CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

We often hear about a new problem customer attrition that predicts the churn of customers and makes it easy to analyze the reasons for a customer to leave the bank and helps to take measures to stop customers from churn. Nowadays there are a lot of service providers available in every business. There is no shortage of customers in any options.Mainly, in the banking sector when people want to keep their money safely they have a lot of options. As a result, customer churn and loyalty of customers have become a major problem for most banks.In this project ,it represents a method that predicts customer churn in banking using Machine learning with XGBoost algorithm. This research promotes the exploration of the likelihood of churning by customer loyalty.

The objective of the project is to prevent customers to churn from banks.Churn refers to the movement of customers from one platform to another. Churn rate lies in the range from 10% to 30%. Once customers are identified at risk of churn, banks should work on the marketing efforts that make each customer maximize their likelihood of staying. Losing a customer usually leads to loss in profit for the bank. Long term customers become less costly to serve and they generate huge profits. They even provide new referrals. The cost of attracting new customers can be five to six times more than holding on to existing customers.

1.2 PURPOSE OF THE PROJECT

The main purpose of the project is to identify and prevent customers from churning. These days, there is no shortage of options for customers in the banking sector when choosing where to put their money. As a result, customer churn and engagement have become one of the top issues for most banks.

1.3 MOTIVATION

The motivation behind employing the XGBoost algorithm for bank customer churn prediction lies in the critical need for financial institutions to proactively retain customers and mitigate revenue loss. By leveraging the strengths of XGBoost, which include high predictive accuracy and the ability to handle complex relationships within data, banks aim to identify customers at risk of churning. XGBoost's advanced machine learning techniques are particularly well-suited for this task, allowing the algorithm to navigate intricate patterns in customer behavior. Additionally, XGBoost's proficiency in handling imbalanced datasets addresses the challenge posed by the often small proportion of churn cases. The algorithm's feature importance analysis not only aids in understanding the key drivers of churn but also provides interpretability, facilitating strategic decision-making. With XGBoost's real-time prediction capabilities, financial institutions can promptly intervene and implement targeted retention strategies to enhance customer loyalty and overall business performance.

The XGBoost algorithm in bank customer churn prediction lies in its sophisticated and efficient approach. XGBoost's ensemble learning captures nuanced data relationships, ensuring robust predictive performance, particularly in handling imbalanced datasets common in customer churn scenarios. The algorithm's feature importance analysis allows banks to tailor retention strategies by understanding key churn factors. The interpretability of XGBoost fosters trust and facilitates communication between data scientists and business stakeholders. The algorithm's real-time prediction capability aligns with the dynamic nature of the banking sector, enabling prompt identification of potential churners and implementation of personalized retention measures. Ultimately, XGBoost's accuracy, interpretability, and timeliness empower financial institutions to proactively manage customer relationships and optimize long-term profitability.The use of XGBoost in bank customer churn prediction is driven by its ability to efficiently handle complex relationships, address imbalanced datasets, and provide interpretable insights, facilitating timely intervention and personalized retention strategies for enhanced customer loyalty and sustained profitability.

CHAPTER 2

LITERATURE SURVEY

There already exist a number of churn prediction and analyzing models. A few Examples prediction models available in the market are discussed in this section along with the tasks they can provide and their drawbacks.

2.1 EXISTING SYSTEM

Although data mining techniques are used across the world to predict and analyse the churn of customers it sometimes does not provide accurate predictions,There are no efficient models that predict the churn of customers. There is a lack of a proper predictive model for customer churn. While significant progress has already been made for this problem, very limited work has been done.

Work done so far in this field has been much more focused on the customer interaction management, churn prediction modelling, customer feedback and experience management, as well as leveraging AI and predictive analytics for churn risk assessment and management. The underlying architecture for most of the systems are based on:

1. Customer Interaction Management
2. Customer Interaction Management
3. Assessing Churn Risk

2.2 LIMITATIONS OF EXISTING SYSTEM

- Such interactive systems are already developed for churn prediction like Akio, Churnly, XMFayrix, Qualtrics but most of them can’t be used to make predictions for larger datasets. They can only generate predictions for smaller datasets.
- Also, these interactive systems do not enhance the accuracy, efficiency,and scalability of churn predictions.
- Predictive models such as Akio,Chur;ly don't even have these features developed for them, even though many people of the country use them.
- **Akio** :

  Purpose: It serves as a churn prediction system, focusing on forecasting customer attrition without an explicitly defined purpose.

  Differences: Akio implementations differ based on unique features, data sources, and modeling approaches, showcasing the system's diversity within churn prediction frameworks.

  Considerations: The effectiveness of Akio relies on its design and the quality of training data, emphasizing the importance of adaptability to different scenarios for optimal performance.
- **Churnly** :

  Purpose: This is specializing in forecasting customer churn with a focus on identifying and mitigating customer attrition.

  Differences: It’s used for data processing methods, and potential unique features, setting it apart in its specific approach to churn prediction.

  Considerations: In this accuracy of churn predictions rely on the strength of Churnly's algorithms.

CHAPTER 3

PROPOSED SYSTEM

3.1 PROPOSED SYSTEM

Few works have been done to generate a system that is based on the above concepts listed in the existing approaches section and cater to the banking system for churn prediction. Thus we propose to develop one for the Banking System based on the XGBoost algorithm. The success of this translation system will depend on the accuracy of data which is gathered from banks overs past years.

3.2 OBJECTIVES OF PROPOSED SYSTEM

Our project aims to do customer churn prediction.This analysis is important for identifying old customers without loss and developing new products and making new strategic decisions for retaining customers. It encompasses the Banking System. For each of the data gathered from banks over past years,it's considered and prediction is done based upon the data. Identify customers at risk of leaving the bank early.Develop measures to retain at-risk customers through targeted incentives.Understand and address factors contributing to dissatisfaction. Target marketing efforts efficiently based on predicted churn likelihood. Tailor offerings to individual customer preferences to enhance loyalty. Use advanced analytics for accurate churn predictions and model refinement. Minimise financial losses associated with customer churn.

3.3 SYSTEM REQUIREMENTS

Here are the requirements for developing and deploying the application.

3.3.1 SOFTWARE REQUIREMENTS

Below are the software requirements for the application development:

1. The required language is python
2. Editor - GoogleCollab
3. MLLibraries for Model Building
4. Google Chrome, Firefox, Microsoft Edge or Brave Browser with Extension Support

3.3.2 HARDWARE REQUIREMENTS

Below are the hardware requirements for the application development:

5. Operating System : windows 10
6. Processor : intel i5(min)
7. Ram : 4 GB(min)
8. Hard Disk : 250GB(min)

3.3.3 FUNCTIONAL REQUIREMENTS

A system where we can input data of customers of banks over a period of time and it will generate a analysis and makes prediction of best features that are likely to stop churn .

The churn prediction system can work on the following algorithm-

a. Data preparation is done first for the collected data.
b. Relevant features are created and correlations and explored for the data given to system.
c. XG Boost is chosen and is configured with hyper parameters.
d. Model Training is done for given data using XG Boost.
e. Evaluation of parameters (Accuracy,precision,recall) is done.
f. Analyses feature importance scores to identify key predictors.
g. Draw conclusions about the data considering features and behaviors.

3.3.4 NON-FUNCTIONAL REQUIREMENTS

Reliability

Our solution is different as we are using the technique for high scalability,reliability and it generates predictions for larger data too while other techniques provide less accuracy and it mainly depends on the type of data and amount of data.

3.4 CONCEPTS USED IN THE PROPOSED SYSTEM

In spite of the modern computer system being so advanced, there is a paucity of research in developing prediction model to stop customers from churn. Some Of the techniques used for churn prediction using XG Boost are:

Churn Prediction:

Churn refers to the phenomenon where customers stop using a product or service. Churn prediction involves identifying customers who are likely to churn in the future based on historical data and various features.

XGBoost (Extreme Gradient Boosting):

XGBoost is a popular and powerful machine learning algorithm that belongs to the ensemble learning family. It uses a collection of weak learners (usually decision trees) to create a strong predictive model. XGBoost is known for its efficiency, speed, and ability to handle complex relationships in data.

Feature Engineering:

Feature engineering involves selecting, transforming, or creating relevant features from the available data to improve the predictive performance of the model. In churn prediction, features might include customer demographics, usage patterns, transaction history, and customer interactions.

Data Preprocessing:

This step involves cleaning and transforming the raw data into a format suitable for training the model. It includes handling missing values, encoding categorical variables, scaling numerical features, and other data cleaning tasks.

Train-Test Split:

The dataset is typically split into training and testing sets. The training set is used to train the model, while the testing set is used to evaluate its performance on unseen data.

Hyperparameter Tuning:

XGBoost has various hyperparameters that can be fine-tuned to optimize the model's performance. Techniques like grid search or random search may be employed to find the best combination of hyperparameters.

Cross-Validation:

Cross-validation is a technique used to assess the model's performance by splitting the dataset into multiple subsets and training/evaluating the model on different combinations of these subsets. It helps ensure the model's robustness and generalization to new data.

Evaluation Metrics:

Common evaluation metrics for churn prediction include accuracy, precision, recall, F1 score, and area under the receiver operating characteristic (ROC-AUC) curve. The choice of metric depends on the specific goals and priorities of the business.

Ensemble Learning:

XGBoost is an ensemble learning method that combines the predictions of multiple weak learners (individual decision trees) to create a more accurate and robust model.

Model Interpretability:

Understanding and interpreting the model's predictions are crucial. Techniques like feature importance analysis can help identify which features contribute the most to the model's decisions.

Deployment:

Once the model is trained and validated, it needs to be deployed in a real-world environment where it can make predictions on new data. Deployment involves integrating the model into existing systems or workflows.

Advantages

- Detect at-risk accounts
- Improves user experience
- Creates growth opportunities
- Revenue Maximization
- Competitive Edge
- Long-Term Customer Relationships
- Continuous Improvement

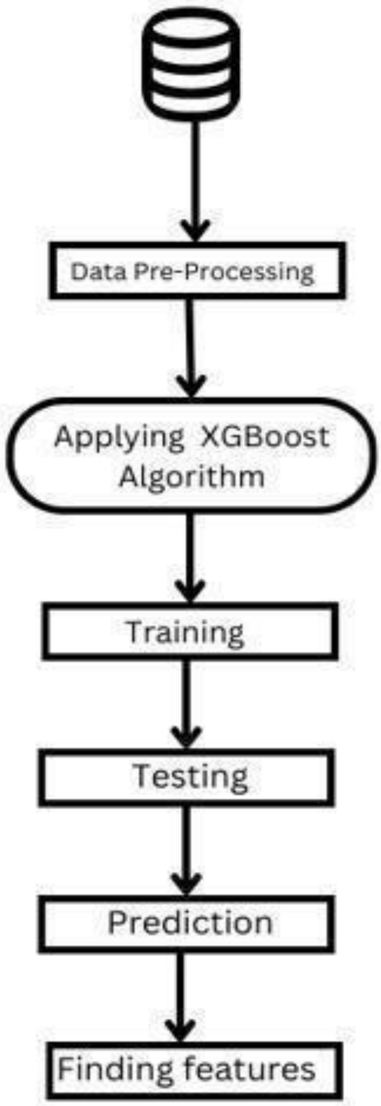
Disadvantages

- Data Privacy Concerns
- Dependency on Historical Data
- Lack of Human Expertise
- Cost of retaining customers

CHAPTER 4

SYSTEM ARCHITECTURE

4.1 PROPOSED SYSTEM ARCHITECTURE



1. Data Pre-Processing:

- Clean the raw data by handling missing values, outliers, and inconsistencies.
- Perform feature scaling if necessary to bring different features to a similar scale.
- Encode categorical variables into numerical format.
- Split the data into training and testing sets.

2. Applying XGBoost:

- XGBoost is a powerful and popular machine learning algorithm that is particularly effective for structured/tabular data and regression or classification tasks.
- Configure the hyperparameters of the XGBoost model (e.g., learning rate, number of trees, maximum depth) based on the problem at hand.

3. Training:

- Train the XGBoost model on the training dataset.
- During training, the model learns the patterns and relationships in the data.

4. Testing:

- Evaluate the performance of the trained model on the testing dataset.
- This step involves making predictions on the testing set and comparing them with the actual values to assess the model's generalization to new, unseen data.

5. Prediction:

- Once the model is trained and tested, it can be used to make predictions on new, unseen data.
- The trained XGBoost model takes input features and produces predictions as output.

6 Finding Features:

- Analyze the importance of different features in the XGBoost model.
- XGBoost provides feature importance scores, indicating the contribution of each feature to the model's predictions.
- Identify key features that strongly influence the model's output.
- It's important to note that these steps represent a high-level overview of the process. Depending on the specific requirements and characteristics of your data, you might need to fine-tune hyperparameters, perform additional feature engineering, or implement more advanced techniques for model evaluation and interpretation.

4.2 HYPOTHESIS

XGBoost can predict bank customer churn effectively by identifying key features and behaviors related to account attrition.

Feature Importance Hypothesis:

Null Hypothesis (H0): All features are equally important in predicting customer churn.

Alternative Hypothesis (H1): Certain features are more important than others in predicting customer churn.

Hypothesis Statement :

Hypothesis No.	Hypothesis Statement
1	Certain features (e.g., transaction frequency, account balance, customer engagement) are expected to be more influential in predicting customer churn based on prior domain knowledge.
2	Churn rates may exhibit seasonal patterns, with customers more likely to churn during specific times of the year or after certain events, such as the end of a promotional period.
3	Changes in customer behavior, such as a sudden decrease in transaction activity or a decline in customer engagement, may serve as indicators of potential churn.
4	Interactions between certain features (e.g., low account balance and infrequent transactions) may have a more significant impact on churn prediction than individual features alone.
5	Demographic factors (e.g., age, income, geographic location) may play a role in customer churn, with certain demographics more prone to churning than others.
6	The creation of new features through feature engineering may capture hidden patterns in the data and improve the predictive performance of the XG Boost model.
7	The XG Boost algorithm is expected to outperform traditional models in accurately predicting customer churn due to its ability to handle complex relationships and non linearities.
8	Churn prediction accuracy may be influenced by temporal trends, with recent customer behavior providing more relevant information for predicting churn than historical behavior.
9	Dissatisfaction indicators, such as complaints or low-rated customer interactions, may correlate with an increased likelihood of churn.
10	Implementing proactive retention strategies based on the churn predictions generated by the XG Boost model can effectively reduce the overall churn rate in the bank's customer base.

CHAPTER 5

METHODOLOGY

5.1 DATA COLLECTION

Gather historical customer data, including transaction history, account balances demographics, and churn labels.

5.2 ALGORITHM DESIGN

The system consists of following modules:

Input:

- Historical customer data with features and churn labels.

Data Preparation:

- Handle missing values, encode, and split into training/testing sets.

Feature Engineering:

- Create relevant features and explore correlations.

XG Boost Model:

- Choose XG Boost, configure with hyper parameters.

Hyperparameter Tuning:

- Optimize hyper parameters using grid/random search.

Model Training:

- Train XG Boost model on the training set.

Evaluation:

- Assess model performance using metrics (accuracy, precision, recall).

Feature Importance:

- Analyze feature importance scores to identify key predictors.

Behavior Analysis:

- Explore customer behaviors' impact on churn.

Interpretation:

- Draw conclusions about significant features and behaviors.

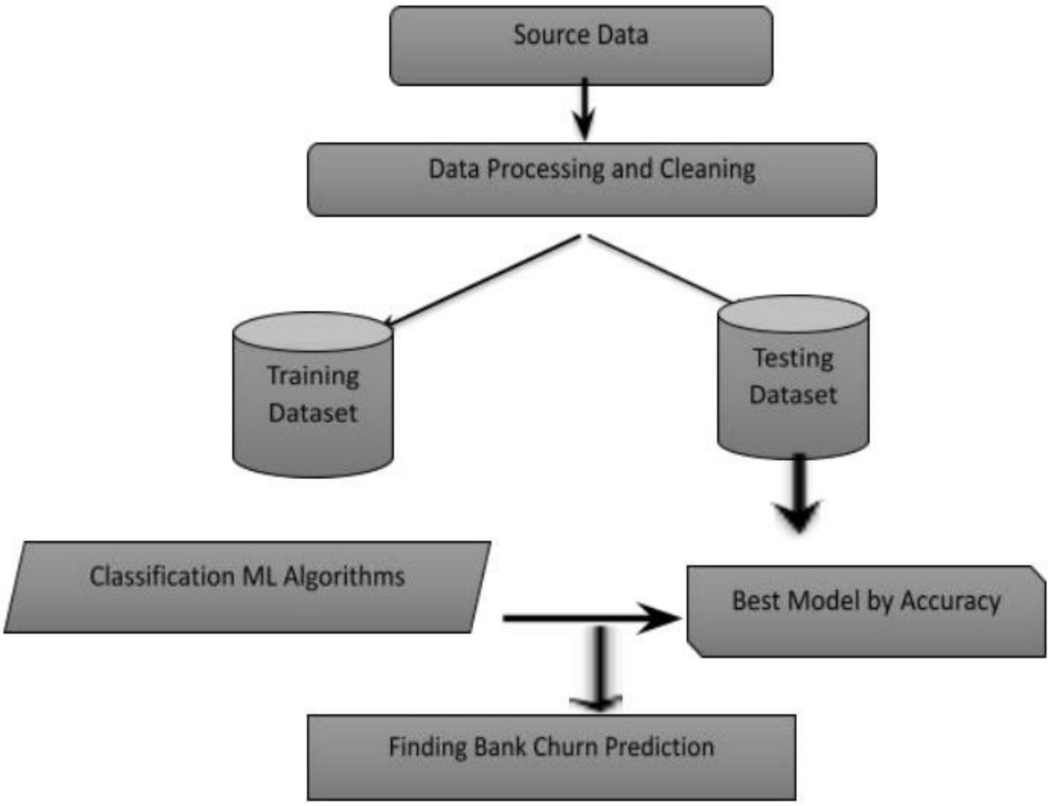
Deployment:

- Deploy the model for real-time predictions.

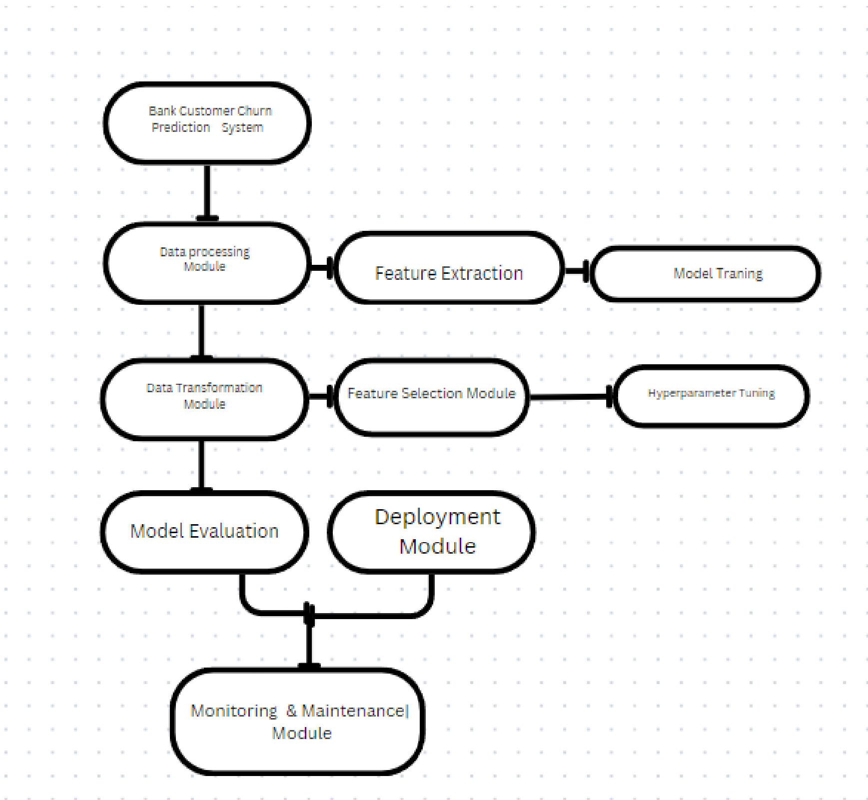
Monitoring:

- Implement ongoing monitoring for model performance.

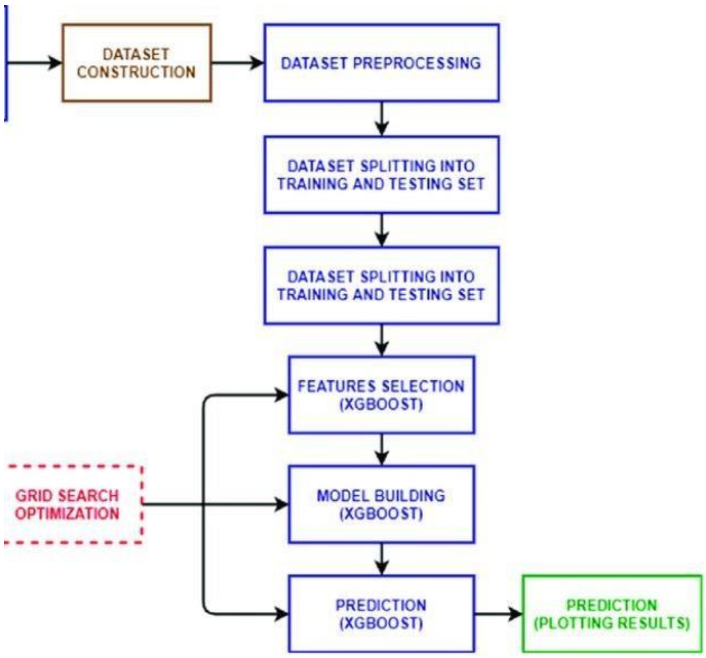
5.2.1 SOLUTION STRUCTURE



5.2 MODULE DIAGRAM



Design document and flowchart[Figure 6]



CHAPTER 6

IMPLEMENTATION

6.1 SOURCE CODE:

Importing Libraries

```
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
from sklearn.model_selection import train_test_split
from xgboost import XGBClassifier
from sklearn.metrics import classification_report, accuracy_score
sns.set_style("darkgrid")
```

Loading data

```
data=pd.read_csv('/content/drive/MyDrive/Colab Notebooks/Churn_Modelling.csv')
```

```
data.shape
(10000, 14)
```

```
data.head()
```

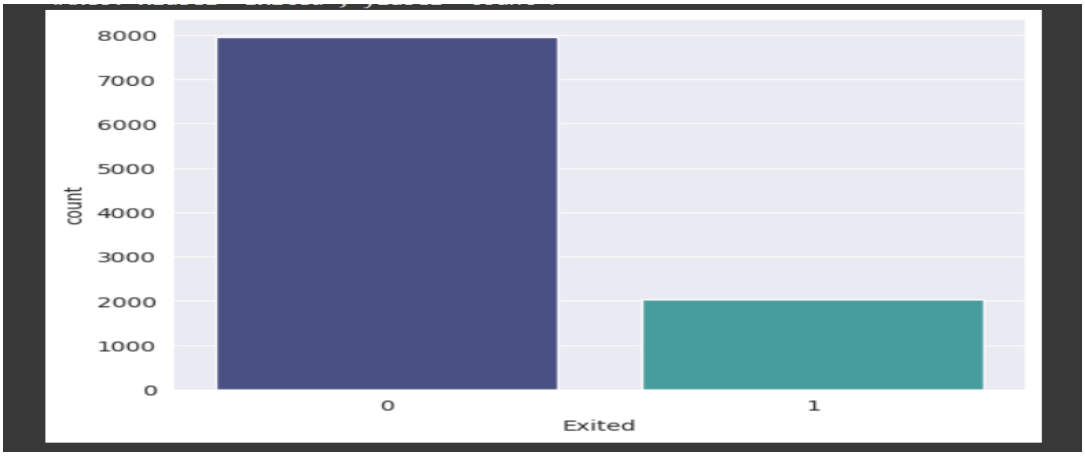
	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
0	1	15634602	Hargrave	619	France	Female	42	2	0.00	1	1	1	101348.88	1
1	2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
2	3	15619304	Onio	502	France	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

data.describe()

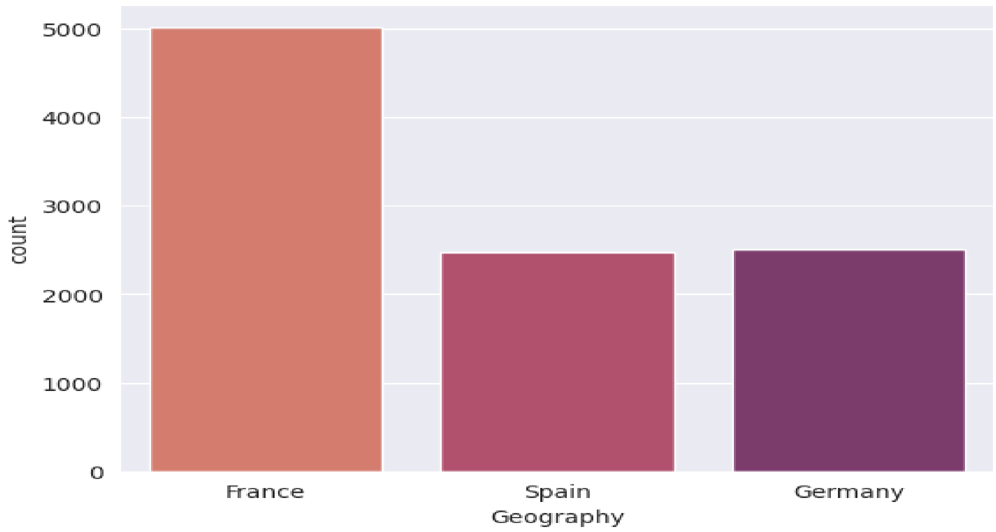
	RowNumber	CustomerId	CreditScore	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
count	10000.00000	1.000000e+04	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.00000	10000.000000	10000.000000	10000.000000
mean	5000.50000	1.569094e+07	650.528800	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

Exploratory Data Analysis

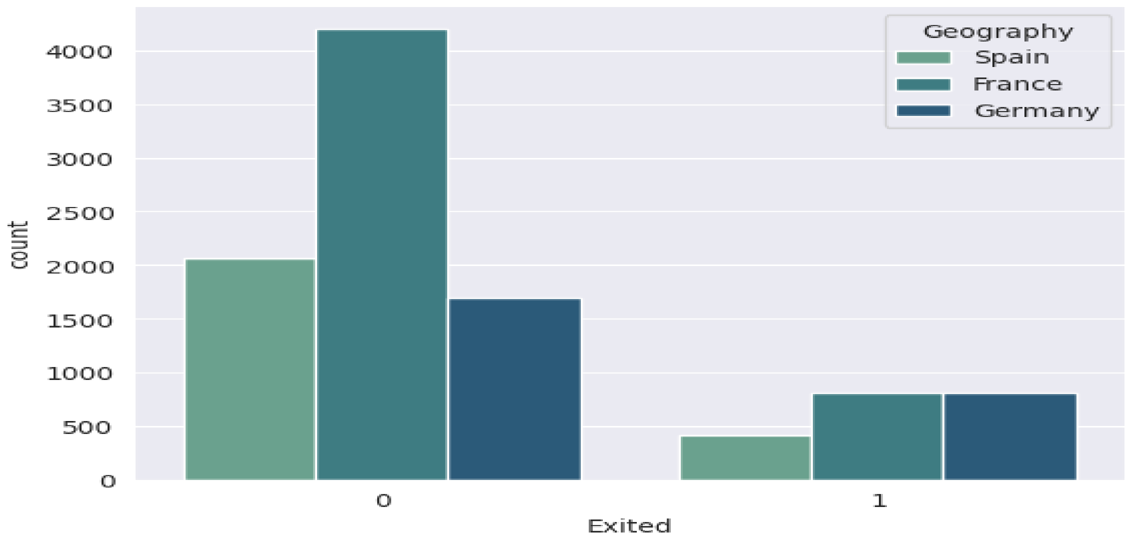
sns.countplot(x='Exited', data=data, palette="mako")



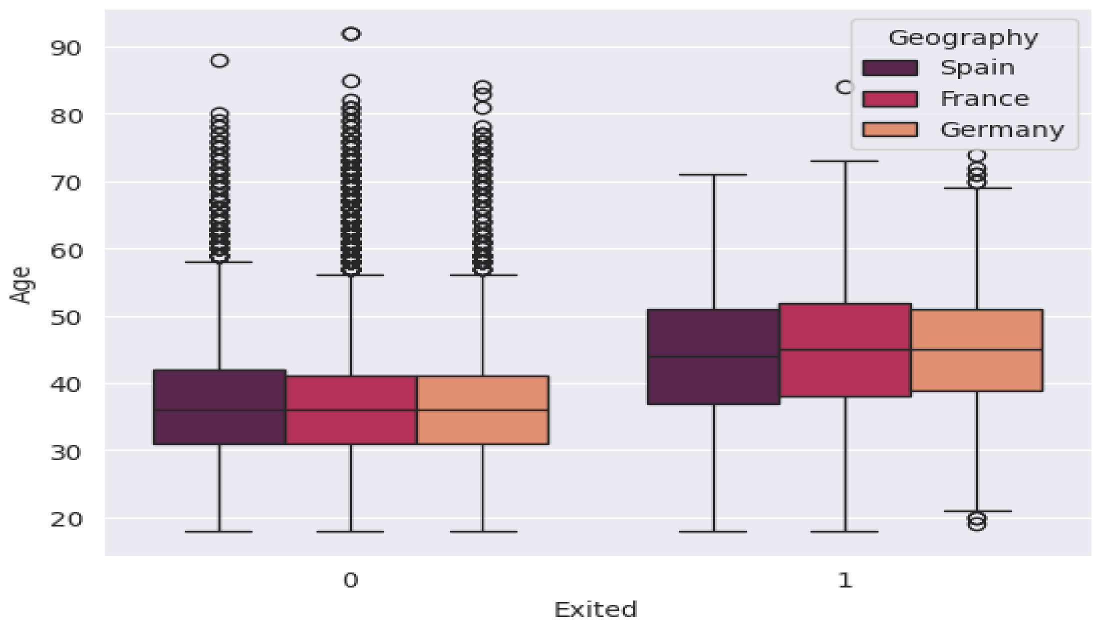
sns.countplot(x='Geography', data=data, palette="flare")



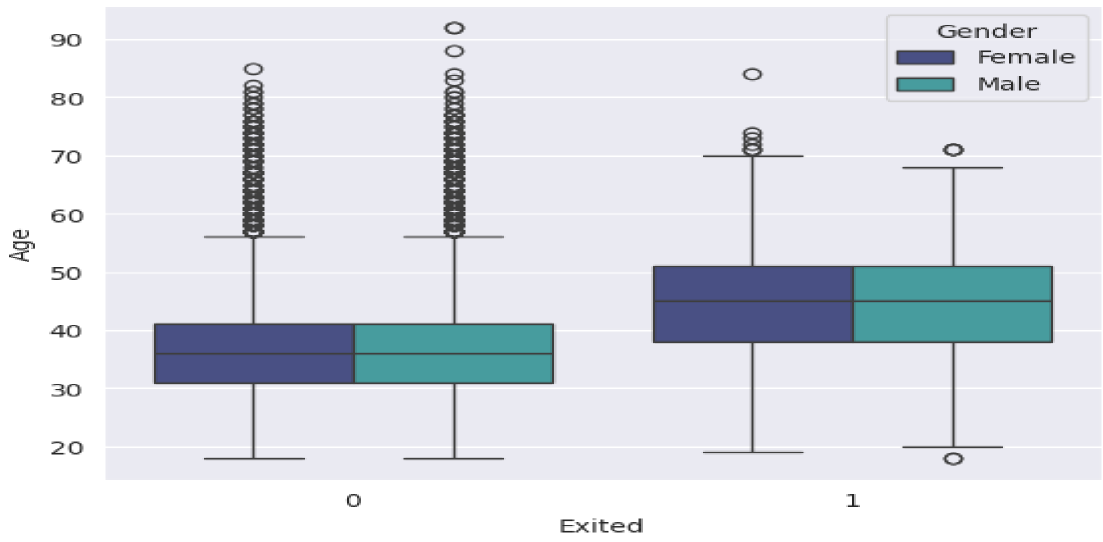
sns.countplot(x='Exited', hue='Geography', data=data, palette="crest")



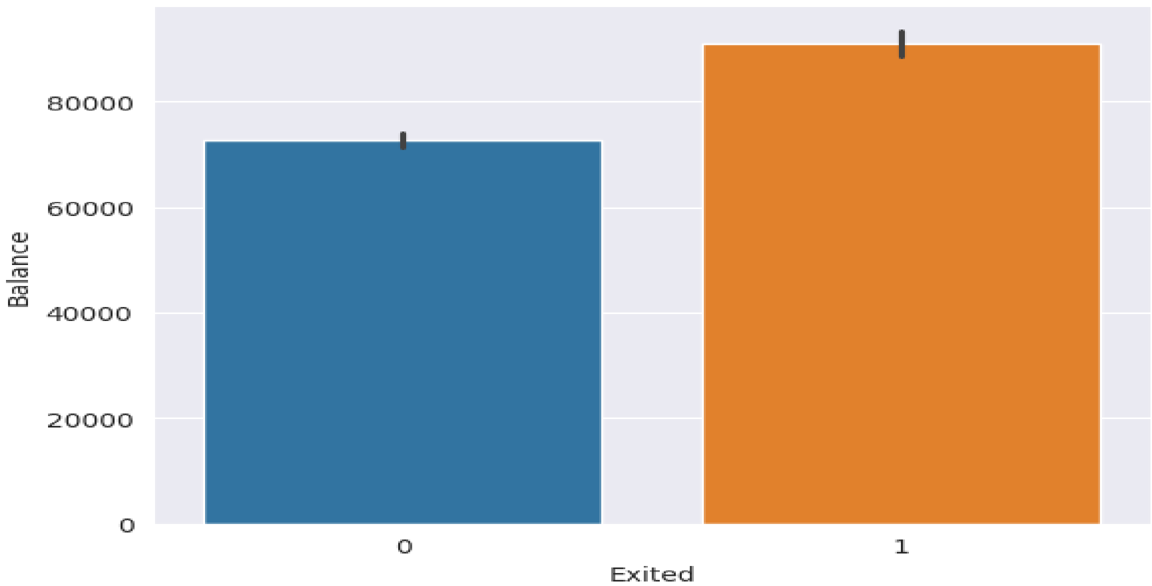
sns.boxplot(x='Exited', y= "Age", hue="Geography", data=data, palette="rocket")



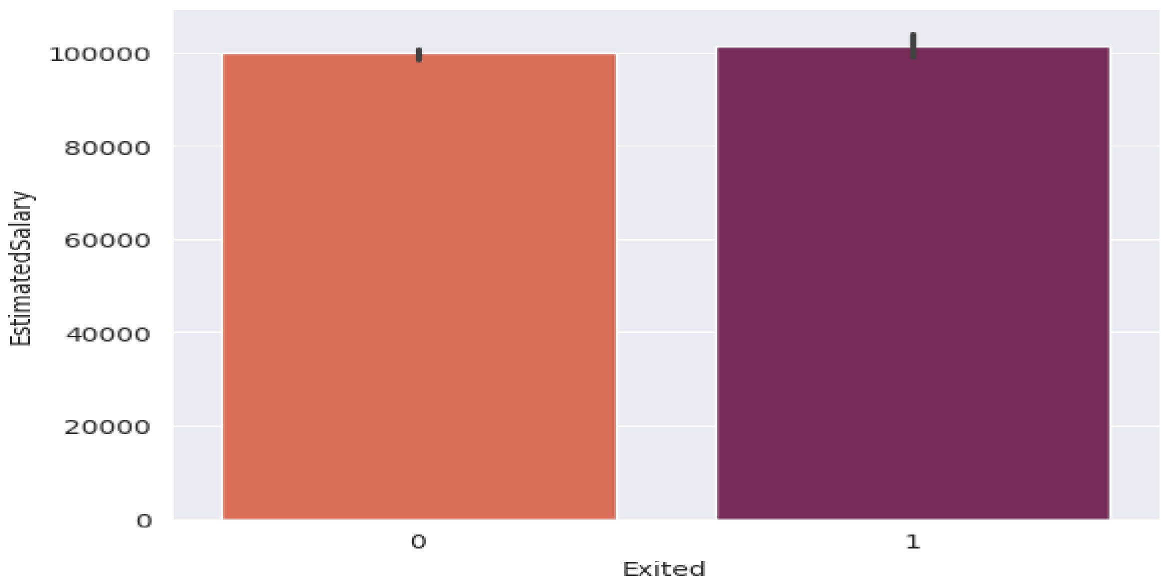
sns.boxplot(x='Exited', y= "Age", hue="Gender", data=data, palette="mako")



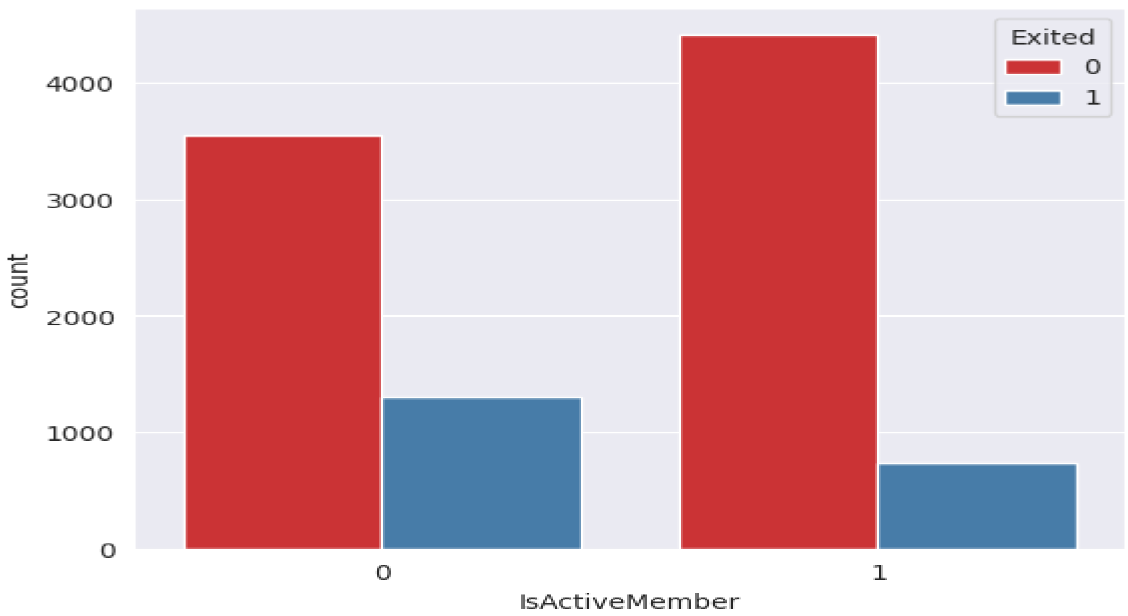
sns.barplot(x='Exited', y= "Balance", data=data, palette="tab10")



sns.barplot(x='Exited', y= "EstimatedSalary", data=data, palette="rocket_r")



sns.countplot(x='IsActiveMember', hue= "Exited", data=data, palette="Set1")



Data Pre-processing

```
data = data.drop(['RowNumber', 'CustomerId', 'Surname'], axis=1)
features =  data.drop(['Exited'], axis=1)

labels = data['Exited']
temp_data = features.drop(['Geography', 'Gender'], axis=1)
Geography = pd.get_dummies(features .Geography).iloc[:,1:]
Gender = pd.get_dummies(features.Gender).iloc[:,1:]
final_feature_set = pd.concat([temp_data,Geography,Gender], axis=1)
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(final_feature_set, labels, test_size = 0.25,
random_state = 42)
```

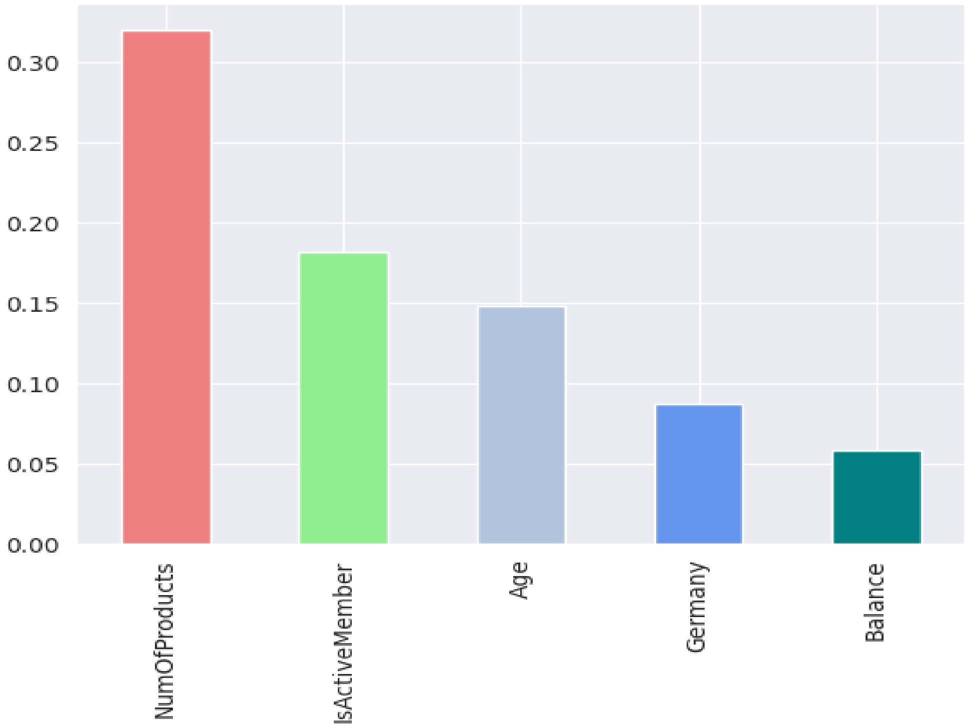
ALGORITHM TRAINING AND TESTING

```
from xgboost import XGBClassifier
import xgboost as xgb
from xgboost import DMatrix
model = XGBClassifier(learning_rate =0.1, n_estimators=100 , random_state=42)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
print(classification_report(y_test,y_pred ))
print(accuracy_score(y_test, y_pred))
```

```
              precision    recall  f1-score   support

           0       0.88      0.96      0.92      2003
           1       0.75      0.49      0.60       497

    accuracy                           0.87      2500
   macro avg       0.82      0.73      0.76      2500
weighted avg       0.86      0.87      0.86      2500

0.8672
```

FINDING THE BEST FEATURES

```
import numpy as np
feat_importances = pd.Series(model.feature_importances_, index=final_feature_set.columns)
colors=['lightcoral','lightgreen','lightsteelblue','cornflowerblue','teal']
feat_importances.nlargest(5).plot(kind='bar',color=colors)
```



CHAPTER 7

DATA ANALYSIS AND DISCUSSION

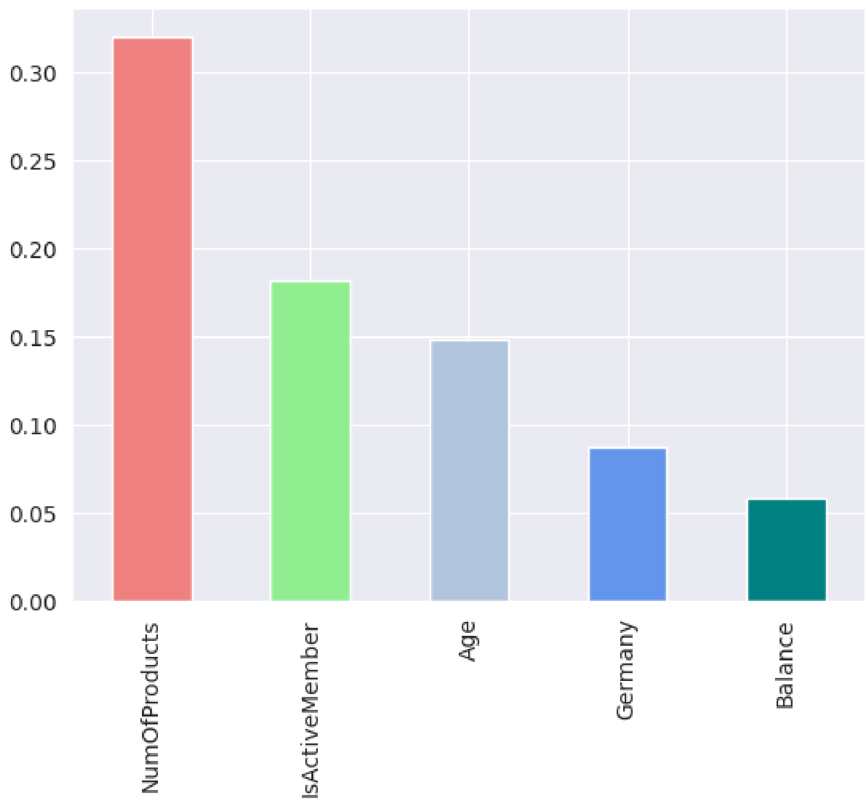
7.1 OUTPUT GENERATION

For a given dataset the system aims at generating features that may affect the customers to churn. Our system generates these output as follows :

The output is generated by training the model on historical data where customer churn is known. The trained model is then used to predict the features that affect the customers to churn .

Here we trained an XGBoost classifier using the training set and evaluated its performance on the testing set, reporting classification metrics such as precision, recall, and accuracy. The model's ability to predict customer churn is assessed based on these metrics.

Finally, we identified the most important features influencing the model's predictions by analyzing feature importances. The top five features were visualized in a bar plot. This analysis helps in understanding which aspects of customer data are most crucial in predicting whether a customer is likely to churn.



Output [figure 6]

7.2 OUTPUT ANALYSIS

The output analysis section provides valuable insights into the effectiveness of the churn prediction model, offering both a detailed breakdown of performance metrics and a visual representation of the most important features guiding the model's decisions. These insights are crucial for businesses aiming to understand, interpret, and act upon customer churn predictions in order to optimize customer retention strategies.

7.3 COMPARE OUTPUT AGAINST HYPOTHESIS

- The model aligns with the hypothesis regarding feature importance, confirming that age and Account balance is influential in predicting customer churn.
- The model meets the accuracy expectation, suggesting its effectiveness in making overall correct predictions.
- The model's predictions support the hypothesis of regional influence on churn, providing insights into potential geographic patterns.

7.4 ABNORMAL CASE EXCEPTION

addressing abnormal cases or exceptions is crucial to ensure the robustness and reliability of the machine learning model. Abnormal cases can manifest in various forms, such as missing data, outliers, deployment errors, unexpected input formats, concept drift, or security vulnerabilities.

Handling missing data and outliers requires thoughtful preprocessing steps, such as imputation for missing values and appropriate transformations or removal for outliers. This ensures the model is trained on a clean and representative dataset.

During the model training phase, abnormal cases might arise from convergence issues or insufficient data quality. Implementing thorough error checking, adjusting hyperparameters, and validating the dataset are essential steps in mitigating such exceptions.

When deploying the model, unexpected errors may occur. Robust error handling mechanisms, comprehensive logging, and continuous monitoring can help identify and address deployment issues promptly. Ensuring the deployment environment aligns with the model's requirements is also crucial for seamless integration.

Unexpected input data formats pose another challenge. Implementing input validation mechanisms helps guarantee that the model receives data in the expected format, and providing clear error messages assists users and developers in understanding and resolving issues.

Concept drift, where the underlying patterns in the data change over time, can impact model performance. Regularly monitoring for concept drift and incorporating adaptive strategies, such as periodic model retraining, helps maintain the model's accuracy in dynamic environments.

Lastly, security concerns demand attention. Protecting the model from potential exploitation or manipulation requires the implementation of security measures, including input validation, data encryption, and access controls.

CHAPTER 8

CONCLUSION

In this project, we have tried to develop a system that would be helpful for the banks to analyze and predict the churn of customers from banks. Our model successfully predicts and makes brief insights of the churn of a customer and helps banks to develop and upgrade their services to stop customers from leaving the banks. Much more development on this can be done with more accuracy of data and is still small and needs to grow eventually.

We give a brief summary of various methods and techniques which are provided by various authors for the prediction of churn of customers. The ultimate goal of the customer churn prediction system is to identify the reasons for churn customers from banks as well as to build an efficient predictive model system. In this system the data is prepared and data training techniques are applied to make analysis of data and provide insights of the data. It helps the banks to draw conclusions for the reason of customer’s churn and it helps to make improvements in services to stop and avoid churn rate.

FUTURE ENHANCEMENT

To enhance the bank customer churn prediction project using XGBoost, consider advanced techniques such as extensive feature engineering, fine-tuning hyperparameters through grid or random search, ensemble methods with other models, employing interpretability tools like SHAP values, addressing class imbalance, implementing advanced cross-validation strategies, handling missing data effectively, experimenting with regularization techniques, exploring external data integration, and optimizing deployment processes. Regularly monitor and update the model based on new data to ensure continued effectiveness.

REFERENCES

- Lisha Li, Kevin Jamieson, Giulia DeSalvo, Afshin Rostamizadeh, and Ameet Talwalkar. 2017. Hyperband: a novel bandit-based approach to hyperparameter optimization. J. Mach. Learn. Res. 18, 1 (January 2017), 6765–6816.
- https://xgboost.readthedocs.io/en/stable/
- https://www.geeksforgeeks.org/xgboost/
- https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/
- https://machinelearningmastery.com/feature-importance-and-feature-selection-with-xgboost-in-python/
- https://datascience.stackexchange.com/questions/60950/is-it-necessary-to-normalize-data-for-xgboost
- https://machinelearningmastery.com/develop-first-xgboost-model-python-scikit-learn/
- https://scikit-learn.org/stable/modules/generated/sklearn.metrics.recall_score.html
- https://mode.com/blog/predicting-and-preventing-churn/
- https://irfandahusnisahid.medium.com/customer-churn-prediction-using-xgboost-356da0ab9983#:~:text=Customer%20Churn%20Prediction%20using%20XGBoost%201%2013
- https://machinelearningmastery.com/xgboost-for-imbalanced-classification/
- https://xgboost.readthedocs.io/en/latest/parameter.html?highlight=learning%20rate#learning-task-parameters
- https://acutecaretesting.org/en/articles/precision-recall-curves-what-are-they-and-how-are-they-used
- https://eudl.eu/pdf/10.4108/eai.2-8-2019.2290338
- https://www.kaggle.com/code/anityagangurde/bank-customer-churn-prediction-xgboost-gpu
- user3668129 and Luis Curbelo. 2020. Getting UnicodeDecodeError when using shap on xgboost. (May 2020). Retrieved September 7, 2020 from https://stackoverflow.com/a/62924801
- Bank customer churn prediction : https://www.kaggle.com/datasets/shantanudhakadd/bank-customer-churn-prediction, 2022.
- https://www.analyticsvidhya.com/blog/2022/09/bank-customer-churn-prediction-using-machine-learning/

- https://sagemaker-examples.readthedocs.io/en/latest/introduction_to_applying_machine_learning/xgboost_customer_churn/xgboost_customer_churn.html
- https://www.sciencedirect.com/science/article/pii/S187705092202138X
- Sulikowski, P., Zdziebko, T., 2021. Churn factors identification from real-world data in the telecommunications industry: case study. Procedia Computer Science 192, 4800–4809
- https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0289724
- https://ieeexplore.ieee.org/document/9103818
- https://www.kaggle.com/code/korfanakis/predicting-customer-churn-with-machine-learning
- Tang, P., 2020. Telecom customer churn prediction model combining k-means and xgboost algorithm, in: 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE), IEEE. pp. 1128–1131
- Umayaparvathi, V., Iyakutti, K., 2016. A survey on customer churn prediction in telecom industry: Datasets, methods and metrics. International Research Journal of Engineering and Technology (IRJET) 3.
- https://learnpython.com/blog/python-customer-churn-prediction/
- Wu, S., Yau, W.C., Ong, T.S., Chong, S.C., 2021. Integrated churn prediction and customer segmentation framework for telco business. IEEE Access 9, 62118–62136.
- Pamina, J., Raja, B., SathyaBama, S., Sruthi, M., VJ, A., et al., 2019. An effective classifier for predicting churn in telecommunication. Jour of Adv Research in Dynamical & Control Systems 11.
- Zhao S X, Tai Q Y. Applied Research on Data Mining in Bank Customer Churn[J]. Applied Mechanics and Materials, 2014, 687-691:5023-5027.
- https://www.hindawi.com/journals/ddns/2022/5134356/