Introduction:

This project describes the details of the Python Capstone Project for ST1 unit within the outlined scope of the project requirements given to us in the assignment handout. We decided to work on the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ and Bank Customer Churn Prediction datasets.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Make your own paragraph here\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

People have been using banks for a long time, staying in one bank or switching from multiple. It can be difficult to keep track of the customers' finances and protect their money. Most banks do not store all their assets in physical cash to ensure the safety of their users’ money in case of emergency situations such as a bank robbery. Banks also struggle to keep track of whether customers are still using their services or not. It would be beneficial to the bank if a tool was made to determine whether a customer has left the bank or is still actively depositing and withdrawing their money.

Methodology:

There are 3 stages in the methodology used to create both software applications outlined below:

1. Design and develop decision support algorithms through data and predictive analyses to identify the best performing algorithm to solve a real-world problem.
2. Implementation of the best performing algorithm as a GUI using Tkinter
3. Deployment of best performing algorithm tool as a web or cloud enabled platform tool

Database set 1: Bank Customer Churn Rate

Stage 1: Algorithm Design Stage:

Stage 1 is the preliminary stage and is the most important stage of the 3. The algorithms used for the Exploratory and Predictive data analysis will vary in scope and complexity based on the problem and dataset being used. The workflow will go through dataset description, exploratory data analysis, predictive data analysis and model preparation and development to complete the algorithm design stage.

Dataset Description

There are 2 datasets being used for the project, both publicly available from the Kaggle website. The Bank Customer Churn rate database consists of 10,000 observations, 11 features and 1 target/class attribute. The 11 features include the collection of customers' activities in the bank along with any relevant banking information. The customer ID feature will not be used as it is merely used as a primary key to identify an individual customer. The class attribute is the churn rate result of the customers represented with 0 and 1. 1 means that the customer has left the bank at some time and 0 means that they have not. The data was originally collected and submitted to the Kaggle datasets by Gaurav Topre.

<https://www.kaggle.com/datasets/gauravtopre/bank-customer-churn-dataset?resource=download>

The aim of the project is to create two software tools to predict the results of the Churn rate in the Bank Customer Churn Rate database and the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Exploratory Data Analysis

After understanding the date, basic exploratory data analysis and visualization for both databases in the first phase, it was decided to use Google Colab to experiment with the dataset. Google Colab can be run directly in the web browser and uses virtual hardware and resources that do not require additional physical hardware. Scripts were all made using the coding language Python and the database is accessed using google drive. Some python libraries necessary to perform the exploratory data analysis have been imported in the google colab as well.

Predictive Data Analysis

To do the predictive data analysis the data must first go through several processing steps which may include pre-processing and performance evaluation.

Model Preparation and Development

All necessary components have been taken and incorporated into a script in Pycharm using Python. The Pycharm project will be available in the Github repository.

Stage 2: Algorithm Implementation Stage

Now that the best performing algorithm has been identified for both the Bank Customer Churn rate and \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ datasets in stage 1, the algorithm can now be implemented as a GUI using tkinter in Pycharm. The Pycharm project will be available in the Github repository.

Stage 3: Software Deployment Stage

The use of the software in Pycharm in stage 2 limited the applicable users and does not allow for the users in the bank for the bank customer churn prediction dataset or the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ . To prescribe the use of the application to these users the software has been implemented as a web-based platform using Flask API. Flask allows us to turn the model into a website application which you can view in the Github repository.

Conclusion

This report is for the ST1 Capstone project for design, development, implementation and deployment of prediction software using Python and the two datasets, Bank customer churn rate and \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. These platforms can be used to provide insight and historical data, allowing the user to predict the future risk or probability of an event occurring. Both databases have been designed through a desktop software tool and a web-based tool, allowing a wide application of the tools towards banks looking to find ways to keep customers and for \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

References

Gaurav Topre (2022), Bank Customer Churn Dataset, Kaggle, <https://www.kaggle.com/datasets/gauravtopre/bank-customer-churn-dataset?resource=download>

University of Canberra (2023), ST\_Capstone\_Project\_Report\_Template, University of Canberra Canvas page, <https://uclearn.canberra.edu.au/courses/13733/assignments/107529>