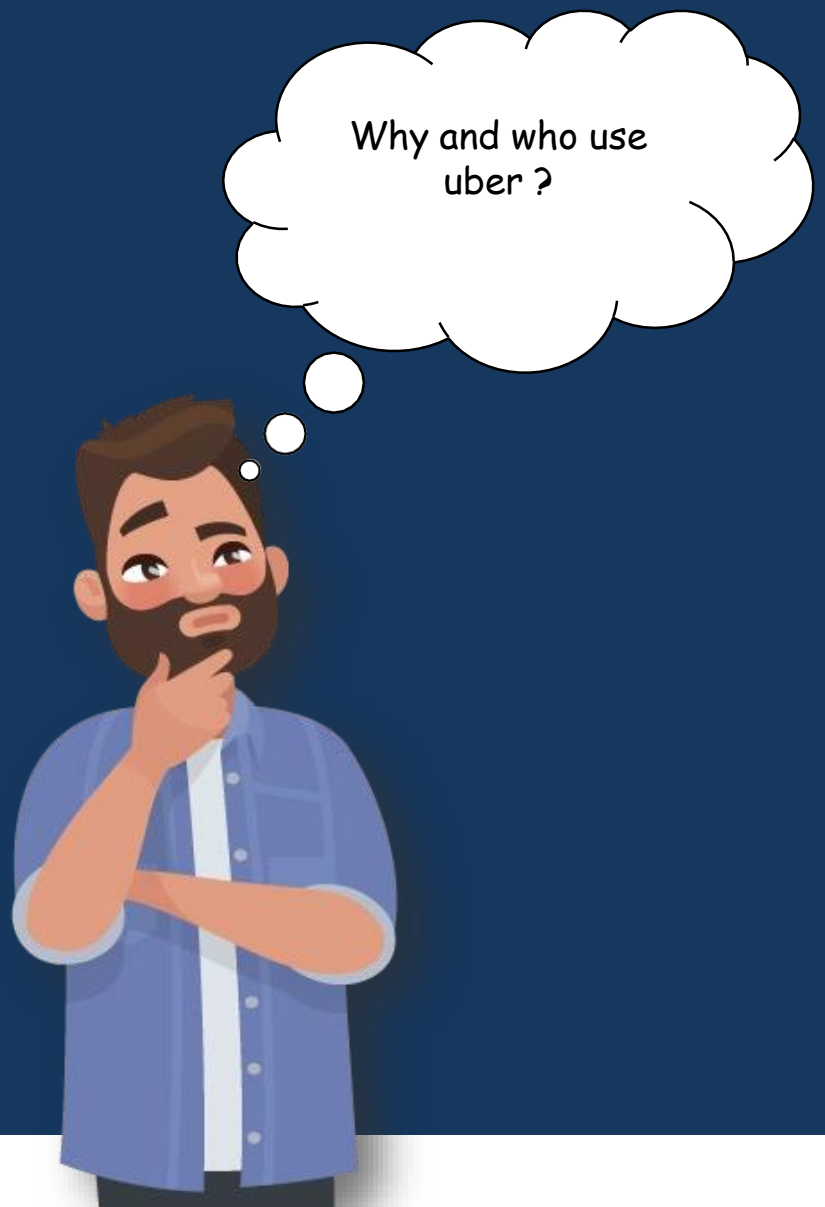
Regression stands for to model a target value based on independent variables and Linear Regression is used to find the relationship between dependent(y) and independent variable(x)

- Linear regression is a supervised machine learning algorithm
- Always works with continuous value
- Formula: y = mx+c m=slope of line and c= intercept
- Main target for linear regression to find the best value for X and Y

ABSTRACT :

Uber was founded just eleven years ago, and it was already one of the fastest-growing companies in the world. In Delhi *, UberX claims to charge 30% less than taxis – a great way to get customers' attention. Nowadays, we see applications of Machine Learning and Artificial Intelligence in almost all the domains so we try to use the same for Uber cabs price prediction. In this project, we did experiment with a real-world dataset and explore how machine learning algorithms could be used to find the patterns in data. We mainly discuss about the price prediction of different Uber cabs that is generated by the machine learning algorithm. Our problem belongs to the regression supervised learning category. We use different machine learning algorithms, for example, Linear Regression, Decision Tree, Random Forest Regressor, and Gradient Boosting Regressor but finally, choose the one that proves best for the price prediction. We must choose the algorithm which improves the accuracy and reduces overfitting. We got many experiences while doing the data preparation of Uber Dataset of* Delhi *of the year 2015.*

Introduction



- Uber has been a major source of travel for people living in urban areas. Some people don’t have their vehicles while some don’t drive their vehicles intentionally because of their busy schedule. So different kinds of people are using the services of Uber and other taxi services.

Introduction

By analyzing Uber trips, we can draw many patterns like which day has the highest and the lowest trips or the busiest hour for Uber and many other patterns. The dataset we are using here is based on Uber trips from Delhi , with a very complex transportation system with a large residential community.

REQUIREMENTS

ENVIRONMENT AND TOOLS:

1. Google colab
2. Numpy
3. Pandas
4. Seaborn
5. Matplotlib

SOFTWARE:

Platform – Windows, Linux or MacOS
Operating system – Window, Linux or MacOS
Technology – python machine learning
Scripting language – Python
IDE – Google colab

conclusion

By analyze the Uber trips for Delhi. Some of the conclusions that we got from this analysis are:

- Monday is the most profitable day for Uber
- On Saturdays less number of people use Uber
- 6 pm is the busiest day for Uber
- On average a rise in Uber trips start around 5 am.
- Most of the Uber trips originate near the Manhattan region in Delhi.

System used

Google colab

Uber Data Analysis

Thank You