PRIYADARSHINI ENGINEERING COLLEGE

Customer Churn Prediction

LOKESH / TEAM BOY 4
contact:lokeshdme234@gmail.com

1.Introduction

Customer churn prediction is a critical component of customer relationship management in today's competitive business landscape. It involves the use of data analytics and machine learning to forecast the likelihood of customers discontinuing their engagement with a company's products or services. By identifying potential churners early in the customer lifecycle, businesses can take proactive measures to retain valuable customers, reduce attrition rates, and optimize resource allocation.

This predictive analysis relies on historical customer data, behavior patterns, and various predictive indicators to generate insights into why customers churn and when it's likely to occur. Armed with these insights, organizations can develop targeted retention strategies, personalized incentives, and proactive customer engagement initiatives to mitigate churn and strengthen customer loyalty.

In an era where customer acquisition costs are high, retaining existing customers is not just cost-effective but also essential for sustainable growth. Customer churn prediction empowers businesses to make data-driven decisions, deliver exceptional customer experiences, and build long-lasting relationships with their customer base. In this document, we delve into the significance, methodology, and benefits of customer churn prediction for businesses aiming to thrive in the competitive marketplace.

2.Problem Statement : Customer Churn Prediction

In the dynamic landscape of modern business, retaining customers is paramount for sustainable growth and profitability. However, organizations often face the challenge of losing valuable customers, known as churn, which can significantly impact their bottom line. To address this issue effectively, it is imperative to formulate a clear problem statement:

The Problem:

Our organization faces the challenge of customer churn, where a portion of our customer base discontinues their engagement with our products or services. Churn not only results in the loss of revenue but also erodes the hard-earned customer relationships. To mitigate this problem, we need to proactively identify potential churners and implement targeted strategies to retain them.

Why It's a Problem:

- **Revenue Loss:** Churn directly affects our revenue, as acquiring new customers is often more expensive than retaining existing ones.
- **Customer Relationships:** Churn impacts the trust and loyalty we've built with our customers over time.
- **Competitive Landscape:** In a competitive marketplace, losing customers to competitors can hinder our market position.

Current Limitations:

- We lack a systematic approach to identify potential churners in advance.
- Our customer retention strategies are reactive rather than proactive.
- Limited utilization of data analytics and predictive modeling for churn prevention.

3.Stalkholder

Stakeholders for customer churn prediction initiatives include individuals and teams from various departments within an organization, as well as external parties who have an interest in or are impacted by churn prediction efforts. Here are the key stakeholders:

Data Analysts and Data Scientists:

- **Role:** Responsible for collecting, cleaning, and analyzing data to develop predictive models.
- **Involvement:** Actively participate in the data preparation and modeling stages of churn prediction.

Marketing Teams:

- **Role:** Utilize churn predictions to design targeted retention campaigns and marketing strategies.
- **Involvement:** Collaborate in the development and execution of marketing initiatives aimed at retaining at-risk customers.

Customer Service Teams:

- **Role:** Use churn predictions to identify and prioritize customers who may need special attention or proactive support.
- **Involvement:** Implement customer service strategies to address customer concerns and prevent churn.

Product Development Teams:

- **Role:** Leverage churn insights to enhance existing products or services and develop features that align with customer needs.
- **Involvement:** Incorporate customer feedback and predictive insights into product roadmaps.

Sales Teams:

- **Role:** Use churn predictions to tailor sales strategies for retaining key accounts and upselling or cross-selling to at-risk customers.
- **Involvement:** Collaborate in designing retention-focused sales approaches.

Executives and Decision-Makers:

- **Role:** Responsible for setting organizational goals and allocating resources.
- **Involvement:** Receive insights from churn predictions to inform high-level strategic decisions.

IT and Technology Teams:

- **Role:** Support the implementation and integration of churn prediction tools and systems.
- **Involvement:** Collaborate in deploying predictive models and ensuring data security.

Finance and Accounting Teams:

- **Role:** Assess the financial impact of churn and the ROI of retention efforts.
- **Involvement:** Analyze financial data related to churn and retention initiatives.

Customers:

- **Role:** The customers themselves can be stakeholders, as they are directly impacted by the churn prediction and retention strategies.
- **Involvement:** Provide feedback and responses that feed into predictive models and influence retention efforts.

Regulatory and Compliance Bodies (if applicable):

- **Role:** Ensure that data privacy and compliance standards are met when handling customer data.
- **Involvement:** Oversee and enforce data privacy and security measures.

Investors and Shareholders (external stakeholders):

4.Impact Analysis

Customer churn prediction has significant implications for an organization across various aspects of its operations and performance. Conducting an impact analysis helps organizations understand how churn prediction efforts can influence different areas. Here's an analysis of the potential impacts:

Revenue and Financial Impact:

- **Positive Impact:** Churn prediction can lead to increased revenue by identifying at-risk customers and implementing targeted retention strategies. Retaining customers is often more cost-effective than acquiring new ones.
- **Negative Impact:** Failure to effectively predict and address churn can result in revenue loss due to customer defection.

Customer Relationships:

- **Positive Impact:** Proactive churn prevention can strengthen customer relationships by demonstrating a commitment to customer satisfaction. Satisfied customers are more likely to become loyal advocates.
- **Negative Impact:** High churn rates can erode trust and loyalty, damaging customer relationships and the organization's reputation.

Marketing and Sales:

- **Positive Impact:** Churn predictions enable marketing and sales teams to design targeted campaigns and strategies to retain at-risk customers, improving customer retention rates.
- **Negative Impact:** Without accurate predictions, marketing efforts may not effectively address churn, leading to wasted resources.

Operational Efficiency:

- **Positive Impact:** Churn prediction allows for more efficient resource allocation, as resources can be focused on retaining high-value customers rather than chasing new leads.
- **Negative Impact:** Inefficient resource allocation due to inaccurate predictions can lead to suboptimal operational costs.

Product Development:

- **Positive Impact:** Churn insights inform product development by identifying areas for improvement and feature enhancements, leading to higher customer satisfaction.
- **Negative Impact:** Neglecting customer feedback and churn data may result in products or services that do not meet customer expectations.

5.Goals

Setting clear and achievable goals for customer churn prediction is essential to ensure that the efforts are focused, measurable, and aligned with the organization's objectives. Here are key goals for implementing a customer churn prediction system:

Reduce Churn Rate: The primary goal is to decrease the churn rate by identifying at-risk customers early and implementing effective retention strategies. The specific target for churn reduction should be defined.

Increase Customer Retention: Increase the percentage of customers who continue using the products or services over a specific period, thereby enhancing customer lifetime value.

Enhance Revenue: Boost revenue through the prevention of revenue loss due to churn and the potential upselling or cross-selling of products or services to retained customers.

Improve Customer Satisfaction: Use churn prediction to address customer issues and concerns promptly, leading to improved satisfaction levels and reduced complaints.

Optimize Marketing Spend: Make marketing efforts more efficient by targeting retention campaigns toward customers most likely to churn, reducing marketing costs while maintaining or increasing effectiveness.

Enhance Product/Service Quality: Utilize churn insights to identify areas for improvement in products or services, ensuring they align with customer expectations.

Increase Cross-Selling and Upselling: Identify opportunities to cross-sell or upsell additional products or services to existing customers, thereby increasing average revenue per customer

6.Design Thinking Process

Design thinking is a human-centered approach that encourages creative problem-solving and innovation. Applying design thinking principles to customer churn prediction involves empathizing with customers, defining specific needs and challenges, ideating creative solutions, prototyping predictive models, testing their effectiveness, and iterating to continuously improve. Here's how the design thinking process can be applied:

Empathize (Understand Customer Needs):

- **User Research:** Conduct in-depth user research to understand customer behaviors, pain points, and reasons for churn.
- **Surveys and Interviews:** Gather qualitative data through surveys and interviews to empathize with customers' experiences.
- **Data Analysis:** Analyze historical customer data to identify patterns and trends related to churn.

Define (Problem Statement and User Needs):

- **Problem Statement:** Clearly define the problem of customer churn, incorporating insights gained during the empathize phase.
- **User Needs:** Identify specific user needs and challenges related to churn prediction and prevention.

Ideate (Generate Creative Solutions):

- **Brainstorming:** Encourage cross-functional teams to brainstorm creative ideas for predicting and addressing customer churn.
- **Concept Development:** Develop innovative concepts and strategies, considering both technological and non-technological solutions.

Prototype (Create Predictive Models):

- **Model Development:** Collaborate with data scientists and analysts to develop predictive models using machine learning algorithms.
- **Feature Engineering:** Identify relevant features (data variables) that contribute to churn prediction.
- **Visualization:** Create visual prototypes and dashboards to present churn predictions in an understandable format.

7.Conclusion

In a dynamic business landscape where customer loyalty and retention are paramount, the implementation of customer churn prediction stands as a strategic imperative. This process, driven by data, empathy, and innovation, has far-reaching implications for organizations seeking to thrive in competitive markets. As we conclude our exploration of customer churn prediction, it becomes evident that it is not merely a tool but a transformative approach to preserving valuable customer relationships.

Customer churn prediction, rooted in the principles of design thinking, enables organizations to anticipate and address customer attrition with precision and care. Through empathetic understanding, data-driven insights, and iterative refinements, it empowers businesses to achieve the following:

7.1 **Enhanced Customer Retention:** By identifying potential churners in advance, organizations can implement tailored strategies to retain customers, nurturing long-term relationships.

7.2 **Increased Revenue and Profitability:** Churn prediction reduces revenue loss and opens doors to upselling and cross-selling opportunities, bolstering financial sustainability.

7.3 **Improved Customer Satisfaction:** Proactive responses to customer concerns foster satisfaction, trust, and loyalty, solidifying the foundation for enduring relationships.

7.4 **Resource Optimization:** Resources are allocated efficiently, channeling efforts toward retaining high-value customers and reducing acquisition costs.

7.5 **Data-Driven Decision-Making:** A culture of data-driven decision-making is cultivated, enabling organizations to stay agile and responsive to changing customer behaviors.

7.6 **Competitive Advantage:** Organizations gain an edge by staying ahead of competitors in customer relationship management, bolstering market position.

As we navigate an era where customer expectations continue to evolve, the journey toward effective churn prediction does not conclude but evolves. It demands continuous adaptation, innovation, and a commitment to nurturing customer-centricity. By keeping customer needs and experiences at the heart of our strategies, we ensure that customer churn prediction remains an ever-reliable compass guiding us toward sustainable growth and enduring success.

PHASE 2 : INNOVATION

Customer Churn Prediction

Priyadarshini Engineering College

Introduction:

Customer Churn Prediction is the process of collecting, examining, and interpreting data related to the visitors and interactions on a website. It provides invaluable insights into user behavior, preferences, and trends.

Problem Statement:

The problem of "Customer Churn Prediciton" lies in the need for organizations to effectively understand and leverage user behavior on their websites

OBJECTIVES

1.Data Acquisition:

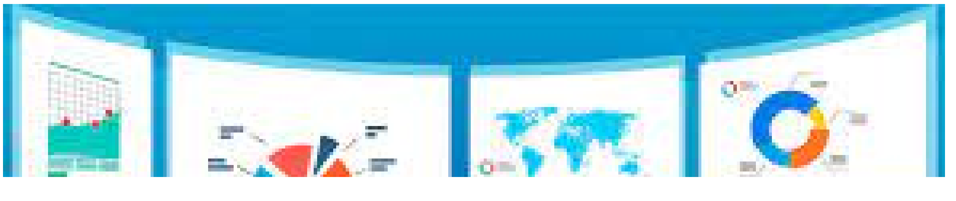
Download and import the dataset from kaggle into your analytics environment.

2.Exploratory Data Analysis (EDA):

Conduct EDA to understand the dataset's characteristics, patterns, and correlations. Visualize key metrics and trends in website traffic.

3.Analysing the Data:

Regularly analyze data collected from the real-time analytics tools and experiments. Adjust strategies based on insights derived from the dataset.



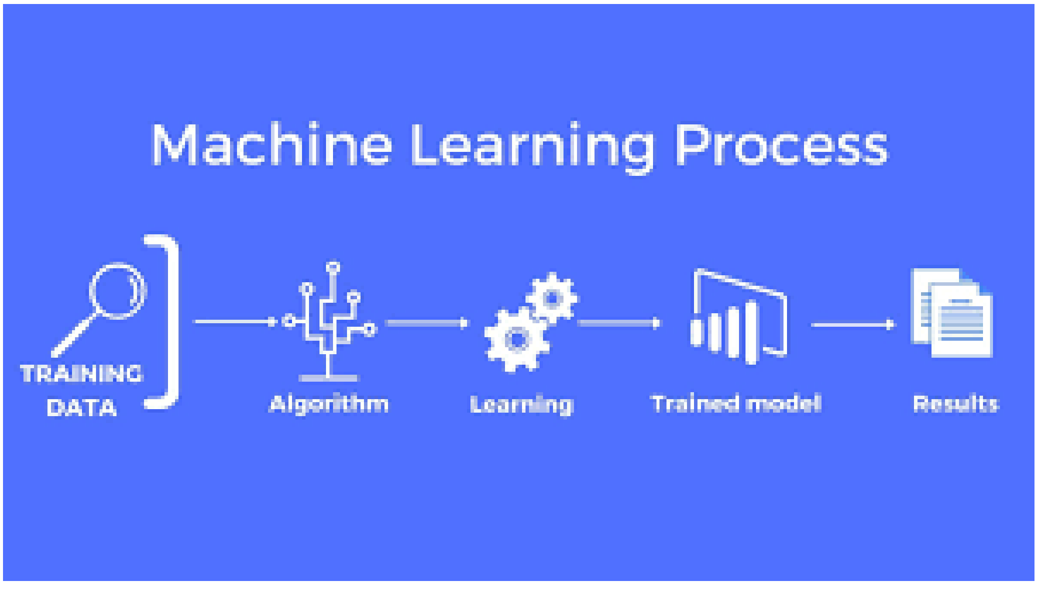
MACHINE LEARNING MODEL

- Regression Analysis: Linear regression or more advanced methods like polynomial regression can be used to model and predict traffic trends over time.

- Time Series Analysis: Techniques like ARIMA (AutoRegressive Integrated Moving Average) can help in forecasting website traffic based on historical data.

- Classification Algorithms: These can be used to categorize website visitors, such as decision trees, random forests, or support vector machines, to identify different user segments or traffic sources.

- Recommendation Systems : Collaborative filtering or content-based recommendation algorithms can be be used to suggest content to users based on their behavior.



VISUAL METHOD

1. Bar Charts: Bar charts are useful for displaying metrics like the number of page views, unique visitors, or bounce rates over a specific time period. You can create bar charts to compare different time intervals or website sections.

2. Line Charts: Line charts are effective for showing trends in website traffic data, such as changes in visitor numbers over time. They can help identify seasonal patterns and long-term trends.

3. Pie Charts: Pie charts can be used to represent the distribution of traffic sources, showing the percentage of traffic coming from direct visits, search engines, social media, etc.

4. Area Charts: Area charts are similar to line charts but can be used to display the cumulative effect of website traffic data over time. They are good for visualizing total page views or unique visitors.

Dashboards:

Dashboards in IBM Cognos allow you to combine various visualizations and key metrics on a single screen, offering a comprehensive overview of website traffic.

Phase 3

Customer Churn Analysis Prediction Priyadrashini engineering college

Introduction

Financial institutions have many clients close their accounts or migrate to other institutions. As a result, this has made a significant hole in sales and may significantly impact yearly revenues for the current fiscal year, leading stocks to plummet and market value to fall by a decent percentage. A team of business, product, engineering, and data science professionals has been assembled to halt this decline.

The objective of this tutorial is that we want to build a model to predict, with reasonable accuracy, the customers who are going to churn soon.

A customer having closed all their active accounts with the bank is said to have churned. Churn can be defined in other ways as well, based on the context of the problem. A customer not transacting for six months or one year can also be defined as churned based on the business requirements.

The Dataset

In the customer churn modeling dataset, we have 10000 rows (each representing a unique customer) with 15 columns: 14 features with one target feature (`Exited`). The data is composed of both numerical and categorical features:

`Exited` — Whether the customer churned or not.

- `CustomerId`: A unique ID of the customer.
- `CreditScore`: The credit score of the customer,
- `Age`: The age of the customer,
- `Tenure`: The number of months the client has been with the firm.
- `Balance`: Balance remaining in the customer account,
- `NumOfProducts`: The number of products sold by the customer.
- `EstimatedSalary`: The estimated salary of the customer.

Categorical Features:

- `Surname`: The surname of the customer.
- `Geography`: The country of the customer.
- `Gender`: M/F
- `HasCrCard`: Whether the customer has a credit card or not.
- `IsActiveMember`: Whether the customer is active or not.

If you don't have a Kaggle account, the dataset can be downloaded here or here.

Questioning the Data

Here, the objective is to understand the data further and distill the problem statement and the stated goal. In the process, if more data/context can be obtained, that adds to the result of the model performance:

- No date/time column. A lot of useful features can be built using date/time columns.
- When was the data snapshot taken? There are certain customer features like `Balance`, `Tenure`, `NumOfProducts`, `EstimatedSalary`, which will have different values across time.
- Are all these features about the same single date or spread across multiple dates?
- How frequently are customer features updated?
- Will it be possible to have the values of these features over some time as opposed to a single snapshot date?
- Some customers who have exited still have a balance in their account or a non-zero `NumOfProducts`. Does this mean they had churned only from a specific product and not the entire bank, or are these snapshots just before they churned?
- Some features like number and kind of transactions can help us estimate the degree of activity of the customer, instead of trusting the binary variable `IsActiveMember`.
- Customer transaction patterns can also help us ascertain whether the customer has churned or not. For example, a customer might transact daily/weekly vs. a customer who transacts annually.

We don't have answers to all of these questions, and we will only use what we have.

Once the dataset is downloaded, put it in the current working directory.

Let's install the dependencies of this tutorial:

```
$ pip install ipython==7.22.0 joblib==1.0.1 lightgbm==3.3.1 matplotlib numpy pandas scikit_learn==0.24.1 seaborn xgboost==1.5.1
```

Copy

Let's import the libraries:

Phase 3

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder, OneHotEncoder,
StandardScaler
from sklearn.feature_selection import RFE
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC
from sklearn.decomposition import PCA
from sklearn.tree import DecisionTreeClassifier, export_graphviz
from sklearn.pipeline import Pipeline
from sklearn.model_selection import train_test_split
from lightgbm import LGBMClassifier
from sklearn.metrics import roc_auc_score, recall_score,
confusion_matrix, classification_report
import subprocess
import joblib
# Get multiple outputs in the same cell
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"
# Ignore all warnings
import warnings
warnings.filterwarnings('ignore')
warnings.filterwarnings(action='ignore', category=DeprecationWarning)
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
```

Copy

Reading the dataset:

```
# Reading the dataset
dc = pd.read_csv("Churn_Modelling.csv")
dc.head(5)
```

Copy

Phase 3

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
0	1	15634602	Hargrave	619	France	Female	42	2	0.00	1	1	1	101348.88	1
1	2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
2	3	15619304	Onio	502	France	Female	42	8	159660.80	3	1	0	113931.57	1
3	4	15701354	Boni	699	France	Female	39	1	0.00	2	0	0	93826.63	0
4	5	15737888	Mitchell	850	Spain	Female	43	2	125510.82	1	1	1	79084.10	0

Copy

Related: *Recommender Systems using Association Rules Mining in Python*.

Exploratory Data Analysis

Let's see the dimension of the dataset:

```
# Dimension of the dataset
dc.shape
```

Copy

```
(10000, 14)
```

Copy

Let's see some statistics of the data. The first line describes all numerical columns, where the second describes categorical columns:

```
dc.describe(exclude= ['O']) # Describe all numerical columns
dc.describe(include = ['O']) # Describe all categorical columns
```

Copy

Phase 3

	RowNumber	CustomerId	CreditScore	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
count	10000.00000	1.000000e+04	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.00000	10000.000000	10000.000000	10000.000000
mean	5000.50000	1.569094e+07	650.528800	38.921800	5.012800	76485.889288	1.530200	0.70550	0.515100	100090.239881	0.203700
std	2886.89568	7.193619e+04	96.653299	10.487806	2.892174	62397.405202	0.581654	0.45584	0.499797	57510.492818	0.402769
min	1.00000	1.556570e+07	350.000000	18.000000	0.000000	0.000000	1.000000	0.00000	0.000000	11.580000	0.000000
25%	2500.75000	1.562853e+07	584.000000	32.000000	3.000000	0.000000	1.000000	0.00000	0.000000	51002.110000	0.000000
50%	5000.50000	1.569074e+07	652.000000	37.000000	5.000000	97198.540000	1.000000	1.00000	1.000000	100193.915000	0.000000
75%	7500.25000	1.575323e+07	718.000000	44.000000	7.000000	127644.240000	2.000000	1.00000	1.000000	149388.247500	0.000000
max	10000.00000	1.581569e+07	850.000000	92.000000	10.000000	250898.090000	4.000000	1.00000	1.000000	199992.480000	1.000000

	Surname	Geography	Gender
count	10000	10000	10000
unique	2932	3	2

Phase 3

```
╟────────┼─────────┼───────────┼─────────╢
║ top    │ Smith   │ France    │ Male    ║
╟────────┼─────────┼───────────┼─────────╢
║ freq   │ 32      │ 5014      │ 5457    ║
╚════════╧═════════╧═══════════╧═════════╝
```

Copy

Checking the number of unique customers:

```
# Checking number of unique customers in the dataset
dc.shape[0], dc.CustomerId.nunique()
```

Copy

```
(10000, 10000)
```

Copy

This means each row corresponds to a customer.

Let's see the churn distribution:

```
# churn value Distribution
dc["Exited"].value_counts()
```

Copy

```
0    7963
1    2037
Name: Exited, dtype: int64
```

Copy

The data set is imbalanced from the above result, with a significant chunk of existing customers relative to their churned peers.

Below, we group by `Surname` to see the average churn value:

```
dc.groupby(['Surname']).agg({'RowNumber':'count', 'Exited':'mean'}

).reset_index().sort_values(by='RowNumber', ascending=False).head()
```

Copy

```
╔══════╤═════════╤═══════════╤═════════╗
```

	Surname	RowNumber	Exited
2473	Smith	32	0.281250
1689	Martin	29	0.310345
2389	Scott	29	0.103448
2751	Walker	28	0.142857
336	Brown	26	0.192308

Copy

Or grouping by `Geography`:

```
dc.groupby(['Geography']).agg({'RowNumber':'count', 'Exited':'mean'}
).reset_index().sort_values(by='RowNumber', ascending=False)
```

Copy

	Geography	RowNumber	Exited
0	France	5014	0.161548
1	Germany	2509	0.324432
2	Spain	2477	0.166734

Copy

From what we see above, customers from Germany have a higher exiting rate than average.

Univariate Plots of Numerical Variables

- Plotting `CreditScore` as a boxplot:

Data Preprocessing

So, in summary, here's what we're going to do:

Phase 3

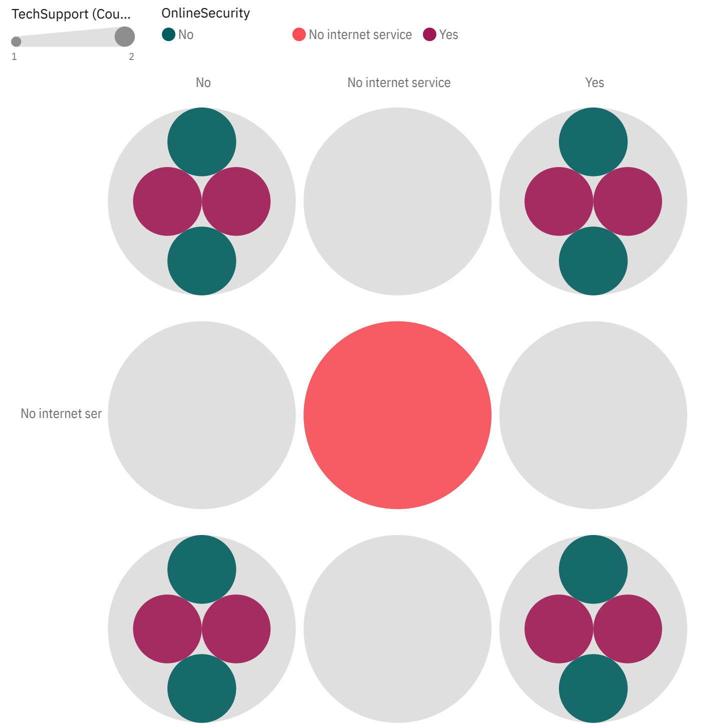
- We will discard the `RowNumber` column.
- We will discard `CustomerID` as well since it doesn't convey any extra info. Each row pertains to a unique customer.
- Features can be segregated into non-essential, numerical, categorical, and target variables based on the above.

In general, `CustomerID` is a handy feature based on which we can calculate many user-centric features. Here, the dataset is not sufficient to calculate any extra customer features.

Tab 1

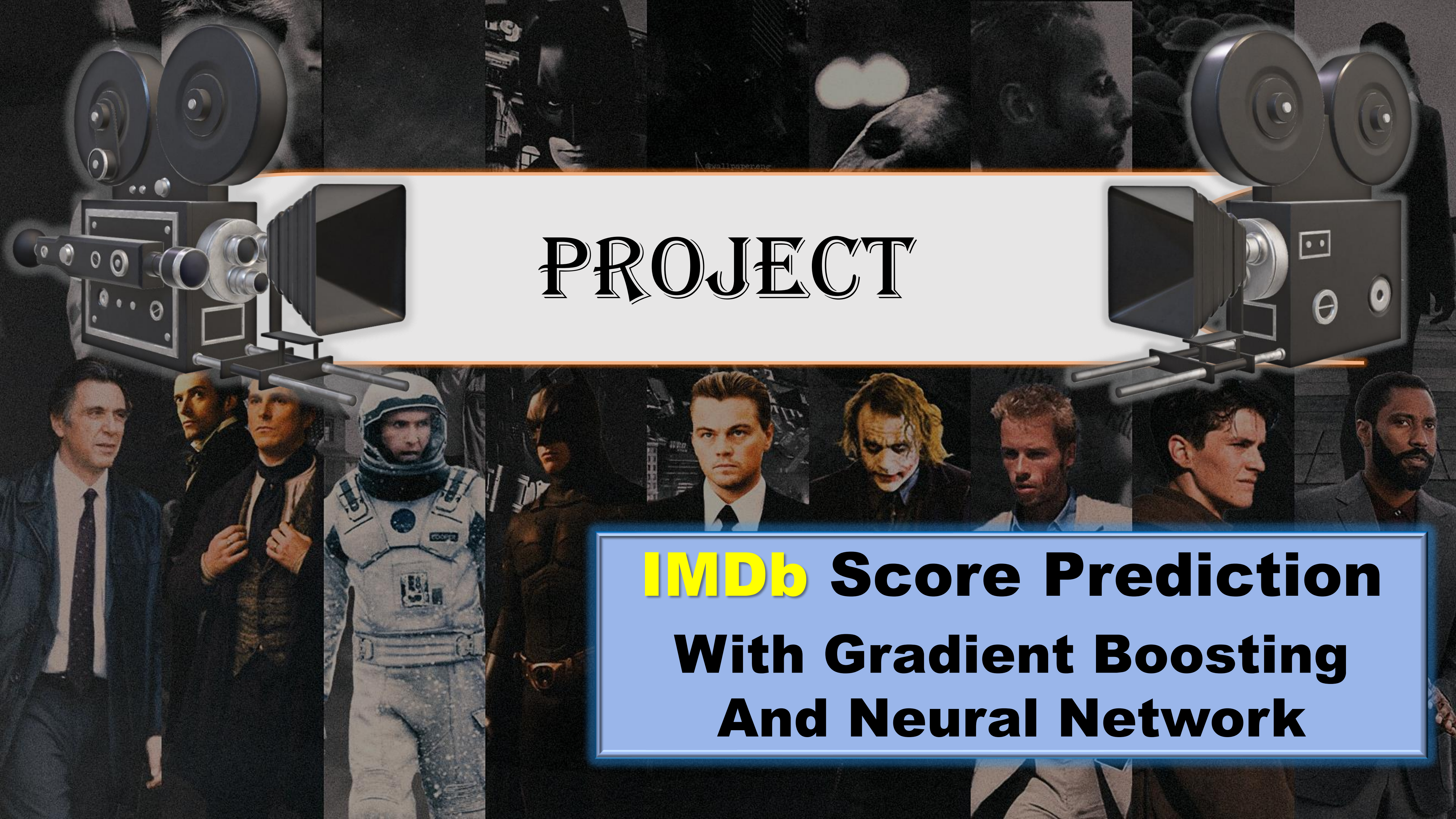
MultipleLines by customerID colored by gender

InternetService hierarchy colored by OnlineSecurity and sized by TechSupport



PaymentMethod, Churn and PaperlessBilling

PaymentMethod	Churn	PaperlessBilling
Bank transfer (automatic)	No	No
		Yes
	Yes	No
		Yes
Credit card (automatic)	No	No
		Yes
	Yes	No
		Yes
		No



PROJECT

IMDb Score Prediction

With Gradient Boosting And Neural Network

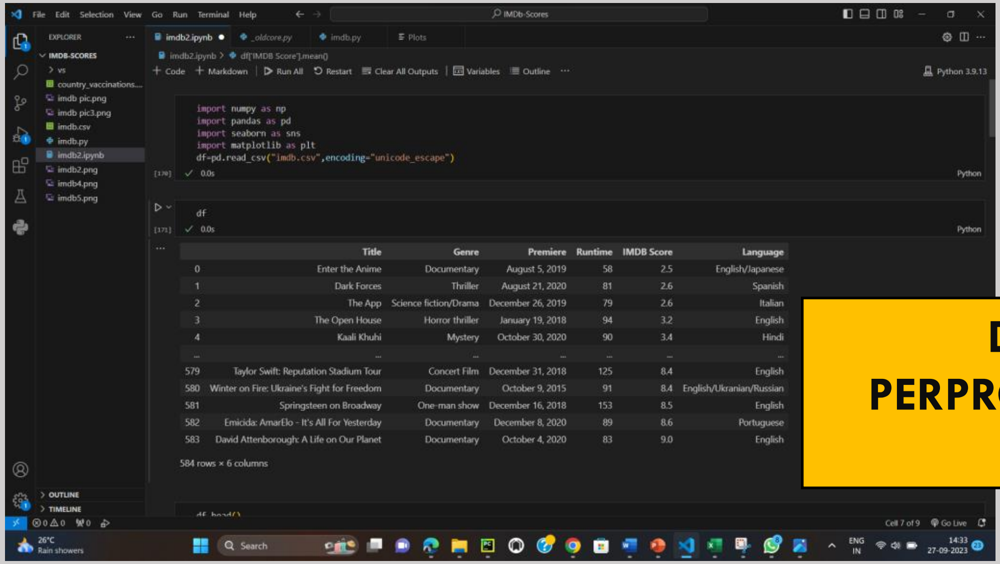
Develop a machine learning model to predict the IMDb Scores of the movies available on Films Based in their genre, premiere date, runtime and language . The model aims to accurately finds the popularity of the movies to the assist users in discovering highly rated films that align with their preferences

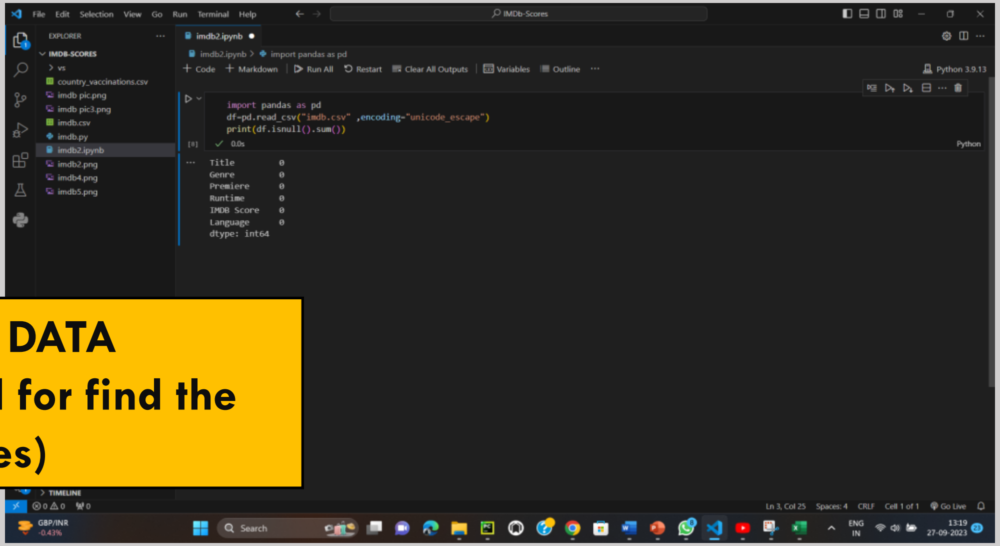
Design the project based on :

* Data Source
* Data Preprocessing
* Feature Engineering
* Model Selection
* Model Training
* Evaluation

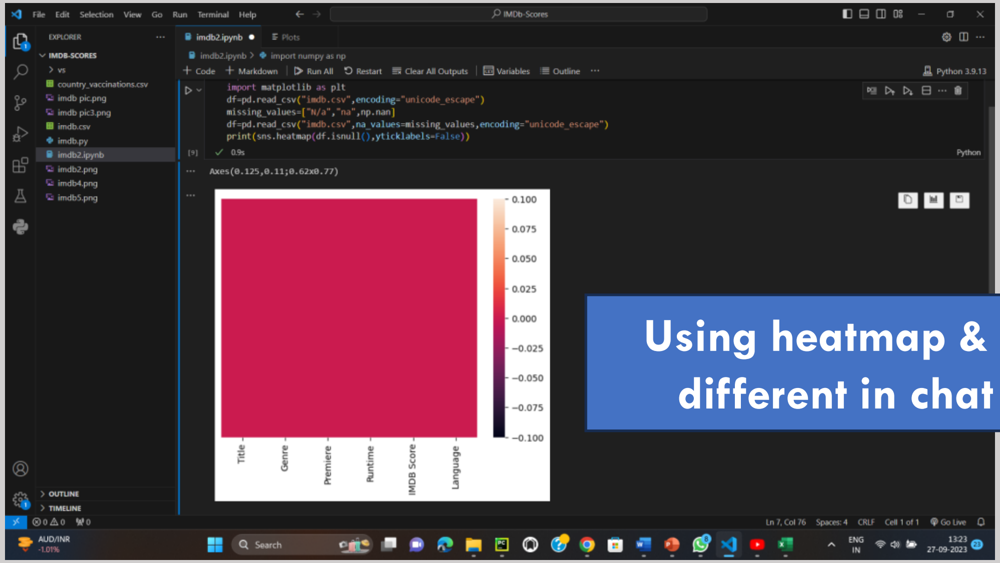
Main algorithms & ML are :

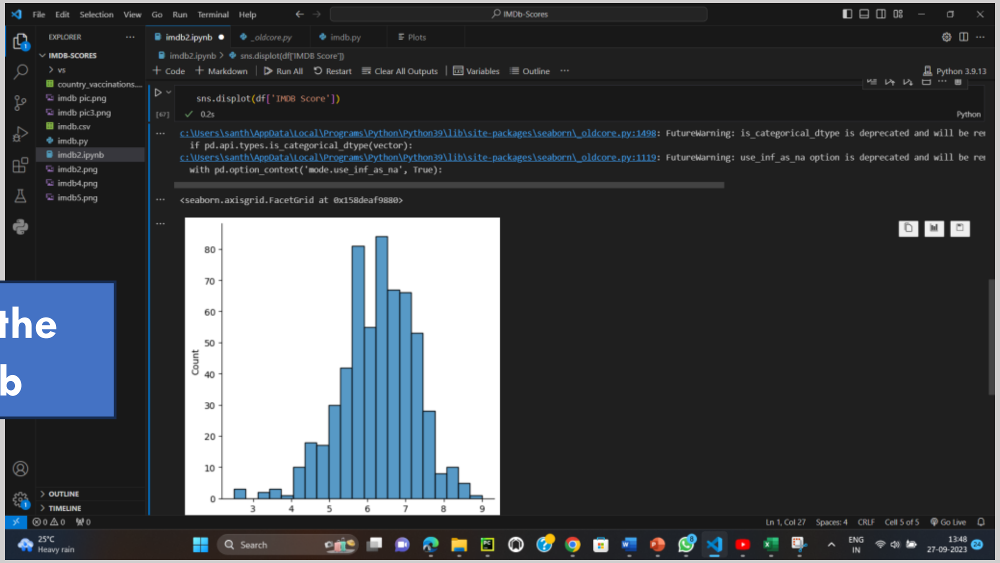
linear Regression ,Random Forest Regression to Predict IMDb Scores .Train the selected model using preprocessing data Regression metrics like MAE ,MSE,R-squared,Gradient Boosting and Neural Network.





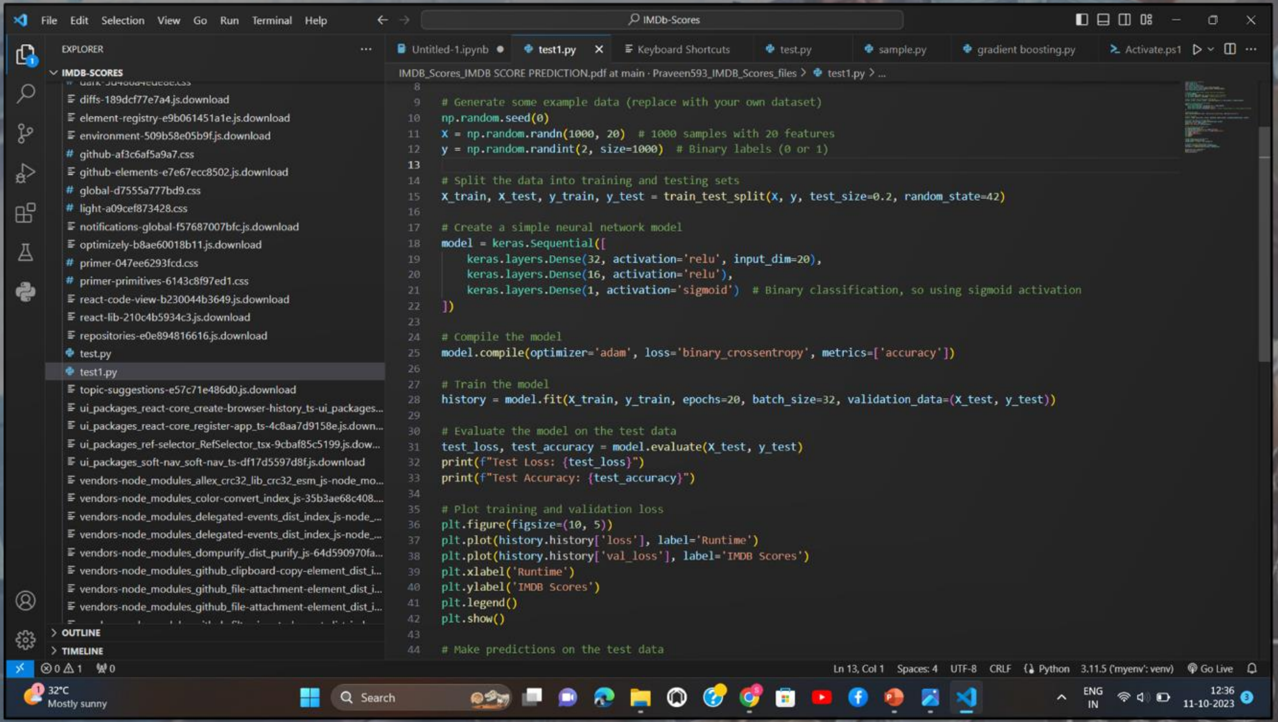
DATA SOURCE & DATA PERPROCESSING (used for find the missing values)



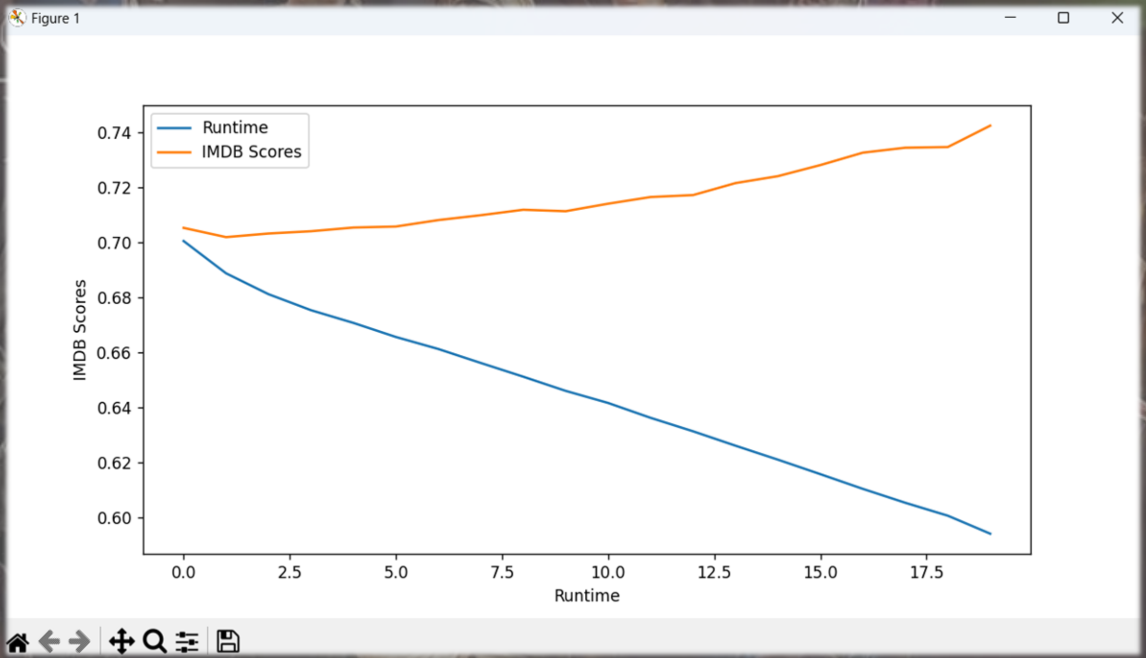


Using heatmap & displot can find the different in chat and predict IMDb

NEURAL NETWORKING



- ➢ Neural Networking is an ensemble learning technique that combines the predictions of several weak learners to create a strong predictive model.
- ➢ Neural Networking,with its iterative learning approach, stands as a powerful tool for predictive modeling, providing accurate results across diverse dataset



- ➢ **Hyperparameter Tuning:** Proper tuning of hyperparameters is crucial for optimal performance.
- ➢ **Potential for Overfitting:** Care must be taken to avoid overfitting, especially if the weak learners are too complex.

Key Components For Gradient Boosting

- **Weak Learners**: Typically shallow decision trees are used.
- **Boosting**: Models are built sequentially, and each subsequent model corrects errors made by the previous ones.

Use Cases For Gradient Boosting

- **Commonly Applied**: Used in various domains, including finance, healthcare, and Kaggle competitions.

Key Components For Neural Network

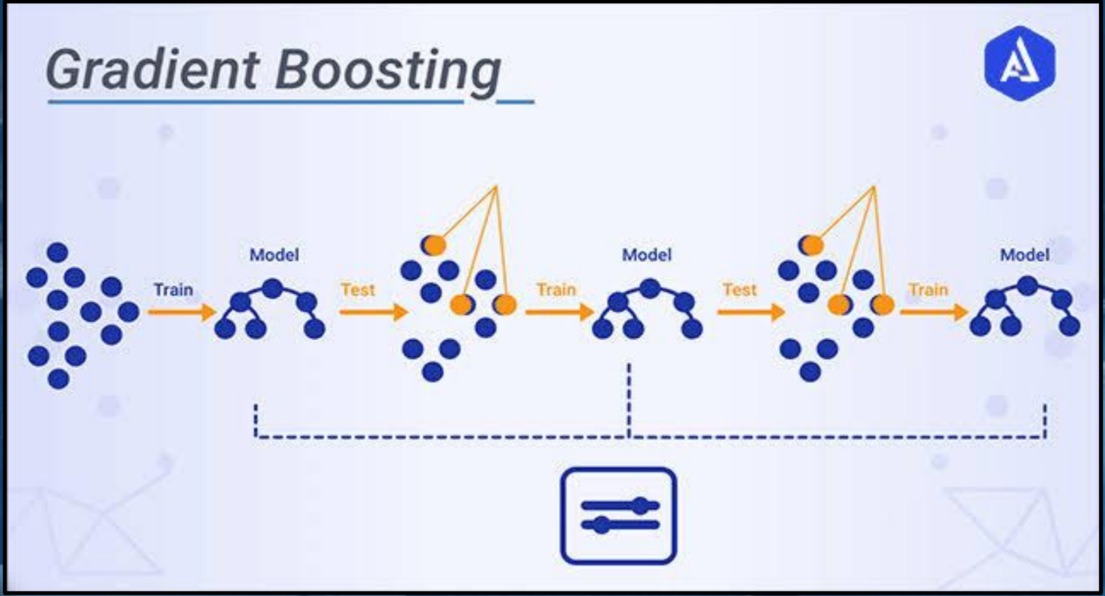
- **Neurons**: Basic computational units that process information.
- **Layers**: Organized in input, hidden, and output layers.
- **Activation Functions**: Determine the strength of connections between neurons.

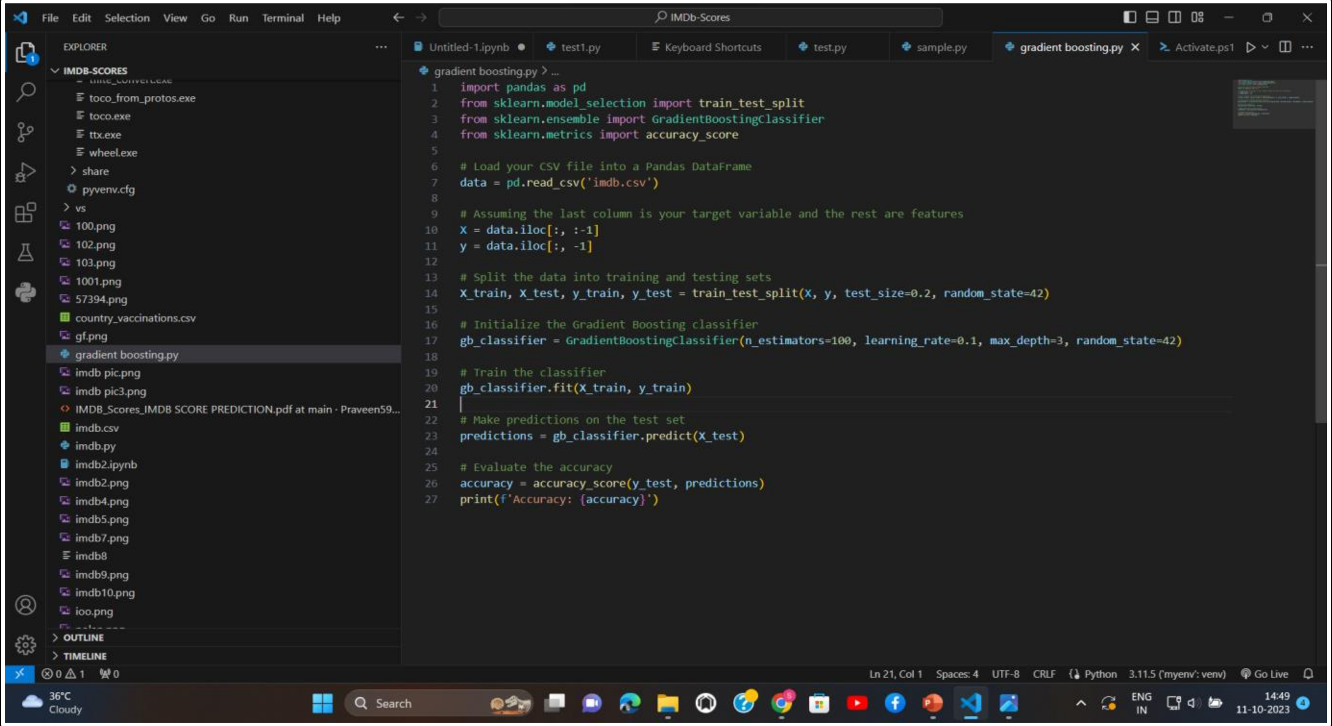
Use Cases For Gradient Boosting

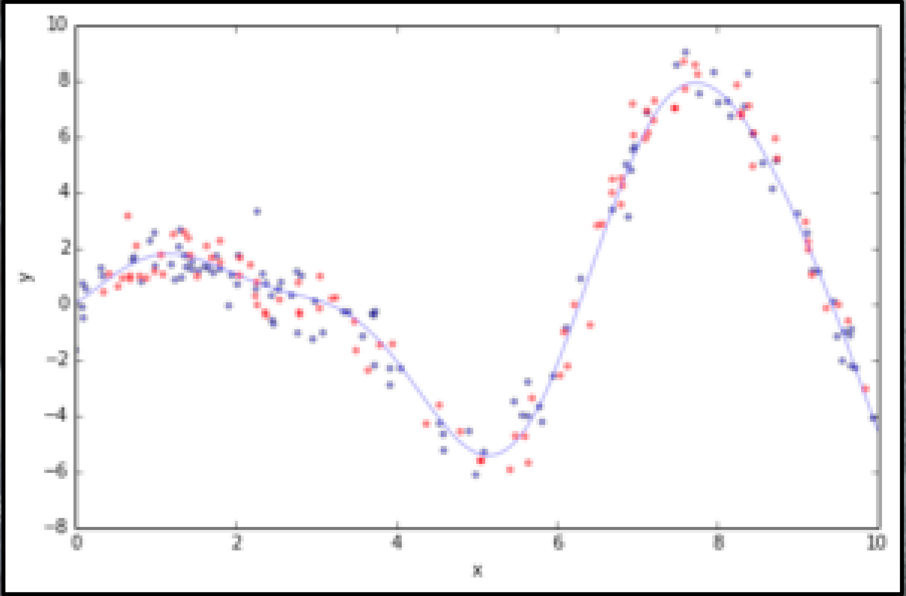
- **Image and Speech Recognition**: Neural Networks excel in tasks like image and speech recognition.

GRADIENT BOOSTING

- Gradient Boosting is an ensemble learning technique that combines the predictions of several weak learners to create a strong predictive model.
- Gradient Boosting, with its iterative learning approach, stands as a powerful tool for predictive modeling, providing accurate results across diverse datasets.







THANK YOU