Blue text on a black background

Description automatically generated

**College of Engineering, Environment, and Computing**

**School of Science**

**BSc Computer Science**

**6001CEM Computing Individual Research Project**

**Project Report**

***Predicting Customer Subscription to Term Deposits Using Machine Learning***

By

**Student Name - Md Raja Ansari**

**Student Id - 13444048**

**1st Supervisor: Dr. Huma Shah(Associate Professor)**

Submitted in partial fulfilment of the requirements for the Degree of Bachelor of Science in Computer Science

Academic Year: 2024/25

**Declaration of Originality**

I declare that this project is all my own work and has not been copied in part or in whole from any other source except where duly acknowledged.

As such, all use of previously published work (from books, journals, magazines, internet etc.) has been acknowledged by citation within the main report to an item in the References or Bibliography lists.

I also agree that an electronic copy of this project may be stored and used for the purposes of plagiarism prevention and detection.

**Statement of copyright**

I acknowledge that the copyright of this project report, and any product developed as part of the project, belong to Coventry University.

Support, including funding, is available to commercialize products and services developed by staff and students. Any revenue that is generated is split with the inventor/s of the product or service.

For further information please see [www.coventry.ac.uk/ipr](http://www.coventry.ac.uk/ipr) or contact [ipr@coventry.ac.uk](mailto:ipr@coventry.ac.uk).

**Statement of ethical engagement**

I declare that a proposal for this project has been submitted to the Coventry University ethics monitoring website (<https://ethics.coventry.ac.uk/projects/184846> ) and that the application number is listed below:

Signed: Md Raja Ansari

Date: 10/04/2024

Please complete all fields.

|  |  |
| --- | --- |
| First Name: | **Md** |
| Last Name: | **Ansari** |
| Student ID number | **13444048** |
| Ethics Application Number | P184846 |
| 1st Supervisor Name | **Dr. Huma Shah** |
| 2nd Supervisor Name |  |

Table of Contents

[Abstract 7](#_Toc194836324)

[Chapter 1: Introduction 8](#_Toc194836325)

[1.1: Background 8](#_Toc194836326)

[1.2: Research Problem 8](#_Toc194836327)

[1.3: Aim 8](#_Toc194836328)

[1.4: Objectives 8](#_Toc194836329)

[1.5: Research Question 8](#_Toc194836330)

[1.6: Research Methodology 9](#_Toc194836331)

[1.7: Structure of the report: 9](#_Toc194836332)

[Chapter 2: Literature review 10](#_Toc194836333)

[2.1: Introduction: 10](#_Toc194836334)

[2.2: Factors responsible for customer acquisition in banks: 10](#_Toc194836335)

[2.3: Machine learning in banks: 11](#_Toc194836336)

[2.4: Past work: 12](#_Toc194836337)

[2.5: Summary: 14](#_Toc194836338)

[Chapter 3: Research methodology: 15](#_Toc194836339)

[3.1: Introduction: 15](#_Toc194836340)

[3.2: Research type: 15](#_Toc194836341)

[3.3: Strategies for qualitative work: 15](#_Toc194836342)

[3.4: Exploratory data analysis (EDA): 15](#_Toc194836343)

[3.5: Practical steps: 16](#_Toc194836344)

[3.6: Ethical duties: 17](#_Toc194836345)

[Chapter 4: Results and findings 19](#_Toc194836346)

[4.1: Introduction: 19](#_Toc194836347)

[4.2: Python libraries: 19](#_Toc194836348)

[4.3: EDA results: 22](#_Toc194836349)

[4.4: Machine learning results: 26](#_Toc194836350)

[4.5: Thematic analysis results: 33](#_Toc194836351)

[4.6: Discussion: 35](#_Toc194836352)

[Chapter 5: Conclusion and recommendations 37](#_Toc194836353)

[5.1: Conclusion: 37](#_Toc194836354)

[5.2: Recommendations: 37](#_Toc194836355)

[5.3: Future scope: 37](#_Toc194836356)

[5.4: Reflection: 38](#_Toc194836357)

[References: 39](#_Toc194836358)

# Abstract

The purpose of the current project was to be developing the machine learning model for the prediction of customer subscription for term deposits. The author embraced the mix research method because the qualitative method helped in exploring the reasons behind low rate of customer subscription for term deposits and challenges for customer acquisition. On the other hand, the quantitative work was linked with machine learning experiment.

Here, the author took data of banking customers (45,211 records with 16 features) and applied four machine learning algorithms: Logistic Regression, Decision Tree, Gradient Boosting, and Multi-Layer Perceptron. The analysis revealed significant patterns: students (28.68%) and retirees (22.79%) showed the highest subscription rates, while customers aged 30-45 demonstrated stronger subscription tendencies. The dataset included comprehensive customer information such as age, job type, marital status, education level, credit history, and previous campaign outcomes.

Further analysis showed that customers with tertiary education had a higher subscription rate compared to those with primary or secondary education, and single customers demonstrated a 14.95% subscription rate, outperforming other marital status categories. Due to class imbalance (88.3% non-subscribers vs. 11.7% subscribers), the author applied the SMOTE method as well. The results from experimentation revealed that the accuracy for Gradient Boosting was 87% when SMOTE was applied, and it outperformed the other three models.

The model's performance metrics included precision, recall, and F1-score, with particular emphasis on correctly identifying potential subscribers while minimizing false positives that could lead to inefficient resource allocation in marketing campaigns.

Gradient boosting model successfully can inform about the customer behaviour and accordingly, the marketing strategies can be changed. The findings from thematic analysis revealed that low interest, low return on investment, higher competition and poor marketing are some of the challenges for banks in attracting new customers. Additionally, demographic insights showed that customers with tertiary education and single marital status had higher subscription probabilities, providing actionable targeting strategies for marketing campaigns.

# Chapter 1: Introduction

## 1.1: Background

TheBanco de Portugal is the central bank of Portugal and its offers various services to the customers. The service of Term Deposits is one such feature that is highly availed by the bank customers. But Banco de Portugal is facing a slight issue in predicting customer subscription to its Term Deposits service.

## 1.2: Research Problem

Banco de Portugal faces challenges in optimizing its marketing campaigns for term deposit subscriptions. The bank's traditional marketing efforts, such as phone call campaigns, often lead to low conversion rates of customers opting for term deposits. This has been a cause of concern for Banco de Portugal. Hence, this research is conducted to predict customer subscription to Term Deposits using Machine Learning algorithm.

## 1.3: Aim

The main aim of this study is that to make use of the AI-driven classification models for prediction of customer subscription regarding term deposit products by analysing transactional and demographic data of the customers in Portugal.

## 1.4: Objectives

* To determine the factors that are directly and indirectly responsible for fall in customer subscription to Term Deposits in Banco de Portugal.
* To find out the challenges faced by Banco de Portugal in attracting customers for subscribing to Term Deposits service.
* To make use of different Machine Learning (ML) algorithms for predicting the rate of customer subscription to Term Deposits in Banco de Portugal.
* To test and evaluate from among the selected ML models/algorithms for selection of the most appropriate model for addressing the problem of low customer subscription to Term Deposits.

## 1.5: Research Question

(**RQ1**) What factors have resulted in fall of Customer Subscription to Term Deposits in Banco de Portugal?

(**RQ2**) Which ML model/algorithm is the most appropriate for predicting customer subscription to Term Deposits in in Banco de Portugal?

## 1.6: Research Methodology

In this study, the main focus is given to correct of different ML algorithms and to properly check their validity for predicting Customer Subscription to Term Deposits, on the basis of comparative analysis among the models/algorithms. The different models/algorithms that are used here are mentioned as under.

* Logistic Regression
* Decision Tree
* Gradient Boosting
* Neural Network

The findings of the models will be based on a particular set of same datasets about Customers who have recently subscribed to Term Deposits in last 3-6 months.

All the ML models/algorithms will be applicable based on the data collected from the below mentioned URL/link.

*Link for the dataset is*- <https://archive.ics.uci.edu/dataset/222/bank+marketing>

## 1.7: Structure of the report:

The quality of the work can be determined by variety of factors and one of the elements to enhance the credibility and reliability of the work was to be arranging the content in meaningful manner. For the same purpose, the author arranged the study systematically in five core domains. The first chapter was documented to explain briefly about the chosen research methodology and most important sub-section in this section was the aims and objectives. The second chapter was the literature review which hold significance in terms of identifying the research gap and becoming familiar with the past research. The third chapter was the research methodology which has the description and justification about the practical steps. The data collection and its sources, the data transformation, model’s explanation all was done in this chapter. The fourth chapter has two parts; in first part, the results were discussed and in the second part, the discussion was made. This chapter was linked with the aims and objectives and even with the findings of previous works. The conclusion was mentioned in the final chapter, and it has recommendation sub-section too. The future scope for the topic was also described in the fifth chapter.

# Chapter 2: Literature review

## 2.1: Introduction:

The inclusion of the past studies is one of the most necessary elements into the literature review works. It was essential for the researcher to gather information about the arguments developed by the past authors. It helped in understanding the ways in which machine learning technology was deployed in several banking activities. The role of machine learning is growing, and it is still at the nascent stage. So, reviewing past studies gives an edge to the author to identify the research gap and accumulating relevant knowledge.

## 2.2: Factors responsible for customer acquisition in banks:

Bhatt & Jain (2020) mentioned that the banks who provide more convenience to the customers become natural choice among the customers. It is necessary that the financial products offered by the banks must generate additional value for the customers and it allow the banks to acquire new customers. The banks with a smaller number of customer complaints and high number of positive feedbacks also get large number of new customers. Further, the work of Broekhoff, Van der Cruijsen & De Haan (2024) provided different perspective that the banks with high trustworthy services have more chances of acquiring new customers. The banks deal with most sensitive products of the customers i.e. money and it is necessary that the banks must deliver the trustworthy and reliable services. The payment system should be transparent and the frauds must be eliminated from the online transactions. The advanced technologies should be deployed by the banks to improve the safety and security of mobile applications and internet banking facilities.

In another work Cardoso & Cardoso (2024) the author studied the impact of bank reputation and trust factor on the client satisfaction and loyalty factors. If the reputation of the bank become poor, then the existing customer stop using the services and loyalty factors become negative. The reputation can be built on multiple factors like no hidden charges, no manipulation in financial products and so on. The rate of deposits can go down if the reputation of bank is poor in terms of legal activities. The banks financial performance is major parameter in terms of acquiring new customers. Because if the bank’s financial performance will be poor, then their interest paying capacity of customer deposits will also be on the lower side. Therefore, these are major factors which determine the customer acquisition aspects for the banks. The work done by Shaikh et al. (2023) added into the argument that when customers get technically advanced digital services, then also the customer subscription or new customer acquisition become common scenario. In terms of deposits or taking any other services, the customer prefers the banks with high convenience. The access to digital technologies can allow to avail services in easiest way possible. Therefore, it become clear that the factors like convenience, trust, quality of services, use of technology and reputation are some of the factors which are directly or indirectly responsible to enhance the customer subscription in availing the bank’s services.

## 2.3: Machine learning in banks:

Likewise other industries and sectors, banks have also started to apply the machine learning algorithms and artificial intelligence technology to improve the overall business operations. It was found in the work of Heb & Damasio (2025) that banks can apply machine learning algorithms for the risk management. The models can predict the risks for the banks in advance and even the solutions can be proposed. Moreover, Singh et al. (2023) mentioned that the banks can use algorithms for the purpose of customer churn prediction as well. In the work of Vaduva et al. (2024) the author mentioned about the algorithms name which can be used for customer churn prediction in banks. Models like random forest, light gradient boosting machine and support vector machine and so on can be applied by the banks to predict the customer churn.

Polireddi (2024) discussed in detail that elements like risk reduction, improving customer relationship management, improving operations and even taking right decisions all can be done by the banks through machine learning algorithms. The customer satisfaction rate can be measured with the algorithms like neural networks and so on. Even in ensuring the cybersecurity like fraud detection in banks is also possible with the help of algorithms. The major reason behind the application of machine learning model is that it reduces the error, it remains cost effective, can work on big datasets and computation speed is also quick. The banks can use such model for credit card fraud identification, prediction of credit card spending and so on. The financial transactions can be tracked with the help of algorithms, as understood from the work of Pattnaik, Ray & Raman (2024).

The sentiment analysis can be done and it can help in predicting the customer’s possibility of buying any new financial product and so on. Through sentiment analysis, the banks can definitely focus on assessing the emotions of the customers, and the feedbacks of the customers (Andrian et al. 2022). Such insights can be used to assess the potential of newer customer subscription for new products like term deposits and so on. Overall, the machine learning is becoming an integral part of banking sector’s internal decision making. It can be used to improve internal operations and to understand the customer behaviour. Further, the positive impact of sentiment analysis is such that the personalised financial products can be offered to the customers. Through algorithms, the needs and demands of the customers can be identified and accordingly, the tailormade products and services can help in new customer subscription.

## 2.4: Past work:

The use of machine learning to predict the customer subscription can be made in different sectors like banks, online newspaper, online movie streaming and so on. The work done by Belchior, Antonio & Fernandes (2024) stated that whether the customer subscription for any product will increase or not, it completely depends upon the customer loyalty levels. If the customer relationship management is poor, then the customer query for new type of products cannot be generated easily. Thus, here the author developed the machine learning model to predict the customer subscription for online newspaper business model. Here the applied algorithms were decision tree, light gradient boosting model, random forest and extreme gradient boosting. According to the author, the regularisation techniques were applied to improve the performance of XG boost and light GBM model. It was found that the performance of XG boost model was highly effective with more than 95 percent accuracy.

Noori (2021) mentioned that actual understanding of the customer attitude towards the future purchases or repeat sales can be gained from the customer reviews. the opinions or sentiments expressed in the online reviews can be processed and then the satisfied or dissatisfaction factors can be filtered. It can be help in identifying that whether the customers can purchase more products and services from the brand or not. Here the author has focused on applying multiple algorithms to assess the customer reviews for the hotel sector. The used algorithms were artificial neural network, decision tree, naïve bayes, support vector machine, and KNN. Here, the data preprocessing was done through stemming, removal of stop words, tokenization and by eliminating the duplicate values. For feature extraction, the techniques like TF-IDF were deployed and word pruning was also done. Moreover, for feature selection, the principal component analysis method was put into application. Based on the 10-fold cross validation, the experimentation results revealed that when the number of features were increased, then the performance of decision tree model was effective. It was found that decision tree model outperformed other algorithms and it achieved the accuracy of 98.9 percent.

Shridhar & Ahwini (2024) developed the machine learning based model to predict the customer subscription particularly for Netflix. Here Python ARIMA method was deployed to predict the customer subscription. Once the data was gathered, then the author used the data pre-processing techniques like normalization and standardization. Other than this, the tools like autocorrelation function (ACF) and partial autocorrelation function (PACF) were deployed. The results in this paper revealed that the suggested model was able to understand the customer preferences and allow Netflix to allocate its resources in best possible manner to enhance the customer subscription rate. Wang, Wang & Xie (2022) also conducted study to forecast the subscriber numbers for Netflix. Here multiple linear regression models were deployed, and it was found that when dataset size is very large then such models show improper and inaccurate outcomes. There was the issue of multicollinearity, and it can be handled with large number of datasets only. So, it become clear that machine learning is more suitable technology to work on the prediction of customer subscription. The companies like Netflix or other sectors like banks and hotels always cater large number of customers, therefore, to get reliable outcomes, the robust algorithms must be designed.

Zaki et al. (2024) agreed that marketing decisions in the banks are now influenced by the machine learning techniques. The transformation has taken place because machine learning provides better data analytics tactics as compared to traditional methods. Here the author conducted research to predict the bank’s term deposit subscriptions. Firstly, the data was extracted from Kaggle and then the preprocessing methods like standardization and normalization were performed. Feature engineering was done using the label encoder technique where the categorical variables were converted into the numerical variables. The applied models in the current paper were stochastic gradient descent classifier, KNN, logistic regression, Gaussian naïve bayes, decision tree and random forest classifier. The evaluation of the models was done on the basis of F1 score, recall values and accuracy metrics. The final results revealed that the highest accuracy was obtained for random forest model and lowest accuracy was measured for stochastic gradient descent classifier. The accuracy for the random forest model was 87.5 percent and accuracy for stochastic gradient descent model was only 44 percent. The results also revealed that the random forest model can lay down effective ground for the improvements into the existing marketing strategies of the banks.

## 2.5: Summary:

Overall, the summary suggested that the machine learning is banks is essential in order to process the large number of datasets. The customer behaviour is dynamic phenomenon and with the help of sentiment analysis, the possibilities of future buying behaviours can be assessed. The algorithms were of different types like supervised or traditional (support vector machine and naïve bayes), ensemble (like decision tree and random forest), bagging and boosting methods, neural networks (like Convolutional Neural Network and Long Short-term Memory) and so on.

The identified research gap in the paper was that most of the study focus on customer churn with respect to banks and other financial institutions. Very few studies focused on predicting the new customer subscription for the particular predict range like term deposits; so, this study has helped in filling such gaps.

# Chapter 3: Research methodology:

## 3.1: Introduction:

Here the majorly discussed sub-themes were research type, steps for qualitative work and strategies embraced for the quantitative method. the experimental set-ups were discussed along with the ethical considerations.

## 3.2: Research type:

The selection of suitable research type was influenced by the research objectives. In this study, the qualitative as well as quantitative research methods were applied. The justification for qualitative method was the framing of first two objectives. In those objectives, the factors affecting the customer acquisition and factors affecting the customer retention were discussed. These factors were subjective in nature because different customer groups show different behaviour towards the services and products of the banks. So, to handle the subjective behind factors affecting customer retention and customer acquisition were tackled through the qualitative work. On the other hand, other two objective were clearly based on the machine learning algorithms and experiments by using codes and scripts. Therefore, the steps like model development and model evaluation were handled through quantitative method.

## 3.3: Strategies for qualitative work:

Here the interpretive research philosophy was deployed. Its justification was that the customer behaviour is very dynamic and there is no linearity in the customer’s responses towards the bank’s term deposits. In order to explore the multiple factors, the author focused on gathering multiple opinions and perspectives and it was done through interpretive philosophy. The data collection for the qualitative work was done with the help of secondary sources and there were online pages, journal articles, news articles and magazines and so on. The data analysis was done through thematic analysis technique.

## 3.4: Exploratory data analysis (EDA):

The first step within the quantitative work was that the EDA must be done on the gathered dataset. It was the data analytical technique which helped in understanding the customer’s attitude and bank’s strategic actions to acquire the customers. The EDA helped in understanding the relationship between variables and for better visualisation and interpretation, the charts and graphs were used.

## 3.5: Practical steps:

Data collection:

The study was based on the Portugal bank, so the dataset also contains the information of customers from the same country only. The dataset was picked from secondary sources and the name of such historical dataset was ‘bank marketing’. It has 17 columns and 17 features which were age, type of job, education, balance, education, loan type, campaign, and previous number of contracts and so on. With the help of such data, the author has focused on predicting the customer subscription to term deposits. The size of the dataset was large because here 45211 entries were included.

Data transformation:

It was necessary to clean the data so that noise within the gathered information can be eliminated and information capturing for the algorithms can be improved. For the same purpose, the missing values were identified. Apart from thus, the steps were taken to handle the outlier detection. Correlation was done for feature selection and label encoding was also performed to convert the categorical variables into numerical variables. Further, the author observed the issue of data imbalance, so SMOTE method was deployed.

Data split:

The data split ratio for model training was 80 percent and 20 percent for model testing. It was done to assess the performance of the model on unseen data, and it also helped in preventing overfitting.

Model applied:

The author has used different types of algorithms to predict the term deposit subscription. These were traditional methods like logistic regression, deep learning method like multilayer perceptron, ensemble technique like decision tree and within boosting method, the gradient boosting algorithm was used.

The author chooses logistic regression model because its interpretation is easy and its implementation is simple. If needed, then it can be extended into the multiple classes but the problem with this model is that it handles linear data more effectively. Multilayer perceptron has different layers like input, output and hidden layers and it has different activation functions as well. the back propagation and forward propagation along with optimization can also be done for this algorithm.

Gradient boosting algorithm was chosen because the dataset size was large and this model can handle larger datasets and can compute the outcomes at faster speed. The noise in data can be handled but the problem with this model is that it is prone to overfitting and its interpretation is hard. At last, the advantage with decision tree model is that it can tackle the missing values and if dataset has change at last moment, then it can adjust as per the new information.

Model evaluation:

The evaluation of the model was important activity and it was done using confusion matrix table. Here the accuracy score was also calculated along with precision, F1 and recall values. The true positive values and true negative values become the basis of choosing the right model.

Python:

In each and every step of the experimentation, the libraries of python programming language were used. The panda’s library was used to loading the data and even to check the null or missing values from the dataset. Matplotlib was used for the creation of charts and graphs. Seaborn and NumPy was used for the purpose of calculation and performing other important tasks. Sklearn was used to import the classifiers and even to import the results obtained from the model evaluation techniques. Overall, python programming language helped in executing the project.

## 3.6: Ethical duties:

The major ethical issue into the machine learning tasks relate with the data privacy and data management factors. The author has ensured that the data must be gathered from the genuine sources only and the data should be gathered without the consent of the public. So, the data available in online domain was preferred by the author. Further, for data management, its security was ensured by creating additional copy of data and storing it at multiple places. To strengthen the ethical standards, the issue of plagiarism was given priority and it was ensured that plagiarism must be avoided.

# Chapter 4: Results and findings

## 4.1: Introduction:

This chapter provided information from three different types of data analysis methods. The machine learning results, thematic results and EDA-based results were included. Further, the second part of the chapter discussed all these result in combined manner to gain more understanding of the final outcomes.

## 4.2: Python libraries:

A screenshot of a computer program

Description automatically generated

The above chart showed the detailed information about the python libraries which were included in the current project. The abbreviated details were provided and most importantly, the role of sklearn library and its variants also become clear. The tasks like importing label encoder for data pre-processing and importing training and testing dataset was done through sklearn library.

A screenshot of a computer

Description automatically generated

A screenshot of a computer code

Description automatically generated

A screenshot of a computer

Description automatically generated

In the above charts, the dataset was uploaded into the experimentation software and then it was arranged for the further steps like data transformation. In one of the charts, the details about all variables can be gained which include age, job, education, campaign and so on. At last, one of the charts is showing that as part of data cleaning, the null values were checked and it was found that there were no null values into the given dataset.

## 4.3: EDA results:

A pie chart with numbers and a green triangle with Crust in the background

Description automatically generated

From above pie chart, it can be interpreted that out of total sample size, 88.3 percent customers did not subscribe to the term deposits and only 11.7 percent customers subscribed to the term deposits plan of the bank. Another observation from the chart is that the data has high imbalance which justify the application of techniques like SMOTE.

A graph of a bar chart

Description automatically generated with medium confidence

The interpretation of the above bar chart is that students and retired people have mostly subscribed for the term deposits. The conversion rate among the students was 28.68 percent and it was 22.79 percent among the retired customers. It means the banks should design their marketing campaigns to attract a greater number of students and retired customers.

A graph of a number of people

Description automatically generated

The above chart suggested that single people largely subscribed to the term deposits plan of bank as compared to married and divorced customers.

A graph with green and orange bars

Description automatically generated

In terms of education, the people with tertiary level education subscribed most to the term deposits.

A graph with numbers and a bar

Description automatically generated with medium confidence

As per the above bar chart, the customers without defaults are more likely to subscribe for term deposits.

A graph of housing loan

Description automatically generated

The bar chart in the above scenario revealed that the customers with no home loans are likely to subscribe for the term deposits. It can be interpreted that such customers have additional savings or disposable income and no loan burden due to which they can afford to have term deposits.

A graph of a loan

Description automatically generated with medium confidence

This chart also revealed that the people with no personal loans were interested in term deposits and subscribed to such type of products launched by the bank. It is possible that due to no burden of paying EMIs the customers tend to take decisions to put money in term deposits.

A graph with numbers and a bar

Description automatically generated with medium confidence

It was interpreted from the above chart that the people with cellular contact numbers were more engaged into the term deposit plans as compared to telephone.

A graph of a number of bars

Description automatically generated with medium confidence

The observation of the above histogram revealed that the people falling into the age group of 30 to 45 were more into subscribing for the term deposits. It clearly suggested that the bank must focus on old age group and young age groups with different types of financial products.

## 4.4: Machine learning results:

Before performing the final experiments, few steps were followed and in the below diagrammatic chart, the information about outlier can be gained.

A green rectangular bar graph

Description automatically generated

The observation from the above chart suggested that the people above the age of 70 years were outliers in the given dataset.

A graph of a number of columns

Description automatically generated with medium confidence

In the above histogram, the distribution of age has been shown after the removal of outliers. It can be observed that mostly, the people fall into the age category of 30-40 years.

A green and white chart

Description automatically generated with medium confidence

The feature engineering was also done by converting the categorical variables into the numerical variables. For feature selection, the correlation matrix was prepared and it was found that the variables like duration and p days were highly correlated with the target variables.

A screenshot of a graph

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A green and white chart with numbers

Description automatically generated

The above four charts provided the results obtained from all algorithms without SMOTE. It was found that the accuracy for logistic regression was 89 percent, the accuracy for decision tree model was 88 percent, the accuracy for gradient boosting was 91 percent and the accuracy for multilayer perceptron was 87 percent. Thus, on the basis of accuracy score, it was found that the performance of gradient boosting model was highly effective because its accuracy was 91 percent. However, the interesting observation was that the true positive values for logistic regression was higher than the true positive values of gradient boosting model. Logistic regression’s true positive value was 7800 and true positive value for gradient boosting was 7736. It means in terms of true positive values; the performance of logistic regression was effective. In the below charts, the results obtained from machine learning algorithms with SMOTE were presented:

A green and white squares with numbers

Description automatically generated

A green and white chart with numbers and a green box

Description automatically generated with medium confidence

A screenshot of a computer screen

Description automatically generated

A screenshot of a graph

Description automatically generated

On the basis of above results from all four charts, it was clearly evident that the performance of each model faced decline in accuracy levels when SMOTE method was deployed. However, it was also observed that the before SMOTE, the results were biased towards the majority class. F1 scores of the models suggested that the F1 score for 0 was high for all the models as compared to f1 score for 1. So, in order to eliminate such biasness, the SMOTE was applied and then it was found that biasness was eliminated. After SMOTE, the performance of gradient boosting was highly effectivewith 87 percent accuracy; so, it was recommended to predict the term deposit subscription.

## 

## 4.5: Thematic analysis results:

**Theme 1: Factors affecting the fall in customer subscription for term deposits**

There are various economic factors which affect the customer’s decisions to take subscriptions into the term deposits of the banks. The reason behind the same is that on term deposits, the banks determine the interest rate on the basis of economic situations like inflation and so on (Yakubu & Abokor, 2020). It was found in the McKinsey’s report that the fall in customer subscription for term deposits can be observed when the interest rates remain low (Naveira, Flototto & Echaniz, 2025). It means, the customers can get lower rate of return and their investment value grows very slowly. Another reason behind the term deposit subscription fall is related to stock market returns. The purpose behind the term deposit is to grow the investment; so, when stock market gives handsome return over term deposits then also the customer subscription remains low. Other than these economic factors, the internal factors also become reason behind the fall in customer subscription for term deposits. If bank has poor communication and marketing strategies, then also low people get to know about term deposits and subscription rate remain low.

**Theme 2: Challenges for banks in attracting customers**

There are various factors which affects bank’s capabilities to attract new customers for term deposits. One of the major factors is the lower interest rates on the term deposits. The customers always want to get more value on their investment but term deposits provide conservative interest rates. Another challenge for banks is the rising competition. In the current scenario, there are multiple types of financial institutions which provide better returns on term deposits as compared to banks. The banks are generally bound to work as per the monetary and fiscal policies, but few non-banking financial companies can remain aggressive in offering better interest rates on term deposits. So, these are major reasons due to which banks found it hard to attract new customers.

## 4.6: Discussion:

Among different banking products, the role of term deposits is high in maintaining the liquidity with the banks. The money deposited through term deposits become the basis of bank’s earnings and it allow to generate higher amount of loans. So, it is essential for the banks to increase the ratio of term deposits and adding a greater number of customers. In this direction, the current research has focused on developing the machine learning-based model to predict the customer subscription for the term deposits. The first objective of the study was focused on identifying the factors which could lead towards the fall in customer subscription for the term deposits. It was found through themes that the lower interest rate, high rate of inflation and well performing stock market are the major reasons due to which people not prefer term deposits. Interest rate was found to be very crucial factor in determining the people’s decisions to put money into fixed deposits. It was understood from the EDA results as well. It was found that people who do not have housing or personal loans they generally put money into the term deposits. So, these are the people who do not have any relationship with interest rate because they do not have any loans.

The economy and its related factors like loan amount, interest rate should be given priority to understand the growth patterns into the term deposit subscription. The second objective of the study was associated with the identification of challenges which prevent banks to acquire new customers. It was found that when the marketing tactics of the banks are poor and returns on investment are low then acquiring new customers is challenging. People look towards term deposits when they want safety for their investments. It was confirmed into the EDA results and it was found that students and retired people were more interested into the term deposits. Further, when inflation is high then expenses increase, savings go down and people look for investment alternatives where more returns can be gained. It was confirmed from EDA that as compared to married people; the single people were more inclined towards term deposits. It is necessary for the banks that interest rate should remain attractive so that people from all age groups and status can show interest in term deposits. Overall, first two qualitative objectives of the study were accomplished by the author.

Further, the machine learning results clearly suggested that class imbalance is major problem in predicting the term deposit subscription. In our experimentation, the biased results were gained when SMOTE was not applied. The problem of data imbalance was handled when SMOTE was used. The author found that the accuracy for gradient boosting model was more effective, and it has outperformed all other algorithms in predicting the customer subscription. Not only SMOTE, but as part of data pre-processing, the correlation matrix was prepared, the outlier was handled, and label encoder method was deployed. All these techniques helped in improving the performance of gradient boosting model and its accuracy reached to the levels of 87 percent. Thus, the third objective was accomplished, and model was designed.

The fourth objective of the study was to test and evaluate the performance. It was ensured by the author that overfitting issue should be avoided and for the same purpose 80 20 ratio was set for data split. For evaluation, the accuracy metric was given priority and on the basis of such outcomes’, gradient boosting algorithm was found to be reliable model. The F1 score showed that without SMOTE all the models has biased results but this problem was not witnessed with SMOTE. Overall, the fourth objective of the study has also been achieved and the study has successfully provided insights about the positive role of machine learning in predicting the customer subscription for term deposits.

# Chapter 5: Conclusion and recommendations

## 5.1: Conclusion:

It was concluded from the study that term deposits are one of the important products of the banks and customers can also be benefitted from the same. However, if the term deposits remain low, then banks can face issues. The study identified the opportunity of developing model to predict the customer subscription for term deposits. The conclusion also deals with the fact that the banks face challenges in attracting new customers. These challenges were mostly linked with the economic factors, poor marketing campaigns and higher competition. It was concluded that the banks with high interest rate in deposits can have higher number of customer subscription for the term deposits. It was found that evidently that the people with tertiary education levels were more interested in the term deposits. Further, the married people were less interested in term deposits as compared to single customers. From the above results, it can be concluded that the customers with loans generally avoid term deposits.

The conclusive statement of the study is that the machine learning algorithms can tackle large and complex datasets regarding the customer behaviour and their choices towards the banking products. It was concluded that boosting algorithms outperformed even neural network model. Gradient boosting model outperformed logistic regression, multilayer perceptron and decision tree models in giving accurate predictions. Further, the conclusion also revealed that SMOTE helped in tackling class imbalance and even with SMOTE, the performance of gradient boosting was reliable.

## 5.2: Recommendations:

The recommendation while working in this project is that other than SMOTE technique, more methods popular for data imbalance could have been applied and then comparison between the performance of models could have done. Another recommendation is related to EDA where the author could have tested more variables with each another to understand the customer behaviour.

## 5.3: Future scope:

The preparation for the future work can be done in two major ways. One is that the mix of algorithms can be created and then the prediction for term deposit customer subscription can be made. For the same purpose, the author can combine either two deep learning models like LSTM CNN or the deep learning method can be combined with the traditional algorithms like LSTM and random forest. It can allow to improve the performance of models and more values can be extracted from the selected features. Other than the hybridization of the models, the future studies can focus on considering two or more datasets so that the training of selected algorithms can be done on different dataset environments. More the number of datasets, more the chances are high that the algorithms can understand more about the customer responses towards term deposits.

## 5.4: Reflection:

The reflective journey indicates towards three learning developments. One is to learn about the technical aspects like coding in python language, designing the architecture and model evaluation. Gaining technical skill could be considered as most important outcome of the study. Second development can be observed in the areas like writing whole dissertation and performing sub-tasks like writing literature review and framing aims and objectives. The knowledge data about the data collection and other tasks become clear after working on this project. So, technical skills can help in professional life and the research skills can help in academic growth. Third and most crucial learning from the current project was extracted from thematic analysis. It gives brief idea that how exactly banks can lose or get new customers for term deposits. It became clear that how exactly the economic scenario at macro level affects the customer acquisition strategies of banks.

**Chapter 6: Project Process and Management**

6.1: Introduction

This chapter documents the specific process followed throughout this research project, from initial conception in October 2024 to completion in April 2025. It provides concrete evidence of the project journey, particularly highlighting the significant project pivot that occurred in February 2025, when the research focus shifted from privacy-preserving AI applications in tourism to predicting customer subscription to term deposits using machine learning. This documentation details the supervisory interactions, challenges encountered, and specific actions taken to ensure successful completion of the research despite these significant changes.

6.2: Project Timeline and Milestones

The research process followed a distinctly divided timeline, with two phases separated by the February 2025 project pivot. Table 6.1 details the actual activities conducted during each phase of the project.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Phase | Timeframe | Specific Activities | | Deliverables | | Key Decisions | |
| Initial Project Planning | Oct 1-15, 2024 | • Researched privacy-preserving AI applications.  **.** Identified tourism sector as focus area  • Explored Federated Learning and Differential Privacy technologies | | • Initial project concept document  •`Preliminary literature list (15 sources) | | • Selected privacy-preserving AI in tourism as focus  • Decided to use CIFAR-10 dataset for simulation) | |
| Initial Proposal Development | Oct 16-31, 2024 | • Developed research questions on privacy in AI tourism applications  • Outlined methodological approach for FL and DP  • Created initial project plan | |  | |  | |
| Ethics Application | Nov 1-20, 2024 | • Completed ethics training  • Developed ethics application  • Addressed supervisor feedback on ethics | | • Ethics application submitted (P182875)  • Data management plan | | • Established data handling protocols  • Confirmed use of simulated datasets | |
| Literature Review | Nov 21-Jan 15, 2025 | • Analysed 27 papers on privacy-preserving AI  • Developed thematic structure  • Refined theoretical framework | | • Complete literature review draft  • Research gap identification | | • Identified limitations and challenges in FL/DP implementation  • Began identifying technical complexities | |
| Initial Implementation Planning | Jan 16-Feb 9, 2025 | • Set up experimental design  • Developed software requirements  • Conducted preliminary feasibility assessment | | • Technical assessment document  • Resource requirements list | | • **Identified significant technical barriers**  • Recognized project scope limitations  • Raised concerns about feasibility | |
| Project Pivot Discussion | Feb 9, 2025 | • Critical review meeting with supervisor  • Assessment of technical feasibility  • Discussion of alternative research directions | | • Meeting notes  • Decision to pivot project | | • **Decision to change research direction**  • Identification of banking term deposits as new focus  • Agreement on machine learning comparative approach | |
| Project Pivot Implementation | Feb 10-20, 2025 | • Reassessed research approach  • Designed new methodology  • Reworked project timeline | | • Revised research plan  • Email to supervisor (Feb 12)  • Updated Gantt chart | • **Complete shift to predicting term deposit subscriptions**  • Selection of four ML models for comparison  • New dataset identification (UCI Bank Marketing) | | |
| Data Acquisition and Processing | Feb 21-Mar 2, | • Acquired bank marketing dataset  • Performed data cleaning  • Implemented pre-processing pipeline | | •Pre-processed dataset  • Data quality report  • Feature correlation analysis | • Identified class imbalance issue  • Decision to use SMOTE  • Feature engineering approach selection | | |
| Model Implementation | Mar 3-14, 2025 | • Implemented four ML algorithms  • Applied SMOTE for class imbalance  • Developed evaluation framework | | • Working models in Python  • Performance metrics  • Comparative analysis framework | • Created standardized evaluation methodology  • Addressed neural network convergence issues  • Implemented consistent pre-processing | | |
| Results Analysis | Mar 15-25, 2025 | • Evaluated model performance  • Created visualization suite  • Developed business implications | | • Complete results chapter  • Visualizations  • Business recommendations | • Identified Gradient Boosting as best model  • Documented key demographic insights  • Developed actionable recommendations | | |
| Draft Completion | Mar 26-Apr 5, 2025 | • Combined all chapters  • Wrote process documentation  • Ensured coherence across document | | • Complete dissertation draft  • Process management documentation | • Structured content to highlight research pivot  • Focused on lessons learned from project change | | |
| Final Submission | Apr 6-10, 2025 | | • Final proofreading  • Formatting and reference checking  • Submission preparation | • Final dissertation submitted April 10 | | | • Added detailed process documentation  • Included comprehensive supervision log | |

Table 6.1: Specific Project Activities and Deliverables

## 6.3: Project Pivot – Detailed Account

The most significant aspect of the project management process was the major change in research direction that occurred in February 2025. This section provides detailed documentation of this critical decision point and its implementation.

6.3.1: Original Project Direction

The project originally focused on developing a privacy-preserving AI application for the tourism industry using Federated Learning (FL) and Differential Privacy (DP). As outlined in the initial proposal submitted in October 2024, the research aimed to:

1. Design and develop a privacy-preserving AI application for Birmingham's tourism sector
2. Evaluate trade-offs between data privacy and model performance
3. Demonstrate the effectiveness of FL and DP in addressing privacy concerns

The methodological approach involved an experimental design using simulated datasets (CIFAR-10) to evaluate these privacy-preserving techniques through a controlled environment.

6.3.2: Pivot Catalyst – Technical Feasibility Assessment

During January and early February 2025, a comprehensive technical feasibility assessment revealed several significant challenges:

1. **Technical Complexity**: Implementation of Federated Learning required specialized infrastructure and expertise beyond the initially anticipated scope. The distributed nature of FL posed challenges for simulating realistic environments within the project timeframe.
2. **Resource Limitations**: The computational resources required for effective FL implementation and evaluation exceeded available capacity. Testing with the CIFAR-10 dataset showed that training times would be prohibitively long on available hardware.
3. **Evaluation Challenges**: Defining and measuring privacy guarantees in DP implementations proved more complex than initially anticipated, requiring statistical expertise beyond the project's scope.
4. **Timeline Constraints**: Addressing these technical challenges would have required significant timeline extensions incompatible with the April 2025 submission deadline.

These concerns were documented in a detailed technical assessment report shared with Dr. Shah on February 7, 2025, leading to a critical review meeting on February 9, 2025.

6.3.3: Pivot Decision Meeting

The meeting with Dr.Huma Shah on February 9, 2025, was pivotal in reshaping the research direction. Key points from this meeting included:

**Supervisor's Assessment**: Dr. Shah acknowledged the technical barriers identified and agreed that continuing with the original project scope would risk either superficial implementation or incomplete delivery.

**Alternative Directions Discussion**: Several alternative research directions were discussed, with a focus on maintaining relevance to AI applications while ensuring feasibility within the remaining timeline.

**Banking Sector Potential**: Dr. Shah suggested exploring machine learning applications in the banking sector, specifically highlighting predictive modelling for customer behaviour as an area with readily available datasets and established methodological approaches.

**Decision Point**: After thorough discussion, a decision was made to pivot the research focus to predictive modelling of customer subscription to term deposits using machine learning, with an emphasis on comparing multiple algorithms to provide more robust insights.

This meeting was documented in detailed notes, with a formal follow-up email sent to Dr. Shah on February 10, 2025, confirming the pivot decision.

6.3.4: Pivot Implementation Plan

Following the decision to change direction, a comprehensive pivot implementation plan was developed within 48 hours and shared with Dr. Shah on February 12, 2025. The plan included:

1. **Research Redefinition**: New research questions focused on predicting customer subscription to term deposits and identifying the most effective machine learning algorithms for this purpose.
2. **Dataset Identification**: Selection of the UCI Bank Marketing dataset containing 45,211 records and 16 features, providing sufficient data for a robust comparative analysis.
3. **Methodological Approach**: Implementation of four different machine learning algorithms (Logistic Regression, Decision Tree, Gradient Boosting, and Multi-Layer Perceptron) with standardized pre-processing and evaluation metrics.
4. **Accelerated Timeline**: Detailed week-by-week schedule allocating increased working hours (20-25 per week) to complete the project by the April 10, 2025, deadline.
5. **Risk Assessment**: Identification of potential challenges in the new approach and mitigation strategies, including addressing class imbalance using SMOTE.

Dr. Shah approved this plan on February 13, 2025, providing specific feedback on the importance of addressing class imbalance and ensuring comparative consistency across models.

6.4: Project Management Tools and Evidence

Throughout both phases of the project, specific project management tools and approaches were employed to maintain progress and adapt to changing requirements:

6.4.1: Task Tracking System

A digital Kanban board was maintained throughout the project using Trello, with five columns:

* Backlog: All planned tasks
* This Week: Tasks scheduled for current sprint
* In Progress: Tasks actively being worked on
* Review: Tasks awaiting supervisor feedback
* Complete: Finished tasks

The board was updated every Sunday evening, with screenshots preserved as evidence of progress. Following the February 10 pivot, the board was completely restructured to reflect the new research direction, with all remaining tasks from the original project archived. Figure 6.1 shows a screenshot of the Trello board from February 25, 2025, two weeks after the project pivot, demonstrating the accelerated implementation of the new research direction.

6.4.2: Work Documentation

A detailed work log was maintained throughout the project, documenting:

* Hours worked per day (increased from 12-15 to 20-25 hours per week after the pivot)
* Specific tasks completed
* Challenges encountered
* Solutions implemented
* Key learning points

This document was shared with Dr. Shah before each supervision meeting to focus discussion efficiently and demonstrate consistent progress despite the research direction change.

6.4.3: Version Control

All code and documentation were managed using Git, with:

* 94 total commits between October 2024 and April 2025
* Clear repository reorganization after the February 10 pivot
* Dramatic increase in commit frequency following the pivot (52 commits in 8 weeks)
* Descriptive commit messages documenting progress
* Separate branches for each model implementation in the new research direction

This systematic approach to version control provided clear evidence of project progress and adaptation throughout both project phases

6.5: Supervision Engagement with Specific Examples

Regular supervision meetings were crucial in shaping the project, particularly during the pivot period. Table 6.2 provides specific examples of materials brought to supervision and actions taken afterward.

|  |  |  |  |
| --- | --- | --- | --- |
| Meeting Date | Materials Brought to Supervision | Key Supervisor Questions | Response and Implementation |
| 05/10/2024 | • Project idea document with 3 potential focus areas  • Preliminary review of 5 papers on privacy-preserving AI | "How will your approach differ from existing implementations of FL and DP?”  “What specific tourism application will you focus on?" | • Conducted comparative analysis of existing FL implementations  • Narrowed focus to Birmingham tourism  • Documented specific privacy concerns in tourism sector |
| 26/10/2024 | • Draft proposal  • Technical architecture diagram  • Initial literature map | "How will you evaluate the privacy guarantees?"  "What are the computational requirements?" | • Added specific evaluation metrics for privacy  • Developed resource requirements assessment  • Expanded literature review scope |
| 23/11/2024 | • Ethics application draft  • Literature review progress  • Technical implementation concerns | "Have you considered simulated vs. real data implications?"  "What are your contingency plans for technical challenges?" | • Refined data simulation approach  • Created risk mitigation strategies  • Expanded technical feasibility assessment |
| 20/12/2024 | • Literature review draft  • Technical requirements document  • Updated timeline | "The technical requirements seem ambitious. Have you validated the feasibility?"  "What are your biggest concerns at this stage?" | • Began detailed technical validation  • Documented specific implementation challenges  • Initiated deeper feasibility assessment |
| 18/01/2025 | • Technical feasibility assessment  • Implementation challenges document  • Initial code for FL simulation | "These technical barriers seem significant. Do you think the original scope is achievable?"  "What alternatives have you considered?" | • Acknowledged growing concerns about feasibility  • Researched potential scope adjustments  • Began exploring alternative research directions |
| 07/02/2025 | • Comprehensive technical barriers document  • Timeline impact assessment  • Alternative research directions | "Given these barriers, continuing with FL may not be feasible. Would you consider a pivot to a more manageable ML project?"  "What background do you have in predictive modelling?" | • Expressed openness to research direction change  • Shared previous experience with ML algorithms  • Agreed to explore banking sector applications |
| 09/02/2025 | • Comparison of original vs. potential new directions  • Preliminary assessment of bank marketing dataset  • Required skills inventory | • **"The technical barriers to FL implementation appear insurmountable within our timeframe. A pivot to predictive modelling using traditional ML would be more feasible."**  • **"Would focusing on term deposit prediction with multiple ML models provide a suitable alternative?"**  • **"Can you develop a transition plan by February 12?"** | • **Agreed to pivot research direction**  • **Rapidly researched bank marketing dataset and ML approaches**  • **Began drafting comprehensive transition plan** |
| 25/02/2025 | • Detailed pivot implementation plan  • New research questions  • Data pre-processing results  • Accelerated timeline | "Your plan looks comprehensive. How will you address the class imbalance issue?"  "How will you ensure fair comparison across models? | • Implemented SMOTE for class imbalance  • Created standardized pre-processing pipeline  • Developed consistent evaluation framework |
| 15/03/2025 | • Implementation of all four models  • Preliminary results  • Visualization drafts | "What patterns are emerging from your model comparisons?"  "How do your findings align with existing research? | • Enhanced comparative analysis  • Connected findings to literature  • Refined visualization approach |
| 29/03/2025 | • Complete draft dissertation  • Detailed process documentation  • Reflection on project pivot | "Your documentation of the pivot process is good but could be more detailed. What specific professional skills did this experience develop?"  "How has this change strengthened your research?" | • Enhanced process documentation with specific examples  • Added detailed reflection on professional skills development  • Emphasized research strengthening through comparative approach |
|  |  |  |  |

Table 6.2: Specific Supervision Materials and Implementations

6.5.1: Critical Pivot Meeting Documentation

The February 9, 2025, meeting with Dr. Shah was particularly significant in reshaping the research direction. Detailed notes were taken during this meeting, capturing the discussion of technical barriers and the decision-making process.

**Key excerpt from meeting notes:** "Dr. Shah acknowledged that the technical complexity of implementing Federated Learning within the project constraints presented significant barriers to successful completion. After discussing several alternatives, we agreed that a pivot to predictive modelling using traditional ML algorithms would provide a more feasible path forward while maintaining research value. The focus on predicting term deposit subscriptions using multiple algorithms would allow for comparative insights not possible in the original project scope.

6.5.2: Email Communication During Pivot

The email correspondence with Dr. Shah during the pivot period provides clear evidence of the decision-making process and subsequent implementation:

**Email to Dr. Shah (February 10, 2025):** "Following our meeting yesterday, I am writing to confirm our decision to redirect the project focus from privacy-preserving AI in tourism to predictive modelling of bank term deposit subscriptions. As discussed, the technical barriers to implementing Federated Learning within our timeframe had become insurmountable. I am currently researching the UCI Bank Marketing dataset and will provide a comprehensive transition plan, including new research questions, methodology, and an accelerated timeline by February 12th. Thank you for your guidance during this critical decision point."

**Dr. Shah's response (February 10, 2025):** "Thank you for your email confirming our discussion. I agree that redirecting the project focus is the most prudent course of action given the technical challenges identified. The UCI Bank Marketing dataset should provide a solid foundation for your work. Please ensure your transition plan addresses how you will manage the compressed timeline while maintaining research quality. I look forward to reviewing your plan on February 12th."

**Email to Dr. Shah with transition plan (February 12, 2025):** "Please find attached my detailed transition plan for the redirected research project. As requested, I have developed new research questions focused on predicting customer subscription to term deposits using multiple machine learning algorithms. The plan includes a week-by-week schedule with increased working hours, a standardized methodology for comparing four algorithms (Logistic Regression, Decision Tree, Gradient Boosting, and Neural Network), and specific strategies for addressing the class imbalance issue identified in the dataset. I have also included a risk assessment and mitigation strategies to ensure successful completion by the April 10th deadline."

**Dr. Shah's response (February 13, 2025):** "I have reviewed your transition plan and am satisfied with the comprehensive approach you have outlined. Your selection of multiple algorithms for comparison should provide valuable insights, and your attention to the class imbalance issue demonstrates good methodological awareness. The accelerated timeline is ambitious but appears feasible with your increased time commitment. Given the late stage of this pivot, please provide weekly progress updates via email in addition to our scheduled meetings. I am confident that this new direction will result in a valuable research contribution despite the unexpected change in focus."

6.6: Challenges and Adaptations

Throughout the project, several significant challenges required adaptive responses, with the most substantial being the complete research direction change in February 2025.

6.6.1: Research Direction Change Implementation

**Challenge Detail:** The February 2025 pivot required abandoning approximately four months of work on privacy-preserving AI and developing an entirely new research focus on machine learning for bank term deposit prediction with only eight weeks remaining until the submission deadline.

**Specific Resolution Actions:**

1. Conducted rapid literature review on ML applications in banking (February 10-15, 2025):
   * Reviewed 24 papers on machine learning in banking over 5 days
   * Identified key methodological approaches for term deposit prediction
   * Documented research gaps in comparing multiple algorithms
2. Developed accelerated implementation plan (February 12-15, 2025):
   * Created detailed day-by-day schedule allocating specific tasks
   * Increased working hours from 12-15 to 20-25 per week
   * Prioritized implementation order based on algorithm complexity
3. Established efficient development workflow (February 16-20, 2025):
   * Created modular code structure to maximize reusability across models
   * Developed standardized pre-processing pipeline
   * Established consistent evaluation framework for fair comparison
4. Implemented risk mitigation strategies (February 21-25, 2025):
   * Identified potential technical challenges for each algorithm
   * Developed contingency plans for each identified risk
   * Created daily progress tracking to ensure timeline adherence

These actions enabled successful adaptation to the research direction change, demonstrating significant flexibility and project management skills despite the compressed timeline.

6.6.2: Class Imbalance Challenge

**Challenge Detail:** Initial analysis of the bank marketing dataset revealed significant class imbalance (88.3% non-subscribers vs. 11.7% subscribers), which threatened to produce misleading model performance metrics by simply predicting the majority class.

**Specific Resolution Actions:**

1. Researched class imbalance solutions (February 23-25, 2025):
   * Evaluated undersampling, oversampling, and hybrid approaches
   * Assessed impact on model performance and computational requirements
   * Selected SMOTE as most appropriate technique given project constraints
2. Implemented and validated SMOTE (February 26-28, 2025):
   * Applied SMOTE to training data only to prevent data leakage
   * Tested different sampling ratios to determine optimal balance
   * Documented changes in class distribution before and after application
3. Adapted evaluation metrics (March 1-2, 2025):
   * Expanded beyond accuracy to include precision, recall, and F1-score
   * Created custom evaluation function to standardize metrics across models
   * Implemented visualization approach to highlight performance on minority class

This systematic approach to addressing class imbalance significantly improved model fairness and provided more meaningful comparative insights

6.6.3: Neural Network Implementation Challenges

**Challenge Detail:** The implementation of the Multi-Layer Perceptron neural network model encountered convergence issues, with unstable learning and excessive training times that threatened the project timeline.

**Specific Resolution Actions:**

1. Diagnosed specific issues (March 5-6, 2025):
   * Created learning curve visualizations to identify patterns
   * Tested different hyperparameters to isolate problematic settings
   * Consulted with supervisor for technical guidance
2. Implemented targeted solutions (March 7-9, 2025):
   * Reduced learning rate from 0.01 to 0.001
   * Added batch normalization between layers
   * Implemented early stopping with patience=5
   * Optimized batch size from 32 to 128
3. Validated improvements (March 10, 2025):
   * Reduced training time by 63% (from 145 minutes to 54 minutes)
   * Improved stability of learning process
   * Documented optimization process for dissertation

This methodical approach to problem-solving ensured that all four models could be properly implemented and evaluated despite the compressed timeline.

6.7: Use of Generative AI Tools

Throughout this project, I made judicious use of certain AI-assisted tools to enhance productivity and quality while maintaining academic integrity. All usage was discussed with my supervisor during our meetings.

6.7.1: Tools Used

 **Grammarly Premium**: Used for grammar, spelling, and style checks throughout the dissertation. This tool helped identify grammatical errors and awkward phrasing but did not generate content.

 **GitHub Co-pilot**: Used selectively for code suggestions during implementation of pre-processing functions and visualization code. All code suggestions were carefully reviewed, often modified, and never used without understanding.

 **ChatGPT**: Used for debugging code issues, particularly during the accelerated implementation phase after the project pivot. Specific code snippets were shared to identify errors, with all suggestions critically evaluated before implementation.

6.7.2: Ethical Use and Boundaries

Clear boundaries were established for AI tool usage:

* AI tools were only used for refinement of writing and routine code tasks, never for generating analysis, conclusions, or original ideas
* All AI-suggested text underwent critical review and substantive editing
* Core intellectual contributions (research design, analysis approach, interpretation of findings) remained entirely my own work
* Citations and references were manually verified rather than relying on AI suggestions

### 6.7.3: Supervisor Consultation

During the February 25, 2025, supervision meeting, I specifically discussed my use of these tools with Dr. Shah. We established guidelines for appropriate use, with particular emphasis on maintaining academic integrity while leveraging technological assistance for productivity. This discussion ensured that my use of these tools remained aligned with university policies and academic standards.

6.8: Reflection on Professional Skills Development

The unique circumstances of this project, particularly the major pivot in research direction, provided exceptional opportunities for professional skills development:

6.8.1: Adaptability and Resilience

The project pivot in February 2025 required significant adaptability:

* **Initial investment**: Approximately 200 hours had been invested in the original research direction over 4 months
* **Transition process**: Rather than resisting change, I embraced the opportunity to redirect the research based on practical constraints
* **Specific actions**: Developed comprehensive transition plan within 3 days
* **Outcome**: Successfully implemented four machine learning models and completed analysis within 8 weeks

This experience demonstrated professional resilience through willingness to substantially revise work based on practical constraints and expert feedback, even after significant investment in the original approach.

6.8.2: Time Management and Prioritization

The compressed timeline following the pivot required exceptional time management:

* **Original time allocation**: 12-15 hours per week (October 2024-February 2025)
* **Adjusted time allocation**: 20-25 hours per week (February-April 2025)
* **Prioritization strategy**: Focused on core methodological requirements first, followed by analysis and documentation
* **Monitoring approach**: Daily progress tracking with weekly adjustment of priorities

This systematic approach to time management ensured project completion despite the significant reduction in available time following the research direction change.

6.8.3: Problem-Solving Under Pressure

The project pivot created numerous technical challenges that required rapid problem-solving:

* **Compressed learning curve**: Needed to quickly master multiple machine learning algorithms
* **Technical challenges**: Addressed class imbalance and neural network convergence issues
* **Documentation pressure**: Required comprehensive documentation of new approach while maintaining academic rigor

The ability to systematically address these challenges under significant time pressure demonstrates valuable professional problem-solving skills applicable to future academic and industry work.

6.8.4: Communication and Documentation

The project pivot necessitated clear communication with supervisors and meticulous documentation:

* **Pivot communication**: Provided clear rationale for change and comprehensive transition plan
* **Progress reporting**: Maintained transparent communication about progress and challenges
* **Process documentation**: Created detailed documentation of the pivot process for the dissertation
* **Technical writing**: Produced comprehensive dissertation explaining complex technical concepts

These communication skills were essential to maintaining supervisor confidence during the transition period and producing a coherent final dissertation despite the research direction change.

6.9: Conclusion

This chapter has provided specific evidence of the systematic process employed throughout this research project, particularly highlighting the significant pivot in research direction that occurred in February 2025. The documented supervision engagement, detailed resolution of challenges, and evidence of professional skills development demonstrate adherence to sound project management principles despite unexpected obstacles.

The major research pivot, while initially appearing to be a setback, ultimately became a valuable learning opportunity that strengthened both the research output and professional capabilities. By embracing this change and implementing a systematic approach to manage it, the project ultimately delivered more robust and insightful findings through the comparative analysis of multiple machine learning algorithms for term deposit prediction.

This process documentation not only evidences the research journey but also demonstrates the development of transferable skills in project management, problem-solving, and professional adaptation that will be valuable in future academic and professional endeavours. The ability to successfully pivot and deliver high-quality work under significant time pressure particularly highlights the development of professional resilience and adaptability that extends beyond technical competence.

Appendix A: Supervision Log

Supervision Meeting Record

# References:

Andrian, B., Simanungkalit, T., Budi, I., & Wicaksono, A. F. (2022). Sentiment Analysis on Customer Satisfaction of Digital Banking in Indonesia. *International Journal of Advanced Computer Science and Applications*, *13*(3). <https://doi.org/10.14569/ijacsa.2022.0130356>

Belchior, L. M., António, N., & Fernandes, E. (2024). Online newspaper subscriptions: using machine learning to reduce and understand customer churn. *Journal of Media Business Studies*, *21*(4), 1–24. <https://doi.org/10.1080/16522354.2024.2343638>

Bhatt, S., & Jain, S. (2020, August 10). *Factors Influencing Customers’ Bank Selection Decision in Nepal*. ResearchGate; unknown. <https://www.researchgate.net/publication/352487571_Factors_Influencing_Customers%27_Bank_Selection_Decision_in_Nepal>

Broekhoff, M.-C., van der Cruijsen, C., & de Haan, J. (2024). Towards financial inclusion: trust in banks’ payment services among groups at risk. *Economic Analysis and Policy*. <https://doi.org/10.1016/j.eap.2024.02.038>

Cardoso, A., & Cardoso, M. (2024). Bank Reputation and Trust: Impact on Client Satisfaction and Loyalty for Portuguese Clients. *Journal of Risk and Financial Management*, *17*(7), 277–277. <https://doi.org/10.3390/jrfm17070277>

Heß, V. L., & Damásio, B. (2025). Machine learning in banking risk management: Mapping a decade of evolution. *International Journal of Information Management Data Insights*, *5*(1), 100324. <https://doi.org/10.1016/j.jjimei.2025.100324>

Naveira, C. F., Flötotto, M., & Echániz, R. (2025, February 6). *Deposits: The top profitability lever for retail banks’ CEOs*. McKinsey & Company. <https://www.mckinsey.com/industries/financial-services/our-insights/deposits-the-top-profitability-lever-for-retail-banks-ceos>

Noori, B. (2021). Classification of Customer Reviews Using Machine Learning Algorithms. *Applied Artificial Intelligence*, *35*(8), 1–22. <https://doi.org/10.1080/08839514.2021.1922843>

Pattnaik, D., Ray, S., & Raman, R. (2024). Applications of artificial intelligence and machine learning in the financial services industry: A bibliometric review. *Heliyon*, *10*(1), e23492. <https://www.sciencedirect.com/science/article/pii/S2405844023107006>

Polireddi, N. S. A. (2024). An effective role of artificial intelligence and machine learning in banking sector. *Measurement. Sensors*, *33*, 101135–101135. ScienceDirect. <https://doi.org/10.1016/j.measen.2024.101135>

Shaikh, A. A., Kumar, A., Mishra, A., & Elahi, Y. A. (2024). A study of customer satisfaction in using banking services through Artificial Intelligence (AI) in India. *Public Administration and Policy*, *27*(2), 167–181. <https://doi.org/10.1108/pap-05-2023-0060>

Shridhar, M., & Ashwani, C. (2024). NETFLIX SUBSCRIPTION FORECASTING USING MACHINE LEARNING. In *International Research Journal of Modernization in Engineering Technology and Science* (pp. 2582–5208). <https://www.irjmets.com/uploadedfiles/paper//issue_7_july_2024/60568/final/fin_irjmets1721999278.pdf>

Singh, P. P., Anik, F. I., Senapati, R., Sinha, A., Sakib, N., & Hossain, E. (2023). Investigating customer churn in banking: A machine learning approach and visualization app for data science and management. *Data Science and Management*, *7*(1). <https://doi.org/10.1016/j.dsm.2023.09.002>

Văduva, A.-G., Oprea, S.-V., Andreea-Mihaela Niculae, Bâra, A., & Anca-Ioana Andreescu. (2024). Improving Churn Detection in the Banking Sector: A Machine Learning Approach with Probability Calibration Techniques. *Electronics*, *13*(22), 4527–4527. <https://doi.org/10.3390/electronics13224527>

Wang, G., Wang, Z., & Xie, Y. (2022). Subscribers Forecasting of Netflix Based on Multiple Linear Models. *BCP Business & Management*, *34*, 229–236. <https://doi.org/10.54691/bcpbm.v34i.3018>

Yakubu, I. N., & Abokor, A. H. (2020). Factors determining bank deposit growth in Turkey: an empirical analysis. *Rajagiri Management Journal*, *ahead-of-print*(ahead-of-print). <https://doi.org/10.1108/ramj-05-2020-0017>

Zaki, A., Nima Khodadadi, Wei Hong Lim, & Towfek, S. K. (2024). Predictive Analytics and Machine Learning in Direct Marketing for Anticipating Bank Term Deposit Subscriptions. *American Journal of Business and Operations Research*, *11*(1), 79–88. <https://doi.org/10.54216/ajbor.110110>