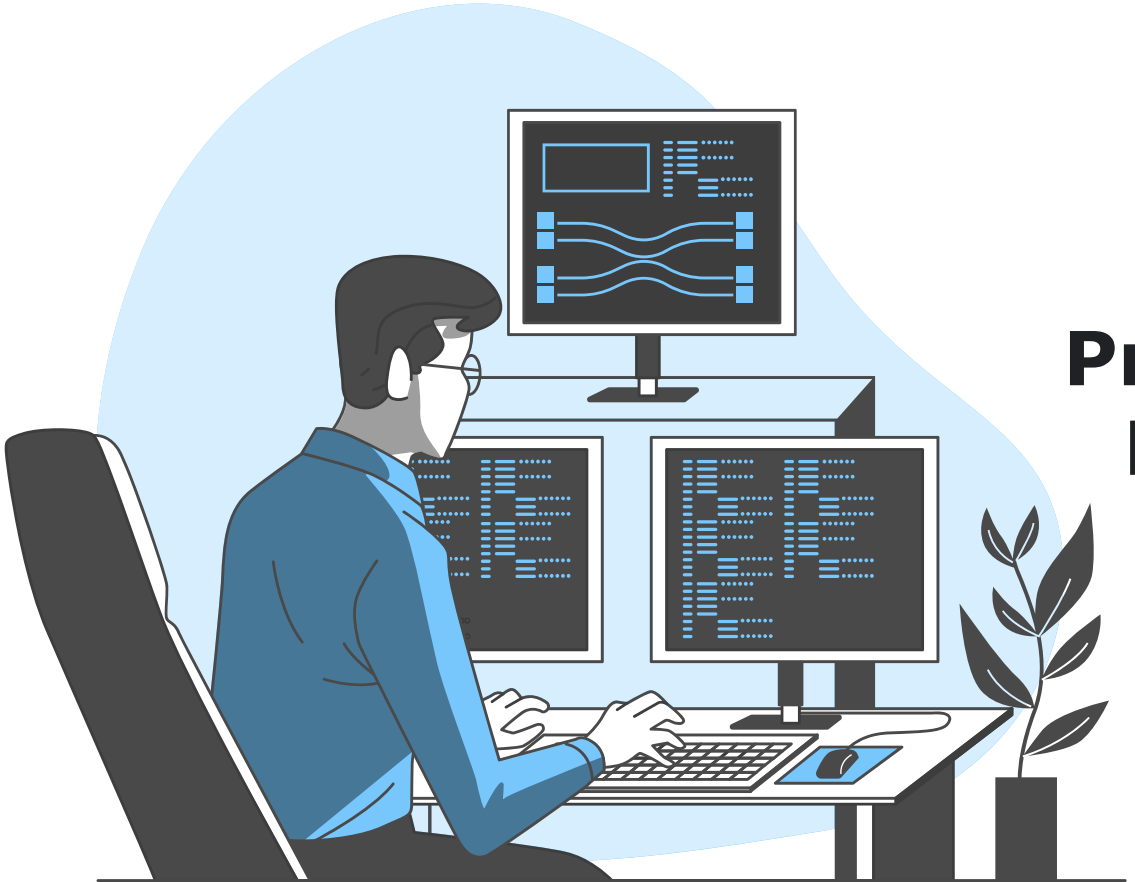


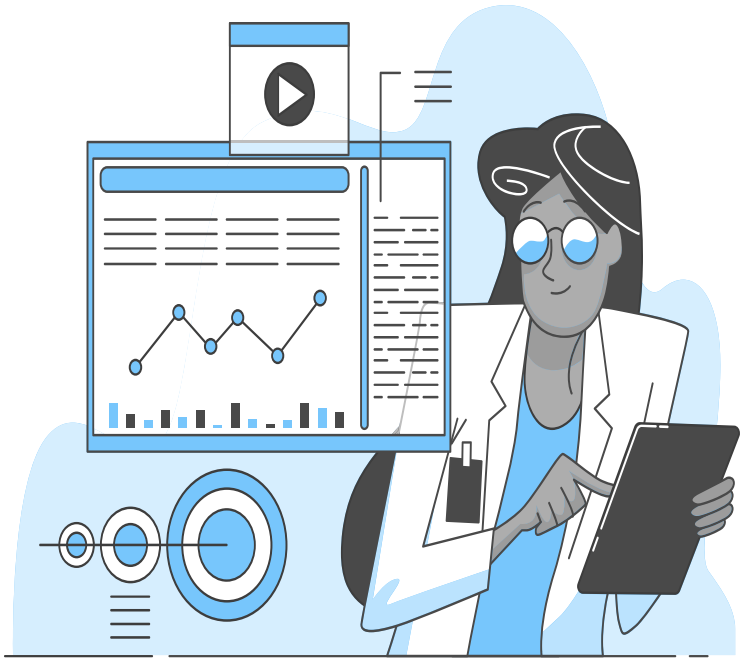
Restaurant Revenue Prediction Using Decision Tree Modelling

By Ankit B Raut

Ivy Professional School, Kolkata

Agenda

1. Objectives
2. Introduction
3. Dataset Dictionary
4. Technology Used
5. System Design
6. Methodology
7. Experiment And Results
8. Continuous & Categorical Data
9. Decision Tree Model
10. Conclusion



Objectives

This dataset contains information about various restaurants and aims to predict the revenue based on several features.

Build a Supervised machine Learning Model for forecasting value of restaurants Revenue based on multiple attributes.

Provide graphical comparisons to provide better view.

Introduction

This is dataset which is including data about Restaurants different features. Dataset shows Which Cuisines, Locations, Average Meal Price etc. Influence factors are affecting in Restaurant’s Revenue Prediction.

Each row represents a unique restaurant with various attributes that may influence its revenue.

Dataset Dictionary

- **Name**: The name of the restaurant.
- **Location**: The location of the restaurant (e.g., Rural, Downtown).
- **Cuisine**: The type of cuisine offered (e.g., Japanese, Mexican, Italian).
- **Rating**: The average rating of the restaurant.
- **Seating Capacity**: The number of seats available in the restaurant.
- **Average Meal Price**: The average price of a meal at the restaurant.
- **Marketing Budget**: The marketing budget allocated for the restaurant.
- **Social Media Followers**: The number of social media followers.
- **Chef Experience Years**: The number of years of experience of the head chef.
- **Number of Reviews**: The total number of reviews the restaurant has received.
- **Avg Review Length**: The average length of reviews.
- **Ambience Score**: A score representing the ambience of the restaurant.
- **Service Quality Score**: A score representing the quality of service.
- **Parking Availability**: Indicates if parking is available (Yes/No).
- **Weekend Reservations**: The number of reservations made on weekends.
- **Weekday Reservations**: The number of reservations made on weekdays.
- **Revenue**: The total revenue generated by the restaurant.

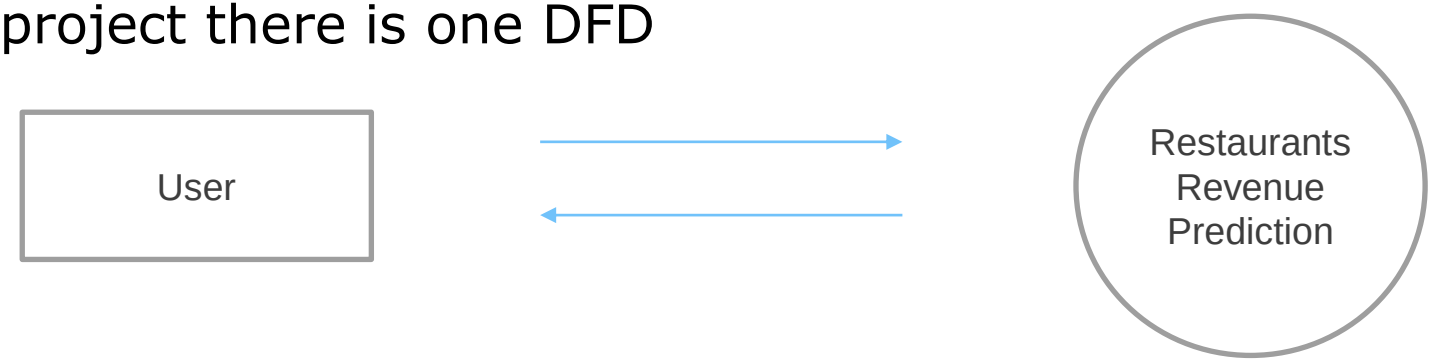
Technology Used

Decision Tree:A decision tree prediction model is a type of supervised learning algorithm used for both classification and regression tasks. It models decisions and their possible consequences in a tree-like graph structure.

Linear Regression: Regression is a method for predicting a dependent component with the help of independent variables. The method is commonly used to predict and calculate correlations between independent and dependent variables. The regression model establishes a linear or exponential connection between independent and dependent variables.

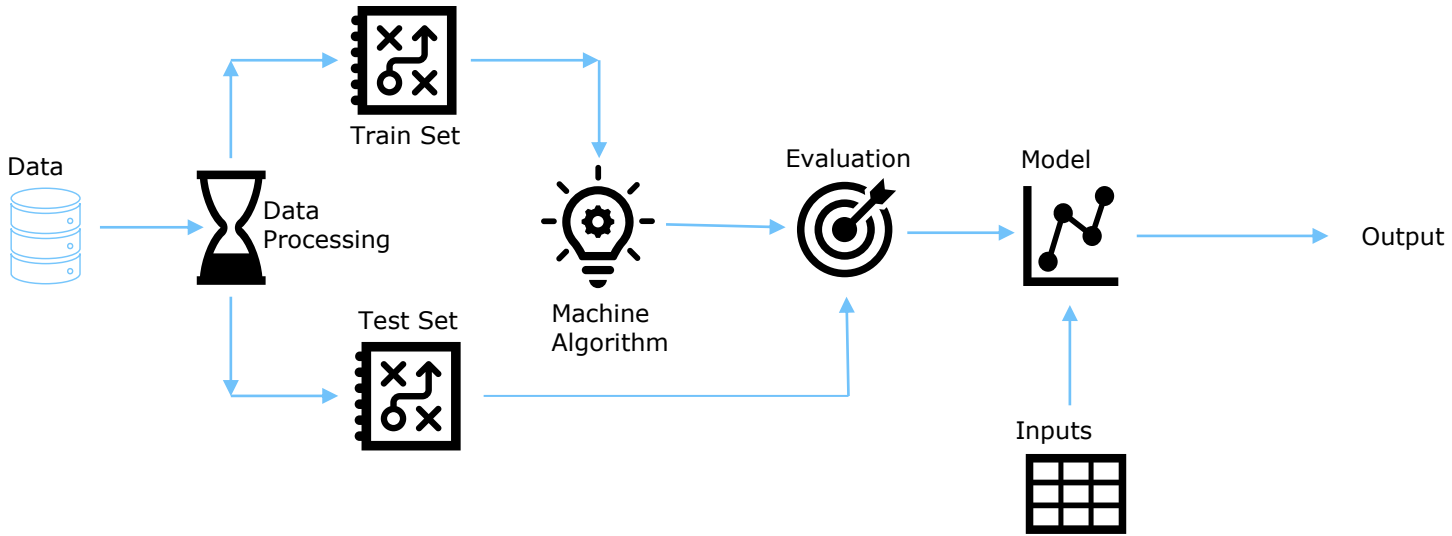
System Design

A data flow diagram shows the way information flows through a process or system. It includes data inputs and outputs; data stores and the various sub processes the data moves through. DFDs are built using standardized symbols and notation to describe various entities and their relationships. In this project there is one DFD

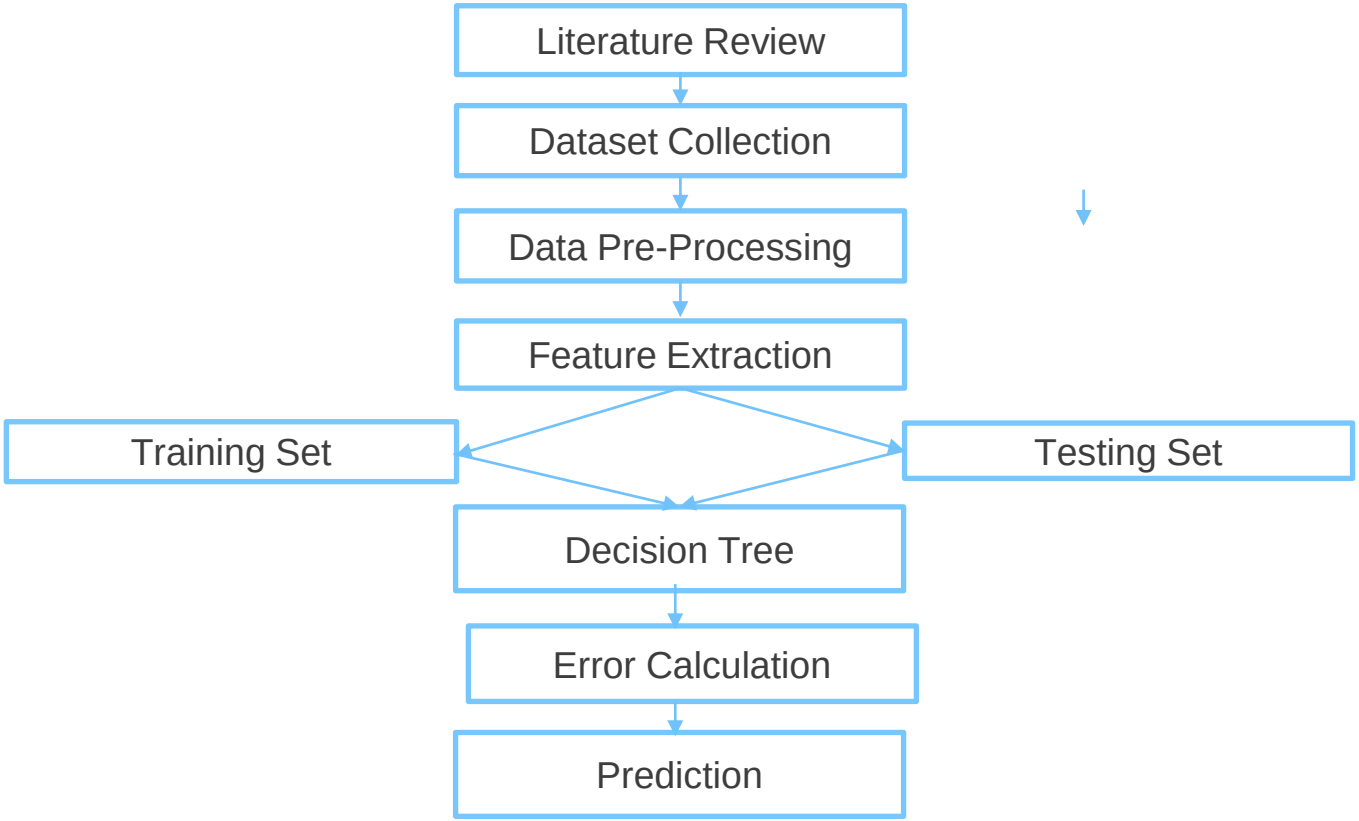


System Design

Different processes are carried out to obtain the actual information such as one hot encoding. Then feature extraction is performed to extract necessary features. User can give an input to the detection model and it will provide an output.



Methodology



Experiment and Results

The dataset is imported and read for understanding. This dataset is taken from Kaggle.

```
resData.head(5)
```

```
Shape before deleting duplicate values: (8368, 17)
```

[1]:

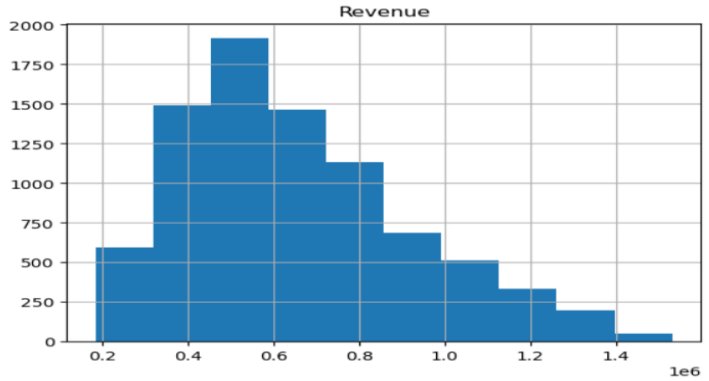
	Name	Location	Cuisine	Rating	Seating Capacity	Average Meal Price	Marketing Budget	Social Media Followers	Chef Experience Years	Number of Reviews	Avg Review Length	Ambience Score	Service Quality Score	Parking Availability	Weekend Reservations	R
0	Restaurant 0	Rural	Japanese	4.0	38	73.98	2224	23406	13	185	161.924906	1.3	7.0	Yes	13	
1	Restaurant 1	Downtown	Mexican	3.2	76	28.11	4416	42741	8	533	148.759717	2.6	3.4	Yes	48	
2	Restaurant 2	Rural	Italian	4.7	48	48.29	2796	37285	18	853	56.849189	5.3	6.7	No	27	
3	Restaurant 3	Rural	Italian	4.4	34	51.55	1167	15214	13	82	205.433265	4.6	2.8	Yes	9	
4	Restaurant 4	Downtown	Japanese	4.9	88	75.98	3639	40171	9	78	241.681584	8.6	2.1	No	37	

```
resData.isnull().sum()
```

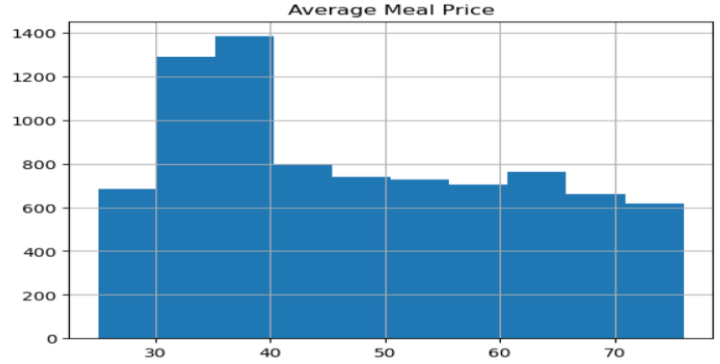
```
Name                        0
Location                    0
Cuisine                     0
Rating                      0
Seating Capacity            0
Average Meal Price          0
Marketing Budget            0
Social Media Followers      0
Chef Experience Years       0
Number of Reviews           0
Avg Review Length           0
Ambience Score              0
Service Quality Score       0
Parking Availability        0
Weekend Reservations        0
Weekday Reservations        0
Revenue                     0
dtype: int64
```

Visualizing Continuous Data

```
%matplotlib inline
# Creating Bar chart as the Target variable is Continuous
resData.hist(['Revenue'])
```




```
resData.hist(['Average Meal Price'])
```



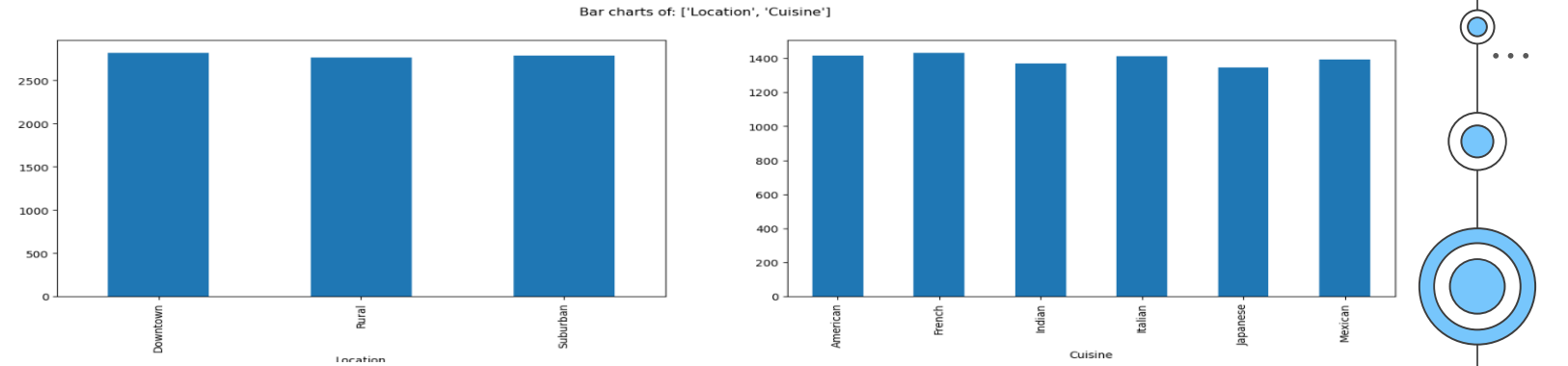
Visualizing Categorical Data

```
def PlotBarCharts(inpData, colsToPlot):
    %matplotlib inline

    import matplotlib.pyplot as plt

    # Generating multiple subplots
    fig, subPlot=plt.subplots(nrows=1, ncols=len(colsToPlot), figsize=(20,5))
    fig.suptitle('Bar charts of: '+ str(colsToPlot))

    for colName, plotNumber in zip(colsToPlot, range(len(colsToPlot))):
        inpData.groupby(colName).size().plot(kind='bar',ax=subPlot[plotNumber])
PlotBarCharts(inpData=resData, colsToPlot=[ 'Location','Cuisine'])
```



Decision Tree Model

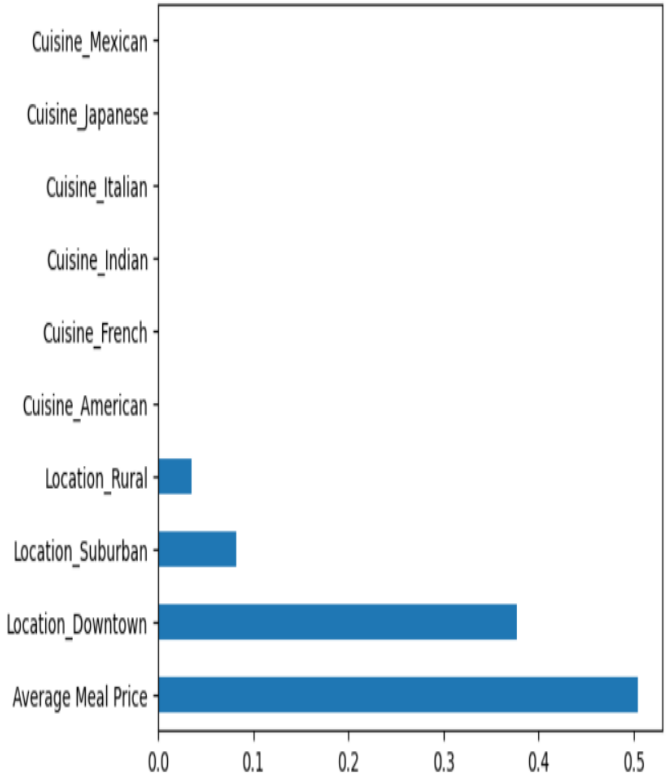
Decision Tree Model which is showing better R2 value than Linear Regression Model.

```
DecisionTreeRegressor(max_depth=4)
R2 Value: 0.9195020149317672

##### Model Validation and Accuracy Calculations ##########
      Revenue  PredictedRevenue
0   704643.06          571433.0
1   543490.90          539140.0
2   697555.64          610814.0
3   568947.60          571433.0
4  1161144.85         1055600.0
Mean Accuracy on test data: 89.40968762826893
Median Accuracy on test data: 91.13354574635768

Accuracy values for 10-fold Cross Validation:
 [89.5555547  89.65051243 89.74530756 89.23368434 89.23987691 89.59450427
 88.84306633 89.14242981 89.80154463 89.11857057]

Final Average Accuracy of the model: 89.39
```



Conclusion

In this project, A Decision Tree Model was successfully implemented employing various prominent algorithms from the Python libraries and modules.

The R2 value of Decision Tree Model is 0.92 and Final Average Accuracy of the Model is 89.39%. So, we are getting more precise and accurate Machine Algorithm Model.

Thanks!

Do you have any questions?

abraut455@gmail.com
+91 9552183297
Ivy Professional School, Kolkata