Capstone Project

Bike Sharing Demand Prediction

Yuvaraj Mahajan
Yashwany Rawat

Table Of Contents

Problem Statement

- Currently Rental bikes are introduced in many urban cities for the enhancement of mobility comfort. It is important to make the rental bike available and accessible to the public at the right time as it lessens the waiting time.

- Eventually, providing the city with a stable supply of rental bikes becomes a major concern. The crucial part is the prediction of bike count required at each hour for the stable supply of rental bikes.

Data Summary

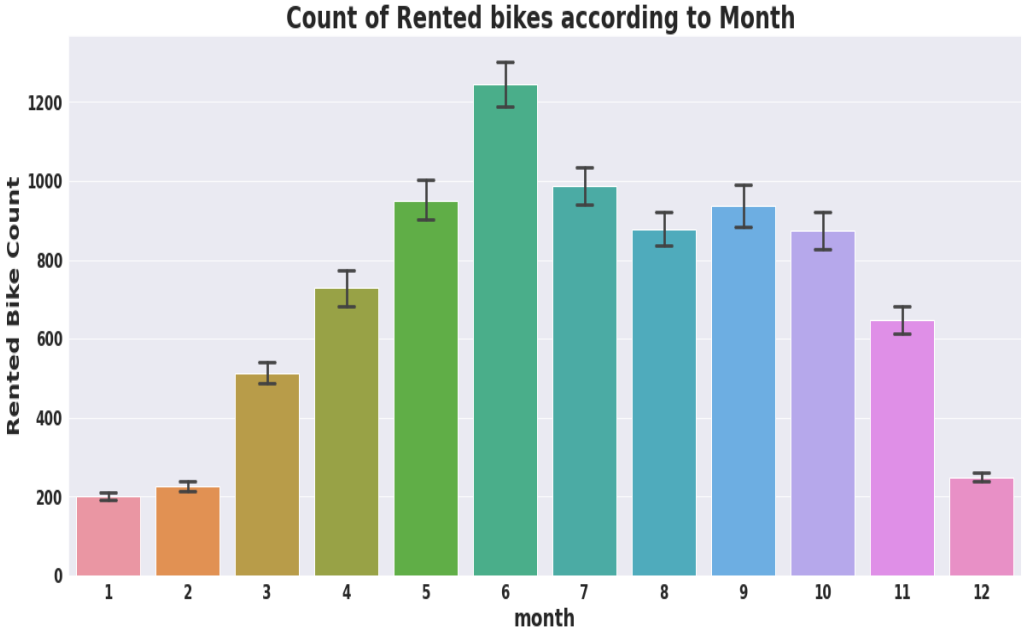
- **Date** : The date of the day, during 365 days from 01/12/2017 to 30/11/2018, formating in DD/MM/YYYY, type : str, we need to convert into datetime format.
- **Rented Bike Count** : Number of rented bikes per hour which our dependent variable and we need to predict that, type : int
  Hour : The hour of the day, starting from 0-
  23 it`s in a digital time format, type : int, we need to convert it into category data type.
- **Temperature(°C)** : Temperature in Celsius, type : Float
- **Humidity(%)** : Humidity in the air in %, type : int
- **Wind speed (m/s)** : Speed of the wind in m/s, type : Float
- **Visibility (10m)** : Visibility in m, type : int
- **Dew point temperature(°C)** : Temperature at the beggining of the day, type : Float
- **Solar Radiation (MJ/m2)** : Sun contribution, type : Float
- **Rainfall(mm)** : Amount of raining in mm, type : Float
- **Snowfall (cm)** : Amount of snowing in cm, type : Float
  Seasons : Season of the year, type : str, there are only 4 season's in data .
- **Holiday** : If the day is holiday period or not, type: str
- **Functioning Day** : If the day is a Functioning Day or not, type : str
- Sum

Data Wrangling

- Columns and their unique values to understand what they contain
- Handling missing values.
- Breaking date column
- Changing data type

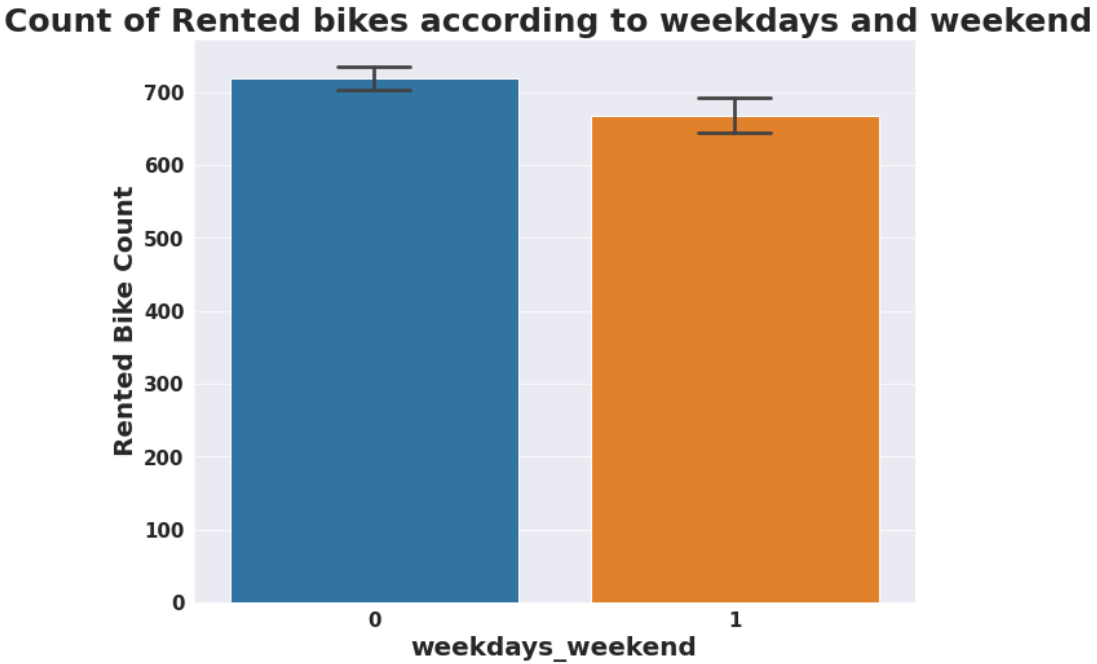
Bike season wise analysis

Count of Rented bikes according to Month



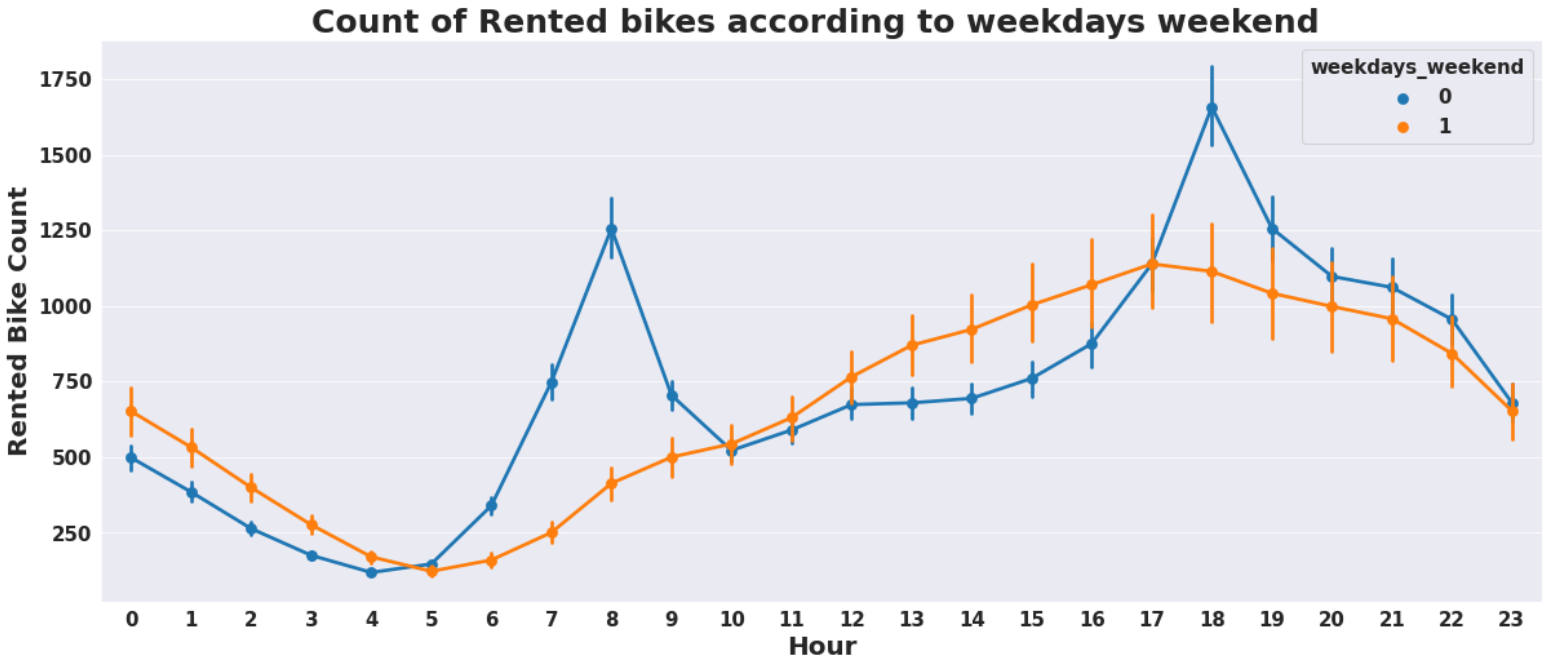
From the bar chart above, it is clear from the 5th to 10th months that the demand for rental bicycles is higher than in other months. These months are the summer season

Count of Rented bikes according to weekdays and weekend

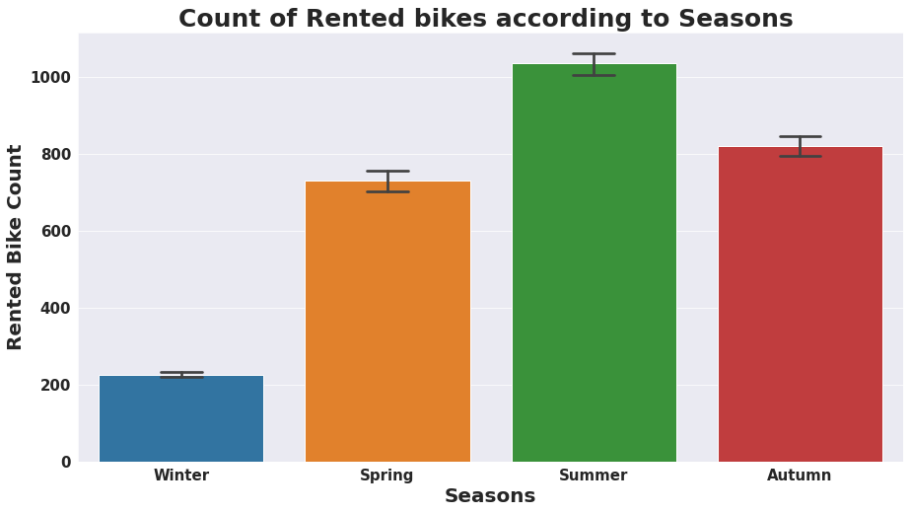


In Weekdays are more use of rental bike booking compare to weekend. blue bar plot are shown weekdays and brown bar plot are shown result of weekend

Count of Rented bikes according to weekdays weekend



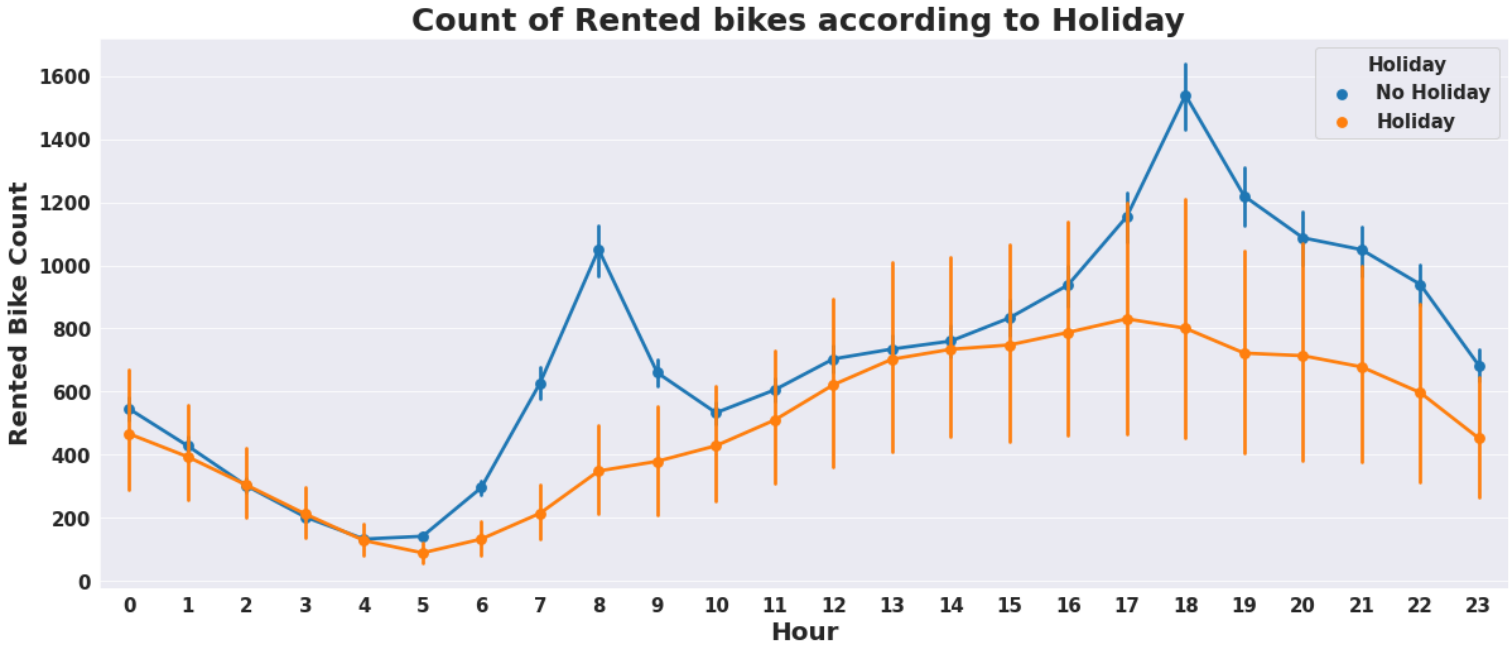
Count of Rented bikes according to Seasons



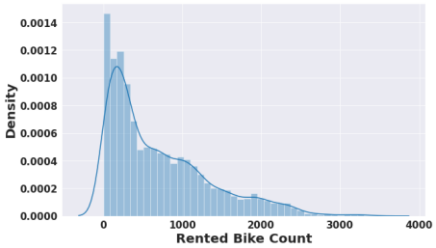
•The bar and dot charts above show the use of rental bicycles in four different seasons and clearly show that:

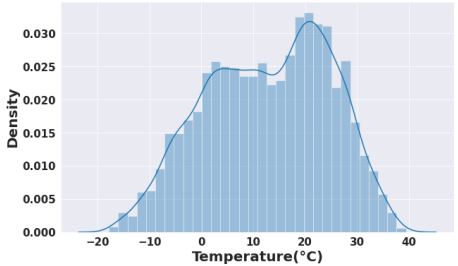
- Rental bicycles are highly used in the summer, with peak hours from 7:00 to 9:00 and 19:00 to 17:00.
- Due to snowfall in winter, rental bicycles are rarely used.

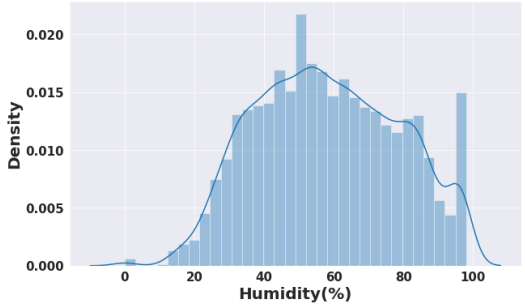
Count of Rented bikes according to Holiday

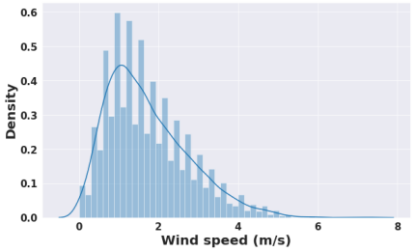


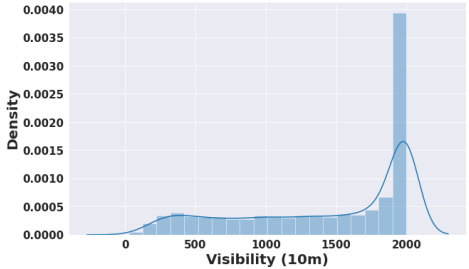
Analyze of Numerical variables distplots

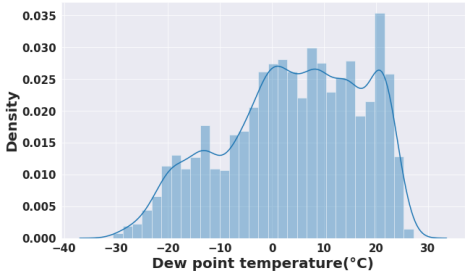


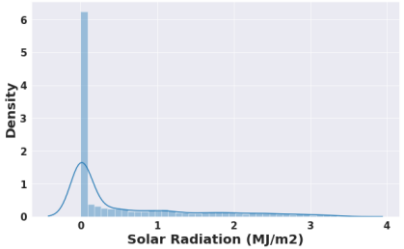


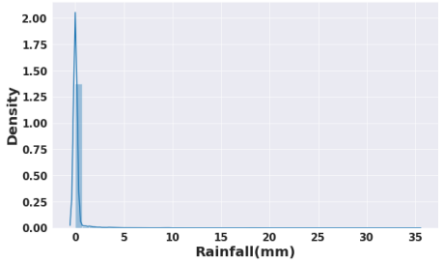


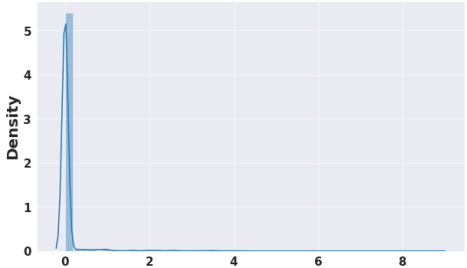








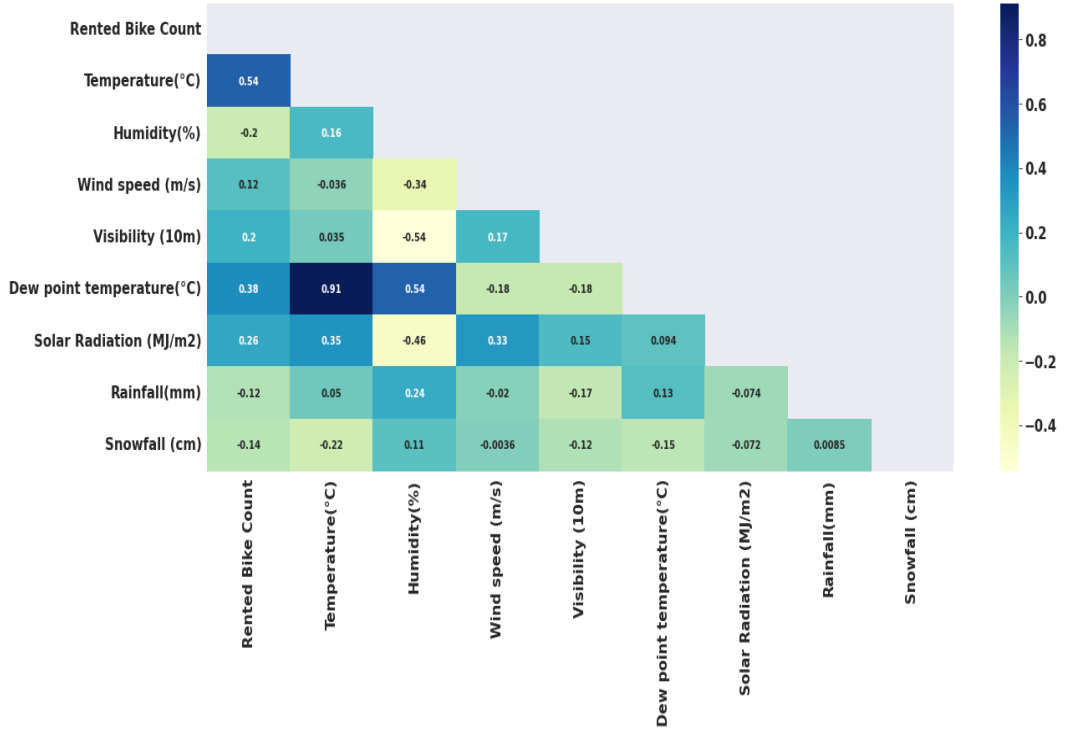




Correlation

	Rented Bike Count	Temperature(°C)	Humidity(%)	Wind speed (m/s)	Visibility (10m)	Dew point temperature(°C)	Solar Radiation (MJ/m2)	Rainfall(mm)	Snowfall (cm)
Rented Bike Count	1.000000	0.538558	-0.199780	0.121108	0.199280	0.379788	0.261837	-0.123074	-0.141804
Temperature(°C)	0.538558	1.000000	0.159371	-0.036252	0.034794	0.912798	0.353505	0.050282	-0.218405
Humidity(%)	-0.199780	0.159371	1.000000	-0.336683	-0.543090	0.536894	-0.461919	0.236397	0.108183
Wind speed (m/s)	0.121108	-0.036252	-0.336683	1.000000	0.171507	-0.176486	0.332274	-0.019674	-0.003554
Visibility (10m)	0.199280	0.034794	-0.543090	0.171507	1.000000	-0.176630	0.149738	-0.167629	-0.121695
Dew point temperature(°C)	0.379788	0.912798	0.536894	-0.176486	-0.176630	1.000000	0.094381	0.125597	-0.150887
Solar Radiation (MJ/m2)	0.261837	0.353505	-0.461919	0.332274	0.149738	0.094381	1.000000	-0.074290	-0.072301
Rainfall(mm)	-0.123074	0.050282	0.236397	-0.019674	-0.167629	0.125597	-0.074290	1.000000	0.008500
Snowfall (cm)	-0.141804	-0.218405	0.108183	-0.003554	-0.121695	-0.150887	-0.072301	0.008500	1.000000

Heatmap



We can observe on the heatmap that on the target variable line the most positively correlated variables to the rent are :

- **the temperature**
- **the dew point temperature**
- **the solar radiation**

And most negatively correlated variables are:

- **Humidity**
- **Rainfall**
- From the correlation heatmap above, you can see that there is a positive correlation between the Temperature and Dew Point columns. H. 0.91. Therefore, omitting this column does not affect the results of the analysis. Also, since the fluctuation is the same, the "Dew point temperature (° C)" column can be omitted

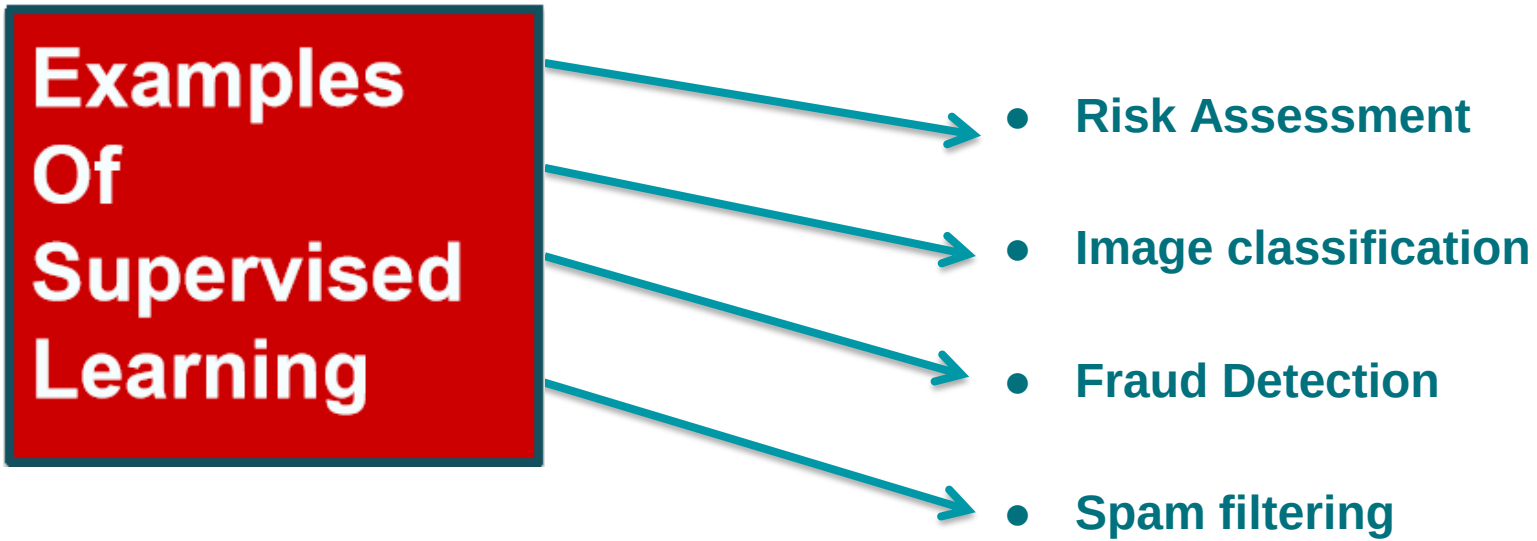
AI

What is Supervised learning

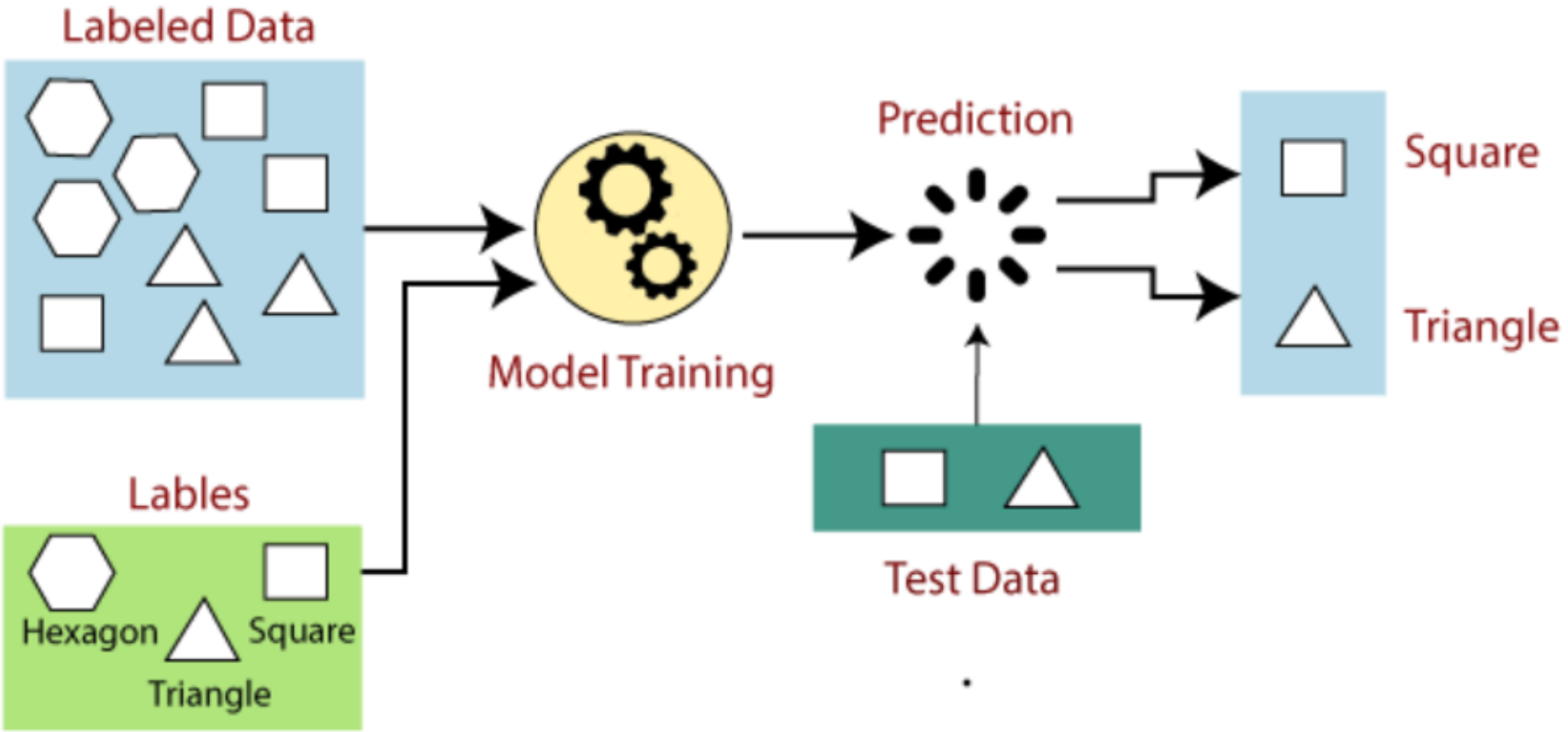
- Supervised learning is the types of machine learning in which machines are trained using well "labelled" training data, and on basis of that data, machines predict the output. The labelled data means some input data is already tagged with the correct output.

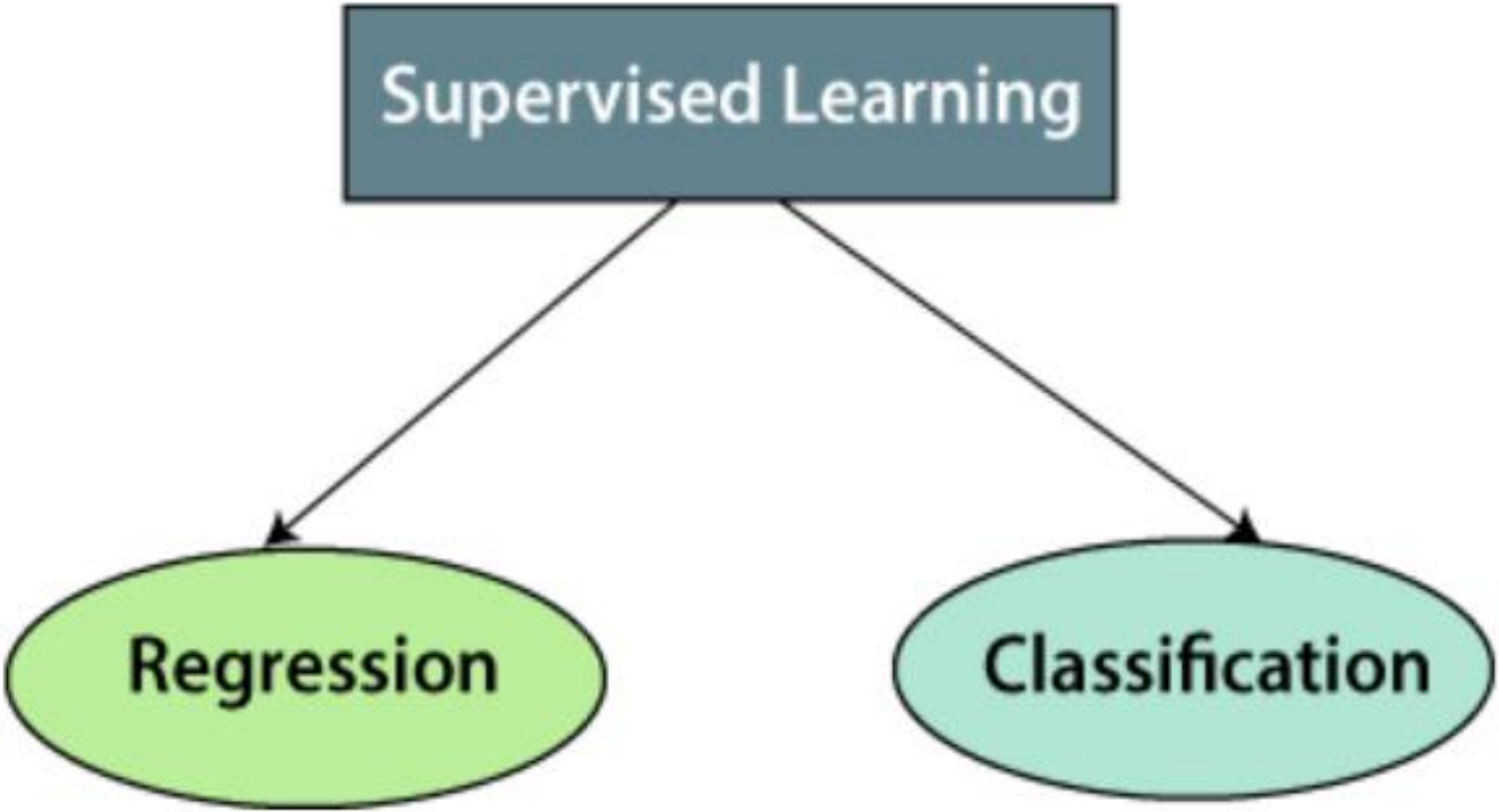
- In supervised learning, the training data provided to the machines work as the supervisor that teaches the machines to predict the output correctly. It applies the same concept as a student learns in the supervision of the teacher.

- Supervised learning is a process of providing input data as well as correct output data to the machine learning model. The aim of a supervised learning algorithm is to find a mapping function to map the input variable(x) with the output variable(y)


Examples
Of
Supervised
Learning
Risk Assessment
Image classification
Fraud Detection
Spam filtering

Working of supervised learning




Supervised Learning
Regression
Classification

Regression

Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as Weather forecasting, Market Trends, etc. Below are some popular Regression algorithms which come under supervised learning:

- Linear Regression
- Regression Trees
- Non-Linear Regression
- Bayesian Linear Regression
- Polynomial Regression

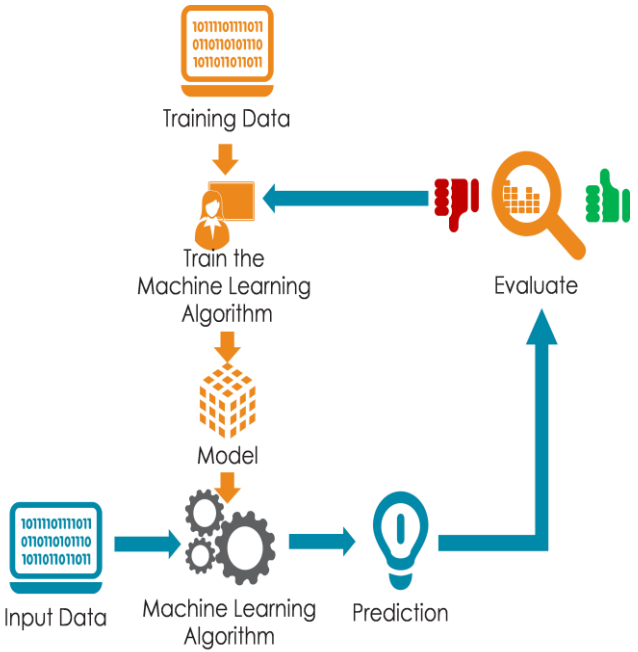
Feature Engineering

Feature Engineering is a machine learning technique that leverages data to create new variable that aren't in the training set . We produce new features with the goal of simplifying and speeding up data transformation while also enhancing model accuracy.

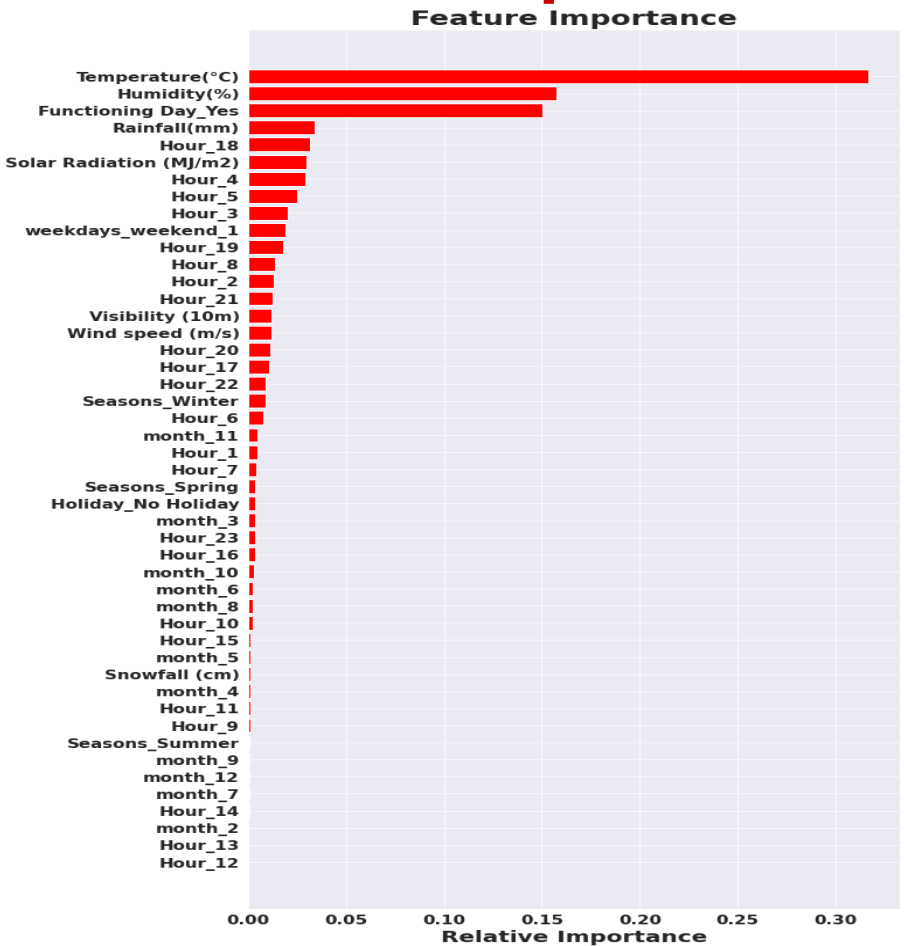
- Rented Bike Count
- Hour
- Temperature(°C)
- Humidity(%)
- Wind speed (m/s)
- Visibility (10m):
- Dew point temperature(°C)
- Solar Radiation
- Rainfall(mm):
- Snowfall
- Seasons
- Holiday
- Functioning Day

Model Training

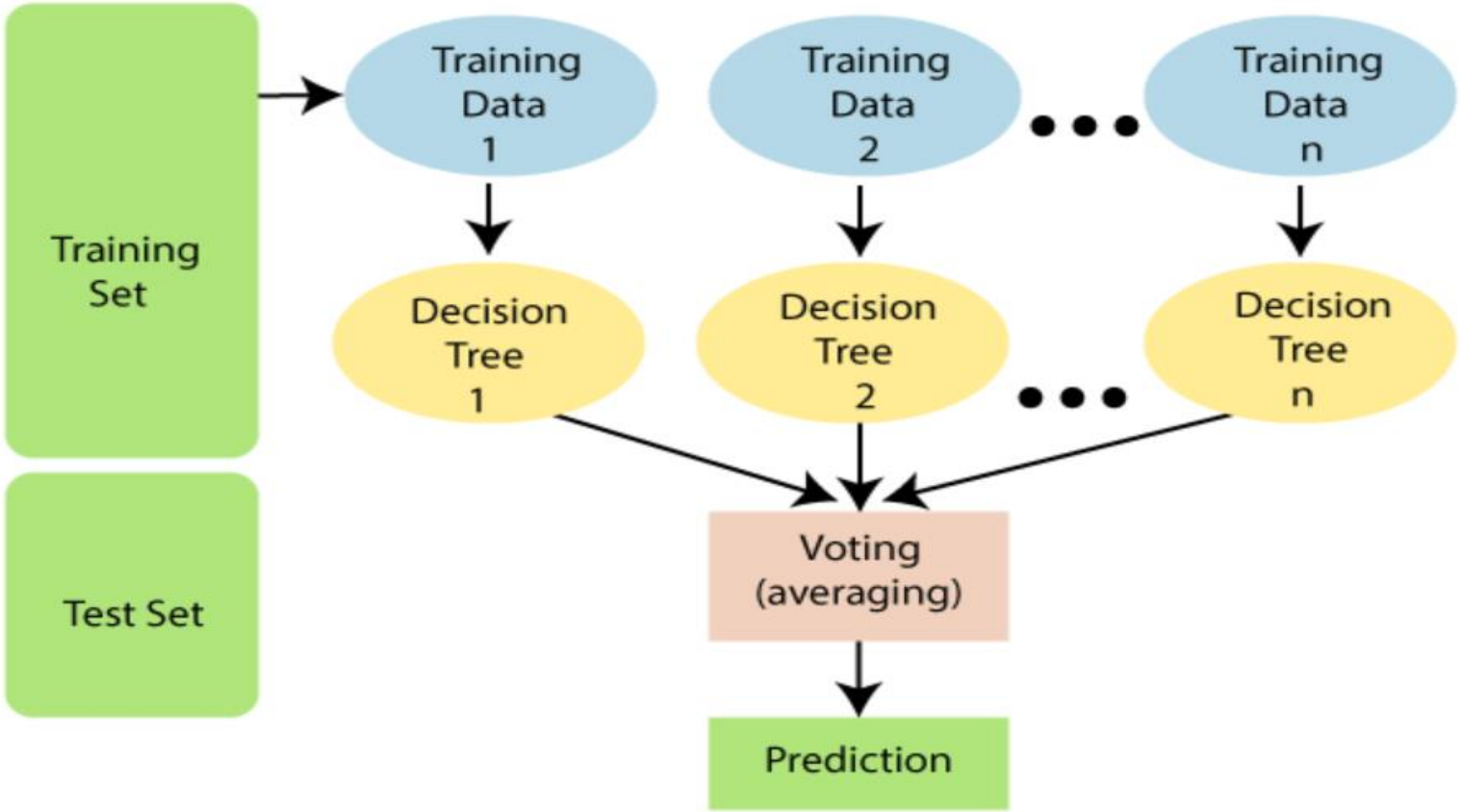
- Model training is the process of fitting a data into machine learning model from which model learns the patterns in data to predict the dependent variable. Model do it so by assigning a weight to each variable.
- After our model is trained we test our model on test data to check how our model is performing.

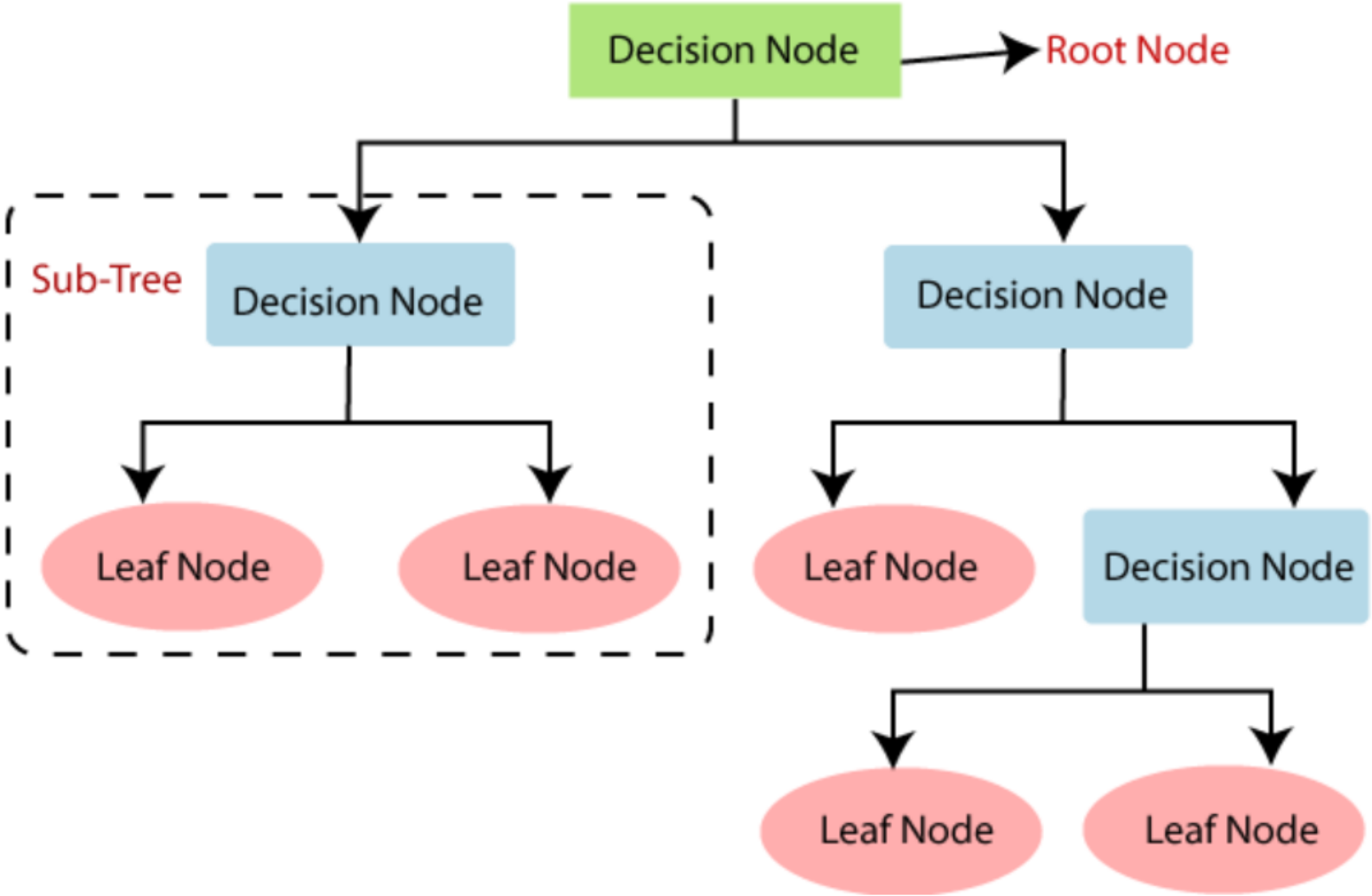


Feature Importance



Random Forest Algorithm




Decision Node
Root Node
Sub-Tree
Decision Node
Decision Node
Leaf Node
Leaf Node
Leaf Node
Decision Node
Leaf Node
Leaf Node

Challenges

- To find the Season wise bike rental booking
- Feature Engineering
- Selecting a features to train a model.
- Model training, tuning and performance improvement.

Conclusion

- During the analysis, we first ran EDA for all the features of the dataset. First, we analyzed the dependent variable "Rented Bike Count" and converted it.
- Then I analyzed the categorical variables and removed the variables with the majority of the class. We also analyzed numerical variables to find out their correlations, distributions, and relationships with dependent variables.
- We've also removed some numeric features, most of which have a value of 0, and hot coded categorical variables. But this is not the final end.
- Because this data is time-dependent, the values of variables such as temperature, wind speed, and amount of solar radiation are not always consistent. Therefore, there are scenarios where the model may not work well.
- Machine learning is an exponentially evolving field, so you need to check your model from time to time in case of any contingency.
- Therefore, having quality knowledge and keeping up with the ever-evolving ML field will certainly help us move forward in the future.

Thank You