OTT REVENUE PREDICTION MODEL (NETFLIX)

VAGDEVI P

VIBHUTI DEMBI

INTRODUCTION

- Over The Top platform

- Challenges

    - Multiple OTT platforms
    - High resolution video/audio
    - Content selection
    - Viewers retention
    - Revenue model

- OTT Revenue Models
    - SVOD (Subscription Video On Demand)
    - AVOD (Advertising Video On Demand)
    - Hybrid

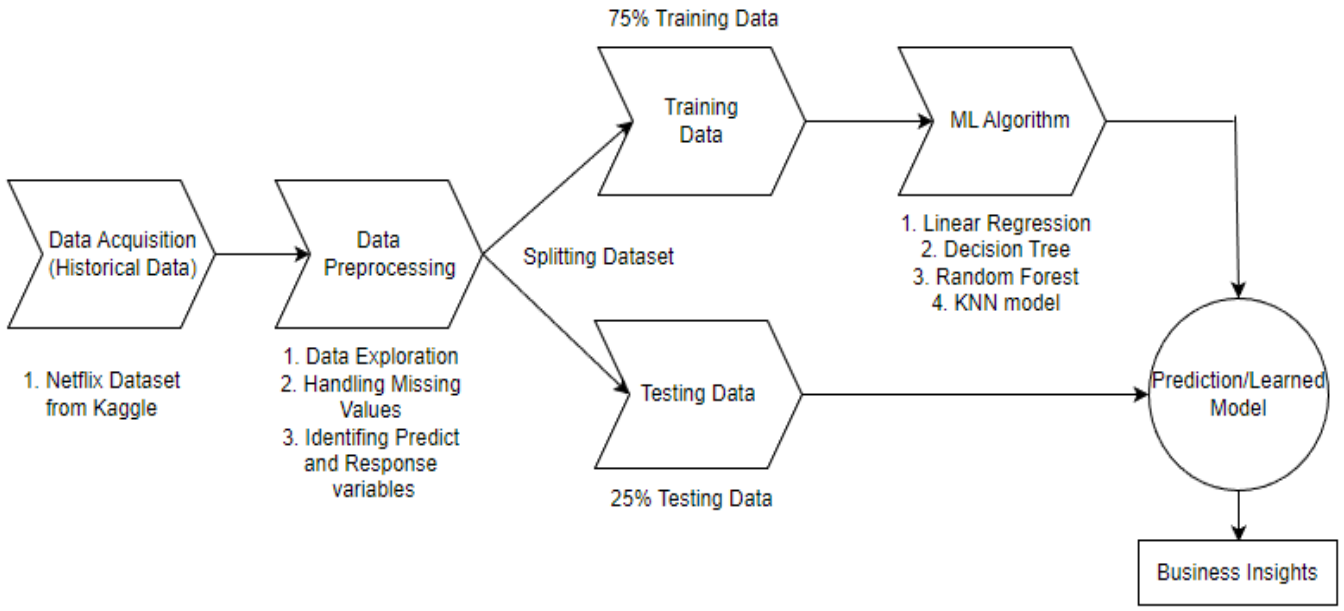
- Problem Statement

OTT prediction model

How much and how quickly a company can intend to grow?
How likely to plan for investments?
How quickly and what are the actions to be taken to improve business if there is a fall in revenue?

PREDICTION PROCESS

- Predict the revenue of Netflix OTT platform using different models: Linear Regression, Decision Tree, Random Forest and KNN models



DATA PRE-PROCESSING AND VISUALIZATION

- Understanding the data
  > Dataset Link: Kaggle-Netflix Revenue and Users.
  > Independent variable X: Subscribers/Year/Content Spend/ in million dollars/Profit in million dollars
  > Dependent variable Y: Overall revenue generated in million dollars

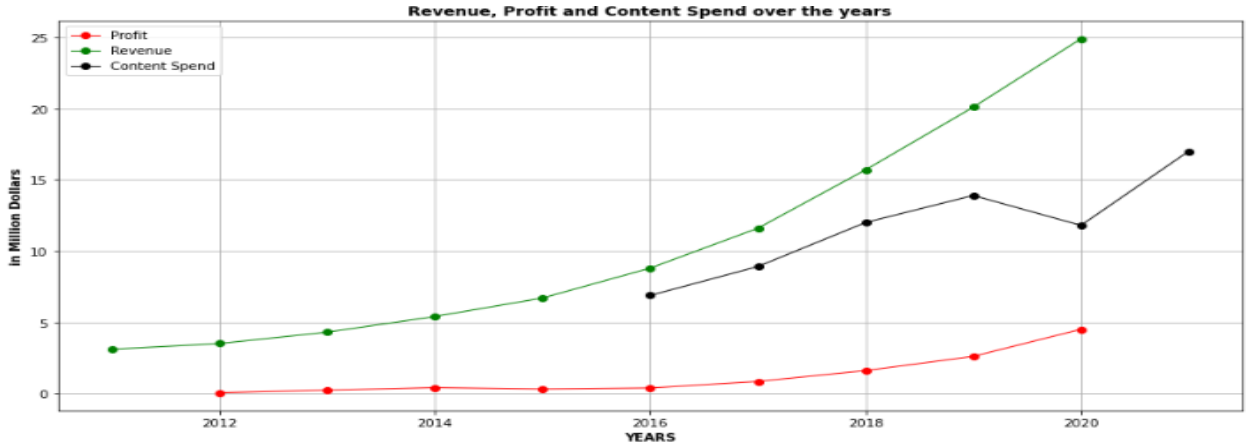
- Import required libraries
  Pandas, NumPy, Matplotlib, Seaborn, Sci-Kit

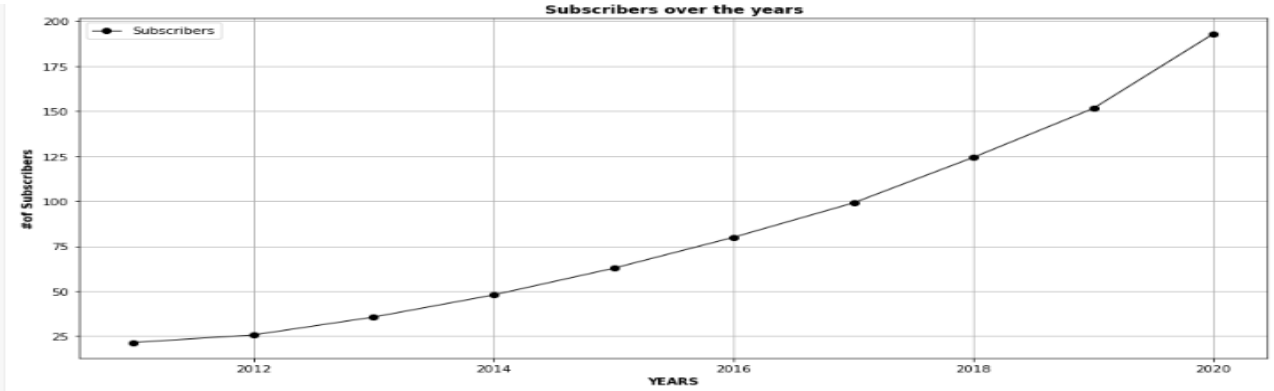
- Import dataset and explore data
  Data object creation and read_csv()

- Identifying and handling missing values
  Impute median

- Extract dependent and independent variable
  Relationship using pairplot () / lmplot () functions

- Split the dataset
  Training: Testing [75:25]





REGRESSION MODELS

1. Linear Regression Model:

- ➔ LinearRegression () is imported from sci-kit library
- ➔ Feature selection is done by checking the MSE of each feature against revenue. The one with lower MSE is chosen for prediction (Graph 1).
- ➔ Training and testing scores are computed along with R2 score, MSE(Mean Square Error) and RMSE (Root Mean Square Error) which help evaluate the linear regression model

2. Decision Tree Regression Model:

- ➔ DecisionTreeRegressor () function is imported from sci-kit library

3. Random Forest Regression Model

- ➔ RandomForestRegressor () function is imported from sci-kit library

4. K- Nearest Neighbors Regression Model

- ➔ KNeighborsRegressor () function is imported from sci-kit library and chose nearest neighbors' parameter to be 2

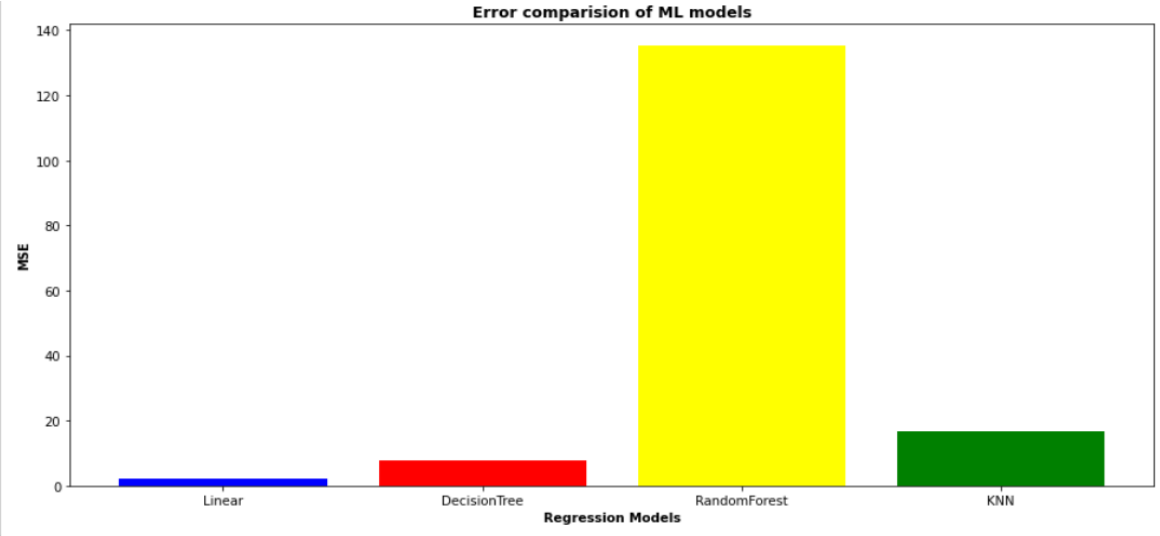
For all the above models,

- ➔ Number of subscribers is taken as an independent variable to predict Revenue (target/dependent variable)
- ➔ Model is evaluated by computing training and testing scores, R2 score, MSE and RMSE

EVALUATION

Linear_Regression_Model	MSE	Training Score	Testing Score
Year	20.17	0.35	0.59
Number of Subscribers	2.33	0.83	0.97
Profit	4.83	0.92	0.89
Content Spend	78.15	0.02	-127.81
X=(year, Number of subscribers, Profit, Content Spend)	105.93	0.95	-0.37

#of_Subscribers	Linear_Regression	Decision_Tree	Random_Forest	KNN_Regressor
MSE	2.33	7.89	135.17	16.71
Training Score	0.83	1.00	0.96	0.87
Testing Score	0.97	0.90	0.83	0.79



CONCLUSION

- From the OTT point of view, these prediction models help in making business critical decisions and provide insights to identify trends in market.
- For a given set of data used in the project, Linear Regression model suits best
- Scope of fine tuning the hyper parameters in other regression models, we might see better predictions and lower errors.
- Deep Neural Networks could help make result more accurate.

REFERENCES

[1]https://www.kaggle.com/datasets/azminetoushikwasi/ott-video-streaming-platforms-revenue-and-users

[2]https://selectra.in/blog/how-ott-earn

[3]https://scikit-learn.org/

[4]https://pandas.pydata.org/

[5]https://matplotlib.org/

[6]https://numpy.org/

[7]https://seaborn.pydata.org/

Any Question??

Thank You!!