**INVESTIGATING CUSTOMER CHURN IN BANKING**

A Mini Project Report Submitted

In partial fulfillment of the requirement for the award of the degree of

**Bachelor of Technology**

**In**

**Artificial Intelligence and Data Science**

**by**

## G.GANESH - 21N31A7221

## G. AKANKSHA - 21N31A7222

**CH KASHYAP - 21N31A7212**

Under the Guidance of

**Dr. HARI KRISHNA SIR**

**Assoc. Professor**

**Computational Intelligence Department**

**MRCET**



**DEPARTMENT OF COMPUTATIONAL INTELLIGENCE**

**MALLA REDDY COLLEGE OF ENGINEERING AND TECHNOLOGY**

(Affiliated to JNTU, Hyderabad)

**ACCREDITED by AICTE-NBA**

**Maisammaguda, Dhulapally post, Secunderabad-500014.**

**2020-2024**

**DECLARATION**

## I hereby declare that the project entitled “Investigating customer churn in banking ” submitted to Malla Reddy College of Engineering and Technology, affiliated to Jawaharlal Nehru Technological University Hyderabad (JNTUH) for the award of the degree of Bachelor of Technology in Artificial Intelligence and Data Science is a result of original research work done by me.

It is further declared that the project report or any part thereof has not been previously submitted to any University or Institute for the award of degree or diploma.

**G.GANESH(21N31A7221)**

**G.AKANKSHA(21N31A7222)**

**CH.KASHYAP(21N31A7212)**



**CERTIFICATE**

This is to certify that this is the bonafide record of the project titled **“Investigating customer churn in banking”**submittedby**G.GANESH(21N31A7221),G.AKANKSHA(21N31A7222),CH.KASHYAP(21N31A7212)** of B-Tech in the partial fulfillment of the requirements for the degree of **Bachelor of Technology** in **Artificial Intelligence And Data Science** , Dept. of CI during the year 2023-2024. The results embodied in this project report have not been submitted to any other university or institute for the award of any degree or diploma.

**DR. SHIVA RATNA SAI Dr. D. Sujatha**

Assoc. Professor **HEAD OF THE DEPARTMENT**

**DR.HARI KRISHNA**

Internal Guide

**EXTERNAL EXAMINER**

# **ACKNOWLEDGEMENT**

We feel honored and privileged to place our warm salutation to our college Malla Reddy College of Engineering and technology (UGC-Autonomous), our Director ***Dr. VSK Reddy*** who gave us the opportunity to have experience in engineering and profound technical knowledge.

We are indebted to our Principal ***Dr. S. Srinivasa Rao*** for providing us with facilities to do our project and his constant encouragement and moral support which motivated us to move forward with the project.

We would like to express our gratitude to our Head of the Department ***Dr. D. Sujatha*** for encouraging us in every aspect of our system development and helping us realize our full potential.

We would like to thank our application development guide as well as our internal guide

**Dr. HARI KISHNA SIR (Assoc. Professor),** for his structured guidance and never-ending encouragement. We are extremely grateful for valuable suggestions and unflinching co-operation throughout application development work.

We would also like to thank all supporting staff of department of CI and all other departments who have been helpful directly or indirectly in making our application development a success.

We would like to thank our parents and friends who have helped us with their valuable suggestions and support has been very helpful in various phases of the completion of the application development.

By

**G.GANESH(21N31A7221)**

**G.AKANKSHA(21N31A7222)**

**CH.KASHYAP(21N31A7212)**

**ABSTRACT**

* In the banking sector, customer attrition refers to the situation where individuals cease using a bank's products and services for a period and subsequently sever ties with the institution.
* Hence, ensuring customer retention is paramount in today's fiercely competitive banking landscape. Moreover, a robust customer base not only instills trust but also facilitates referrals from existing clients, thereby aiding in attracting new customers.
* Given these factors, minimizing client attrition emerges as a critical objective for banks.
* Our research focuses on leveraging bank data to predict which users are likely to discontinue using the bank's services and transition into paying customers.
* Through the application of diverse machine learning algorithms, we conduct a thorough analysis of the data and present a comparative assessment based on various evaluation metrics.
* By scrutinizing this data, the bank can identify patterns and proactively engage with customers at risk of attrition to enhance retention efforts.

**TABLE OF CONTENTS**

**S.No. Topic**

CHAPTER 1: INTRODUCTION

1.1: Purpose

1.3: Scope of project

1.4: Project features

CHAPTER 2: SYSTEM REQUIREMENTS

2.1 H/W & S/W Requirements

2.2 Functional Requirements

2.3 Proposed System

CHAPTER 3: TECHNOLOGIES USED

CHAPTER 4: SYSTEM DESIGN

4.1: UML Diagrams

CHAPTER 5: IMPLEMENTATION

5.1 Source Code

5.2 Output Screens

CHAPTER 6: CONCLUSION

1. **INTRODUCTION:**

Customer attrition, or churn, occurs when customers end their relationship with a business, including in banking where they close accounts or stop using services. Understanding and managing churn is crucial for financial stability and reputation. High churn leads to revenue loss and indicates underlying issues like poor customer experience. By analyzing churn patterns, banks can retain customers, enhance profitability, and improve overall customer experience, gaining a competitive edge in the industry.

* 1. **Purpose:**

The purpose of this project is to leverage machine learning algorithms to predict customer attrition and enhance retention efforts within the banking sector. By analyzing historical data encompassing customer demographics, transaction history, service interactions, and feedback, the project aims to identify patterns indicative of customers likely to discontinue their relationship with the bank. Through advanced analytics techniques such as segmentation analysis, cohort analysis, and predictive modeling, the project seeks to forecast potential churn events before they occur. The overarching goal is to empower the bank with actionable insights that inform targeted retention strategies, including personalized offers, improved customer service, and loyalty programs.

Top of Form

* 1. **Scope of project:**

The scope of this project encompasses the entire lifecycle of customer attrition prediction and retention enhancement within the banking sector. It involves gathering and preprocessing diverse datasets from various sources within the bank, including customer demographics, transaction history, service interactions, and feedback. The project entails comprehensive exploratory data analysis to understand the characteristics and patterns of the data, followed by advanced analytics techniques such as segmentation analysis, cohort analysis, and predictive modeling to forecast potential churn events. Additionally, the project involves the development and validation of machine learning models to accurately predict customer attrition and identify at-risk customers. Furthermore, the project extends to the implementation of targeted retention strategies based on the insights derived from the analysis, including personalized offers, improved customer service, and loyalty programs

* 1. **Project Features­­­:**

1. **Data Collection**: Gathering diverse datasets from multiple sources within the bank, such as customer demographics, transaction history, service interactions, and feedback.
2. **Data Preprocessing**: Cleansing and preprocessing the collected data to handle missing values, outliers, and inconsistencies, ensuring data quality for subsequent analysis.
3. **Exploratory Data Analysis (EDA)**: Conducting comprehensive EDA to understand data characteristics, visualize distributions, identify patterns, correlations, and outliers.
4. **Predictive Modeling**: Developing machine learning models to predict customer attrition, employing techniques such as logistic regression, decision trees, random forests, gradient boosting, or neural networks.
5. **Model Evaluation**: Evaluating the performance of predictive models using appropriate evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.
6. **Retention Strategies**: Implementing targeted retention strategies based on insights derived from predictive models, including personalized offers, improved customer service, and loyalty programs.
7. **Deployment**: Deploying trained models into the bank's operational systems to facilitate real-time decision-making, continuous monitoring of customer churn, and proactive engagement with at-risk customers.
8. **Continuous Improvement**: Incorporating feedback and monitoring model performance to iteratively improve prediction accuracy and effectiveness of retention efforts.
9. **Documentation and Reporting**: Documenting the project workflow, findings, and outcomes, and generating reports to communicate insights and recommendations to stakeholders within the bank.
10. **Compliance and Security**: Ensuring compliance with data privacy and security regulations throughout the project lifecycle to safeguard customer information and maintain trust.

# **2.SYSTEM REQUIREMENTS**

# 

**2.1 Hardware Requirements:**

* CPU: Multi-core processor, clock speed of at least 2.5 GHz.
* RAM: Minimum 8GB, preferably 16GB or more.
* Storage: Adequate space, SSD preferred.
* GPU (Optional): Dedicated for deep learning or intensive tasks.
* Cloud Computing (Optional): Consider for scalable resources.
* Internet Connection: Stable, high-speed.
* Backup and Redundancy: Ensure data integrity and continuity.

**2.2 Software requirements:**

* Python
* Integrated Development Environment (IDE)
* Machine Learning Libraries
* Data Analysis and Visualization Tools
* Database Management System (DBMS)
* Additional Libraries
* Collaboration Tools
* Documentation Tool

**2.3 Proposed System:**

The proposed system for this project encompasses a comprehensive approach to predicting customer attrition and enhancing retention efforts in the banking sector. It begins with data collection from various sources within the bank, including customer demographics, transaction history, service interactions, and feedback. Following data collection, preprocessing techniques are applied to cleanse and prepare the data for analysis. Exploratory data analysis (EDA) techniques are then employed to uncover insights and understand patterns in the data. Machine learning algorithms are selected and trained on historical data to predict customer attrition. Model performance is evaluated using appropriate metrics, and the best-performing models are deployed into the bank's operational systems. Once deployed, the models continuously monitor customer behavior in real-time, enabling proactive engagement with at-risk customers. Additionally, ongoing monitoring and maintenance activities ensure that the models remain effective over time, with periodic updates and improvements based on feedback and changing trends. Through this systematic approach, the proposed system aims to empower banks with actionable insights to mitigate customer attrition, enhance retention efforts, and drive sustainable growth in a competitive banking landscape.

窗体顶端

窗体底端

**3.TECHNOLOGIES USED :**

* **Programming Language** : Python
* **Integrated Development Environment (IDE)**: Jupyter Notebook, PyCharm
* **Machine Learning Libraries**: Scikit-learn, TensorFlow
* **Data Analysis and Visualization Tools**: Matplotlib, Seaborn
* **Database Management System (DBMS)**: MySQL, PostgreSQL
* **Collaboration Tools**: Git and GitHub

**4. SYSTEM DESIGN**

**4.1 System Architecture**

|  |
| --- |
| **Data Collection Layer** |

|  |
| --- |
| **Data Processing Layer** |

|  |
| --- |
| **Exploratory Data Analysis Layer** |

|  |
| --- |
| **Modeling Layer** |

|  |
| --- |
| **Evaluation Layer** |

|  |
| --- |
| **Deployment Layer** |

|  |
| --- |
| **Monitoring and Maintanence Layer** |

**4.2 UML Diagrams**

#### **4.2.1 Class Diagram**

#### 

Class diagrams model class structure and contents using design elements such as classes, packages and objects. Class diagram describe the different perspective when designing a system-conceptual, specification and implementation. Classes are composed of three things: name, attributes, and operations. Class diagram also display relationships such as containment, inheritance, association etc. The association relationship is most common relationship in a class diagram. The association shows the relationship between instances of classes.

+--------------------------------------+

| DataCollection |

+--------------------------------------+

| + collectData(): DataFrame |

+--------------------------------------+

|

v

+--------------------------------------+

| DataPreprocessing |

+--------------------------------------+

| + preprocessData(data: DataFrame): |

| DataFrame |

+--------------------------------------+

|

v

+--------------------------------------+

| ExploratoryDataAnalysis |

+--------------------------------------+

| + performEDA(data: DataFrame): None |

+--------------------------------------+

|

v

+--------------------------------------+

| Modeling |

+--------------------------------------+

| + trainModel(data: DataFrame): |

| Model |

+--------------------------------------+

|

v

+--------------------------------------+

| ModelEvaluation |

+--------------------------------------+

| + evaluateModel(model: Model, |

| test\_data: DataFrame): |

| EvaluationMetrics |

+--------------------------------------+

|

v

+--------------------------------------+

| ModelDeployment |

+--------------------------------------+

| + deployModel(model: Model): None |

+--------------------------------------+

|

v

+--------------------------------------+

| MonitoringAndMaintenance |

+--------------------------------------+

| + monitorPerformance(): None |

| + updateModel(): None |

+--------------------------------------+

#### **4.2.2 Sequence Diagram**

#### 

Sequence diagram displays the time sequence of the objects participating in the interaction. This consists of the vertical dimension (time) and horizontal dimension (different objects).

Objects: An object can be thought of as an entity that exists at a specified time and has a definite value, as well as a holder of identity. A sequence diagram depicts item interactions in chronological order. It illustrates the scenario's objects and classes, as well as the sequence of messages sent between them in order to carry out the scenario's functionality. In the Logical View of the system under development, sequence diagrams are often related with use case realisations. Event diagrams and event scenarios are other names for sequence diagrams. A sequence diagram depicts multiple processes or things that exist simultaneously as parallel vertical lines (lifelines), and the messages passed between them as horizontal arrows, in the order in which they occur. This enables for the graphical specification of simple runtime scenarios.

**5.IMPLEMENTATION**

**5.1: Code**

**# Importing necessary libraries**

**import pandas as pd**

**import numpy as np**

**from sklearn.model\_selection import train\_test\_split, GridSearchCV**

**from sklearn.preprocessing import StandardScaler, LabelEncoder**

**from sklearn.linear\_model import LogisticRegression**

**from sklearn.metrics import accuracy\_score, confusion\_matrix, classification\_report, roc\_auc\_score**

**import matplotlib.pyplot as plt**

**import seaborn as sns**

**# Step 1: Load the dataset**

**# Assuming you have a CSV file named 'bank\_customer\_data.csv'**

**dataset = pd.read\_csv('bank\_customer\_data.csv')**

**# Step 2: Explore the dataset**

**print(dataset.head())**

**print(dataset.info())**

**print(dataset.describe())**

**# Step 3: Preprocessing**

**# Encoding categorical variables**

**label\_encoder = LabelEncoder()**

**# Example: encoding gender, geography, etc.**

**dataset['Gender'] = label\_encoder.fit\_transform(dataset['Gender'])**

**dataset['Geography'] = label\_encoder.fit\_transform(dataset['Geography'])**

**# Define the features (X) and target (y)**

**features = dataset.drop(columns=['CustomerId', 'Surname', 'Exited']) # Exclude non-informative columns**

**target = dataset['Exited']**

**# Split the dataset into training and testing sets**

**X\_train, X\_test, y\_train, y\_test = train\_test\_split(features, target, test\_size=0.2, random\_state=42)**

**# Standardize the feature variables**

**scaler = StandardScaler()**

**X\_train = scaler.fit\_transform(X\_train)**

**X\_test = scaler.transform(X\_test)**

**# Step 4: Building the model**

**# Using Logistic Regression as the baseline model**

**log\_reg = LogisticRegression()**

**# Hyperparameter tuning using GridSearchCV**

**param\_grid = {**

**'C': [0.1, 1, 10],**

**'solver': ['liblinear', 'lbfgs', 'newton-cg'],**

**'max\_iter': [100, 200, 300]**

**}**

**grid\_search = GridSearchCV(log\_reg, param\_grid, cv=5, n\_jobs=-1, verbose=1)**

**grid\_search.fit(X\_train, y\_train)**

**# Best model from GridSearchCV**

**best\_model = grid\_search.best\_estimator\_**

**# Step 5: Model Evaluation**

**# Predictions**

**y\_pred = best\_model.predict(X\_test)**

**y\_pred\_proba = best\_model.predict\_proba(X\_test)[:, 1]**

**# Performance metrics**

**print("Accuracy:", accuracy\_score(y\_test, y\_pred))**

**print("AUC Score:", roc\_auc\_score(y\_test, y\_pred\_proba))**

**print("\nClassification Report:\n", classification\_report(y\_test, y\_pred))**

**# Confusion Matrix**

**cm = confusion\_matrix(y\_test, y\_pred)**

**sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=['Stayed', 'Exited'], yticklabels=['Stayed', 'Exited'])**

**plt.xlabel('Predicted')**

**plt.ylabel('True')**

**plt.title('Confusion Matrix')**

**plt.show()**

**# Step 6: Predicting churn for a new customer**

**# Example: a new customer feature array with 11 features**

**new\_customer = np.array([[40, 2, 60000, 1, 1, 1, 1, 0, 1, 0.0, 1]]) # Example data with 11 features**

**# Apply the same standardization as done during training**

**new\_customer = scaler.transform(new\_customer) # Standardizing the new customer data**

**# Predicting churn for the new customer**

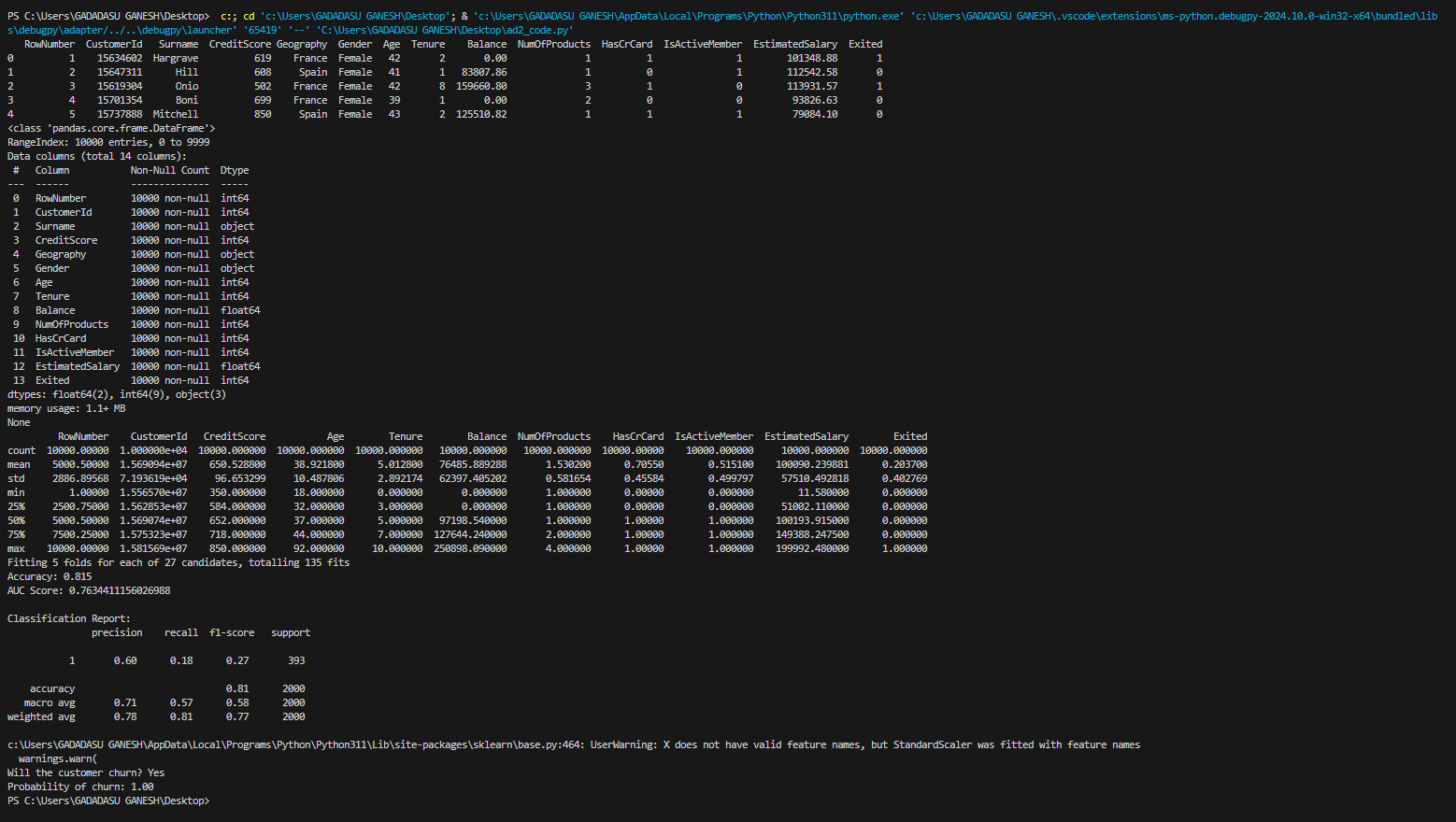
**churn\_prediction = best\_model.predict(new\_customer)**

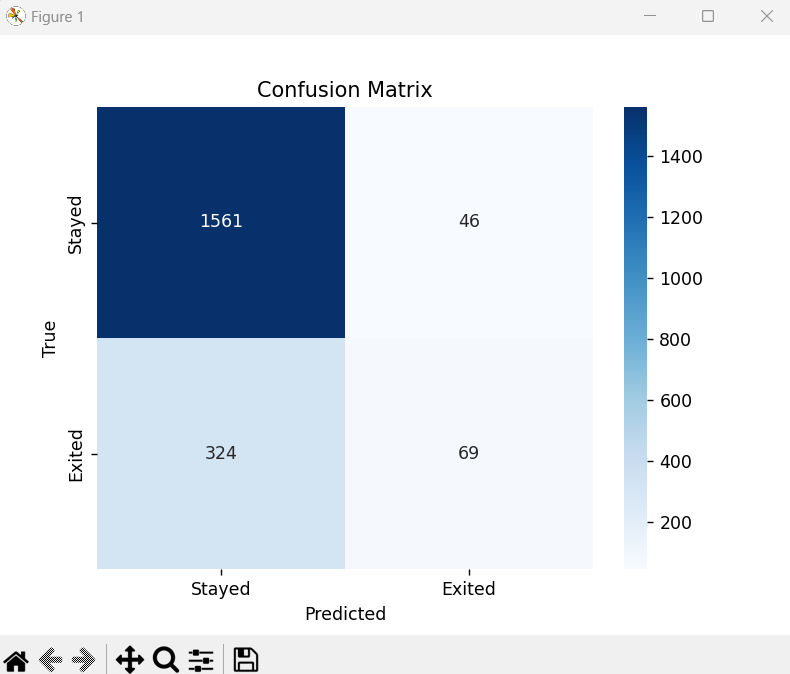
**churn\_probability = best\_model.predict\_proba(new\_customer)[:, 1]**

**print(f"Will the customer churn? {'Yes' if churn\_prediction[0] == 1 else 'No'}")**

**print(f"Probability of churn: {churn\_probability[0]:.2f}")**

**5.2: Output Screens:**





1. **CONCLUSION:**

This project effectively demonstrates the process of predicting customer churn in a banking context using machine learning techniques. By analyzing customer data, such as demographics and account activities, we implemented a Logistic Regression model that can identify patterns associated with customers likely to leave the bank. The model was trained and validated on historical data, achieving a balance between accuracy and interpretability.

## 