National College of Ireland

Facilitating Customers towards Product Subscription in Fin-Tech Application through Behavioural Analysis

MSc Research Project
Data Analytics

Lakshmiraj Natarajan
Student ID: x22173391

School of Computing
National College of Ireland

Supervisor: Jorge Basilio

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Lakshmiraj Natarajan
Student ID:	x22173391
Programme:	Data Analytics
Year:	2023 - 2024
Module:	MSc Research Project
Supervisor:	Jorge Basilio
Submission Due Date:	14/12/2023
Project Title:	Facilitating Customers towards Product Subscription in Fin-Tech Application through Behavioural Analysis
Word Count:	6396
Page Count:	21

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Lakshmiraj Natarajan
Date:	31st January 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Facilitating Customers towards Product Subscription in Fin-Tech Application through Behavioural Analysis

Lakshmiraj Natarajan
x22173391

Abstract

In the current scenario of financial technology as a business aspect predicting customer behavior plays a key role in effective marketing of their premium services. This research aims to identify the premium subscribers on the basis of the user behavior of the financial technology application with a specific concern of user who has an interest but have financial constraints. The idea for this research will help in building the business strategy for premium subscription promotion by utilizing optimal advertising resources and also ethically extending the promotions to the respective circle of users. The primary contribution of this research involves the development and implementation of a machine learning model built by using neural network algorithms to discover the patterns in behavioral aspects of the user which shows their willingness to subscribe the premium services of the application. The built model achieves an accuracy rate of 78.5 % which shows how the behavioral data is analyzed and provides valuable insights to analytics within the financial technology industry. The findings from this research will provide a base for the marketing campaign depending on the data driven environment and enables the financial technology sector to focus thier key resources towards the user shows their willingness towards the premium subscription. Although the model provides progressive further investigation needs to be taken to provide more values including incorporating more features to the dataset to support model progressiveness towards scalability and accuracy. Future works will address these areas to promote values to the fintech industry.

1 Introduction

In recent years technology and finance have coincided and emerged as Financial Technology companies grow at a fast pace and offer their business services digitally to withstand the current digitally evolved generation. Their application evolves in the way that show how customer interacts with their financial service product such as saving plan, credit facilities and loans. The financial technology company critical aspect is to drive the users from using the free version to premium version of application. By this aspect the company has both opportunity of covering larger scale of customer and challenges on strategy for retention of those customer.

1.1 Research Background and Motivation

In the current trend there are more digital banking application which has data about the behaviour of the customer how they use the application. So this can be used to learn the pattern of the user behaviour. Financial technology business will use this information for improving user experience and also helps to segmenting the customers who are likely to pay for the premium version of their application by reaching out them through promotional campaigning. This insights helps to minimise the resources on advertisement and promoting targeting advertisement to the segmented users.

The motivation of this research is to provide significant impact on how Fin-Tech industry sector will approaching the customer relationship management and respective market growth. with the help of model built by using deep learning algorithm can uncover the pattern on user behaviour and predict the user towards the product thus the business can target those user by using their marketing resources to provide promotional discounts for the premium service rather than spending more for marketing resources. The end use of the value insight provided by this research is to increasing the subscription and promotes business growth by optimisation of resource usage.

1.2 Research Question and Objectives

RQ: How can user interaction metrics from the digital banking application be used with deep learning techniques such as neural network to predict the potential premium subscription users, thereby optimizing target promotion strategy and reducing customer acquisition costs.

To address this research question, this research has the following objectives.

- To preprocess the dataset, feature engineered and standardizing the data.
- Visualize the preprocessed data to get preliminary insights.
- Building the model by using deep learning algorithm technique i.e., neural network used in this research.
- Evaluate the model performance by using metrics.
- Improve the model by regularisation, optimization and ensemble learning.

2 Related Work

Based on the business development background using mobile applications to provide service to acquire more customers across all industry sectors became popular. Particularly in the financial sectors they have been offering both free and paid versions i.e., premium versions of the application to figure out how people using the paid version in turn help to increase revenue generation with that application. This research tends to predict which customers were willing to pay for the premium version of the application. There is no previous research paper regarding the usage of this dataset to predict the customers who were willing to purchase the premium subscription. This literature study explores the various studies in which prediction models were developed to predict the results by analyzing customer behavior organized into the following subsections 2.1 Prediction modeling using customer behavior analysis, 2.2 Practical Implementation and Emerging Challenges.

2.1 Prediction modeling using customer behavior analysis

This review explores the prediction modelling on future action based on customer behavior analysis has become major factor for business on strategy operation management. Through this review on various paper based on using the machine learning algorithm shows Artificial Neural Network and Support Vector Machine based model on predicting customer action like fraudulent activity, churn, purchase intention result better accuracy.

VLN and Deeplakshmi (2021)shows the importance of Customer Relationship Management and building model using Support Vector Machine to predict churn in customer actions. Their work reflects the necessity of the organisation to gain insights from the customer behavior inorder to retain them and maintain the customer loyalty. Thus their work concludes with for an company inorder to acquire customer and on strategy for retain those customer building a robust prediction model is essential and SVM provides big benefit on that modeling process. Similarly Karvana et al. (2019) work explores the use of machine learning technique for predicting customer churn in banking sector. In their research they developed five different model for classification on the dataset which contain 57 attribute contain information regarding private bank Indonesia. On that study model that built by using SVM with the sampling ratio of 50:50 found to be more effective and the gained insights will help the company to built strategy to prevent customer churn.

Wen et al. (2018) work explores the experiment conducted based on customer purchasing behavior. In their study they developed a model called STPC-PGM (Spatial, Temporal, Payment, and Category in Probability Graphic Model) for predict the behavior of the customer based on the dataset contain customer financial information i.e., utilizing transaction data for building the model inorder to analyze across spatial, temporal, payment, and product category dimensions and predict the future action of the customer better over by using traditional methods. Through the resulted insights from their work will help to improve marketing strategy by predicting customer action with better accuracy and highlights the significance of mobile payment data for the marketing strategy. Liu et al. (2020) work explores the use of machine learning modeling to predict the user repurchasing behavior on shopping based on online platform. On the various algorithm used for modeling their study usage of XGBoost model gives better accuracy on predicting user repurchasing behavior. Their research also highlights the significance of feature selection and model fusion to optimise model accuracy on prediction.

In the pandemic situation happen during the time of COVID-19 Akter et al. (2022) work explores the utilization of model built by using machine learning algorithm to predict the user behavior on online shopping in Bangladesh. Their study shows the trend on online shopping behavior on upside based on customer behavior on the pandemic and help to build strategy to fulfill the customer need through their service more effectively by neglecting several barriers. Similarly Maheswari and Priya (2017) study explore the significance of using SVM algorithm in predicting the user behavior in online e-commerce shopping like frequency of the purchase, demography of the customer and preference of the product. Their work findings will assist the company to provide the promotional offers and helps in developing marketing strategy which inturn improve the sales.

Li et al. (2019) explores the usage of clustering analysis and classification algorithm to predicting behavior of the customer on scooter purchasing. For this study they using data resources from adventure works cycle which has attributes like user age,mmute distance etc... and highlights the significance of the model for customer segmentation and gain insights for improving sale marketing strategy development. Similarly Cirqueira et al. (2019) study explores model building to predict the customer prediction. The significance of this research by providing novelty on their research agenda and analytical framework for predicting user intent, purchasing decision and buying session on online platform. This paper gives significant understanding on predictive analysis by customer behavior is important for developing strategy customer need and experience of the shopping.

Similar study on financial sectorAmuda and Adeyemo (2019) work explores the utilization of model built by using Multi-layer Perceptron of Artificial Neural Network in predicting customer churn with effective accuracy. Their study highlights the significance of automated feature engineering that eliminates the needs for manual work tested to data source related financial sector in Nigeria. Their insights shows effective than using traditional methods and help user retention strategy development. Review on Baratzadeh and Hasheminejad (2022) works shows the fraudulent transaction detection in the financial sector using analysing the customer behavior. Their work show how deep learning techniques work effectively by comparing with other machine learning model like Random Forest and Support Vector Machine. And also shows the importance of choosing evaluation methods and tackles the issue of data imbalance in the modeling.

Khodabandehlou and Zivari Rahman (2017) works explores customer churn prediction through predictive analysis and also made comparison of various supervised machine learning technique. Their work highlights performance of Artificial Neural Network model better than other model by providing effective insights through better accuracy and understanding. And highlights the benefits of boosting technique. Similarly machine learning model built using Swish Recurrent Neural Network combined with Brownian motion-based Butterfly Optimization Algorithm aids effectively in the process of feature selection for predicting customer churn in telecommunication industry in Sudharsan and Ganesh (2022) provide insights effectively for analysing the user data through unique activation function that help in develop customer retention strategies.

Work by Zhao et al. (2021) shows the utilization Long Short Term Memory (LSTM) Neural Network for classification of customer in retail sector through recognising pattern which is better than traditional methods like logistic regression. support vector machine. And provides insights on customer segmentation and target marketing. Likewise predicting customer shopping pattern in works by Salehinejad and Rahnamayan (2016) shows the model developed by using Recurrent Neural Network algorithm with auto encoder for feature extraction and used rectified linear unit(relu) for activation function for effectively predicting customer shopping pattern. works by SCHUTTE (n.d.) highlights the usage of neural network technique in the complex application scenario i.e., real world prediction by facing challenges in providing better accuracy than the traditional method in prediction customer future purchasing date with animal feed company.

Similarly review on work by Jiang et al. (2021) predicts of customers location within shopping mall through analysing customer behaviors and recognise the customer pattern. They built the model that utilize the data from customers mobile device and WiFi information withing the shopping mall. And also shows the effectiveness of using XGBoost learning algorithm for predicts the customers location which aids to provide tailored service to the customer. And Tanuwijaya et al. (2021) case study showcase the use of machine learning modeling to identify the potential user for Netflix subscription through K-means clustering to identifies the significant feature that influence the prediction. The insights driven from the model valuable in optimise the strategy on marketing to fulfils the customer needs. And in the review of works by Meshrarm and Bekouni (2021) focuses the prediction and classification of user in the network sector through user behavior analysis in the backdrop of marketing strategy. In this work they deploy RFML algorithm to segmenting the user and used Artificial Neural Network with Time series modeling to predict the customer behavior. Their works provides significant insights to understand the user behavior in digital marketing sector.

Similar to this research work, Ghochani et al. (2013) work assist in analyse and predicts the customer behavior.They focused on service based industries deal with large amount of data. The approach is aims to predict the user preference by using Artificial Neural Network to analyse the user previous purchasing history to predict the user behavior which aids in improving the marketing strategy which in turn potentially increase the profits.

Thus, the above work's review highlights the impact of detecting complex user behavior pattern that reflects in the development of business strategy by employing various machine learning algorithm on fraudulent detection, churn predictions etc...

2.2 Practical Implementation and Emerging Challenges

This section covers the certain challenges in their implementation to give practical solution by analysing the customer behavior across different sectors like finance, telecommunication, digital marketing, e-commerce etc...

In the field of telecommunication sector as background, works proposed by VLN and Deeplakshmi (2021) and Amuda and Adeyemo (2019) provides insights to improve strategy on customer retention by carefully evaluate the user satisfaction with maintaining the data accuracy and data privacy by predictive analysis using Artificial Neural Network and Support Vector Machine algorithm on customer churn prediction.

Karvana et al. (2019) works highlights the importance of maintaining the optimal balance between the selection of model and profitability in the background of banking sector. And address the complex challenge of selection of model while considering the customer satisfaction level.

Predcitive modeling in e-commerce sector, reviews on works by Wen et al. (2018) and Liu et al. (2020) gives understanding of user purchase behavior for predict the probability of repurchase by utilizing graphical models and XGBoost models. In these works the complex challenges is the models requires data preprocessing and feature selection for improving the accuracy of prediction. Models used in Akter et al. (2022) and Maheswari and Priya (2017) to predict the customer choice of shopping in e-commerce background highlights the importance of choosing the model adaptability which apt for the market condition and customer preferences were challenges faced. And Li et al. (2019), Cirqueira et al. (2019) works highlights teh requirement of advance machine learning technique to build a comprehensive framework inorder to predict the customer purchases.

Baratzadeh and Hasheminejad (2022) highlights the use of deep learning in detection of fraudulent transaction faces the complex challenges to make optimal balance between the model interpretability and complexity and shows the necessity of model changes to be done to catch with fraud trends.

Implementation of these predictive model on various sector highlights the usage of insights gained by each work for their respective sector like enhancing marketing strategy, maintaining customer relationship management. These works faces several complex challenges like data privacy concerns, continuous model updating due to dynamic nature of customer behavior and complexity of model interpretability.

3 Methodology

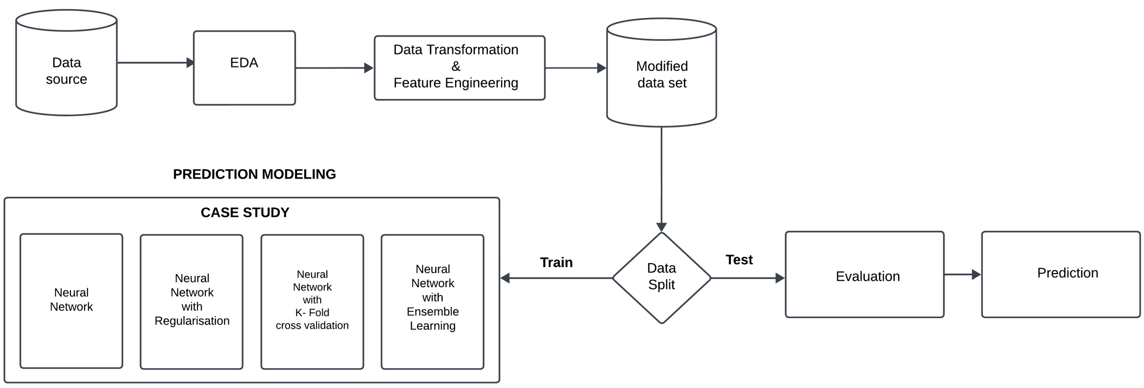


Figure 1: Research Methodology

As business which has the customer as an end point utilized the machine learning techniques to analyse the customer pattern and promote their product to improve the revenue generation has made in recent years.This project is to identify the customer who were willing to subscribe the premium product of the fin-tech application but has financial constraints so that the company can use the insights gained by this research to provide targeting promotion to acquire new customer and optimise the resource allocation on advertising. This work is developed through the machine learning modeling using Neural Network. CRISP-DM which is business centric approach methodology used in this project. A structural approach driven according to research need utilized in this project and explained in Fig. 1. This section aim to develop a methodology that can be replicate to do this project to validate the results.

3.1 Data Source

The source of the dataset for this research is taken from kaggle which is an open source dataset repository for researches. The benefit of using data source from kaggle, resources need to undergo review process to verify all the ethical practices were followed. Dataset used for this research focussed on the behaviors of the customer regarding the fintech mobile application usage. Analyzing this dataset will able to provide significant insights regarding subscription trends. The reason behind choosing this dataset will be useful for effective in customer behavior analysis.

3.2 Detailed Methodology

The methodology applied on this research was to provide the clear view of customer behavior analysis through the dataset obtained from the kaggle. This research project was developed using Google Colab which is a cloud-based development environment that helps to execute complex analytic procedures efficiently even on the machine with less hardware capabilities.

3.2.1 Data Collection and Mounting :

The primary step is to acquire dataset from Kaggle. Then the dataset need to mounted onto Google Colab via Google drive for access for analysis and manipulation.

3.2.2 Exploratory Data Analysis and Visualization :

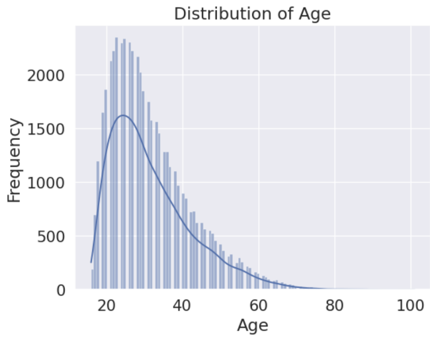
To obtain the primary understanding of the data, an exploratory data analysis (EDA) was done using library tools of python ie., Matplotlib and Seaborn. These tools were used in visualizing various aspects of the dataset.

In Fig.2(a), Observations were made among the distribution on age with the frequency of users with the application give the distribution pattern i.e., customers of age between 20 to 40 using the application than others.

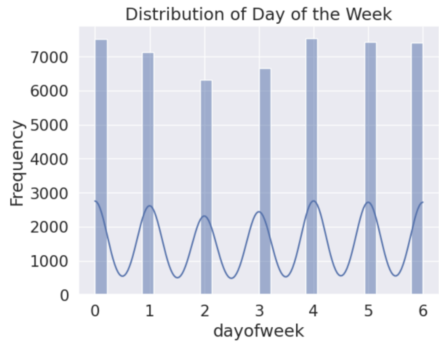
In Fig.2(b), Observations were made among the distribution on the day with the frequency of users with the application give the distribution pattern i.e., the user engagement were uniform over all day in the week.

In Fig.2(c), Observations were made among the distribution on the day with the frequency of users with the application gives the distribution pattern i.e., hours were more customer involved was around 10 am and 6 pm driven through the peak curve and user involvement drops after 8 pm.

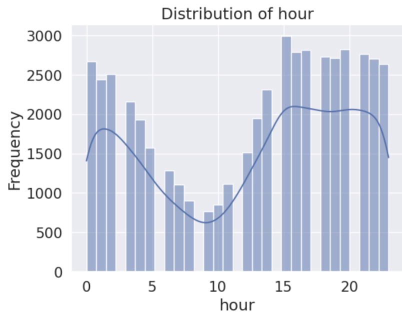
In Fig. 3. Correlation matrix developed and visualized using a heat map to analyse the feature relationship and correlation level with enrollment in premium subscription i.e., response variable. From the matrix observation shows the moderate level of positive correlation between the used premium features and mini game participants and moderate level of negative correlation between age and the number of screen that reflects the younger age user visit and utilise the more number of screens in the application.



(a) distribution on age



(b) distribution on day



(c) distribution on hours

Figure 2: Distribution of Features with Frequency of Users

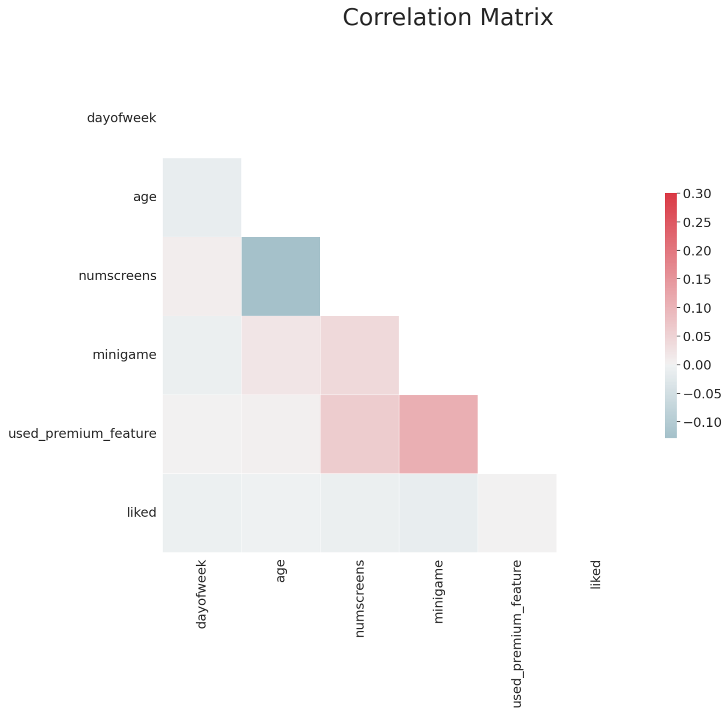


Figure 3: Correlation Matrix

3.2.3 Data Preprocessing and Transformation:

In the aspect of data preprocessing stage several data process and transformation were undergone. Distinct screen string value were recognised from screen list feature. processing involved in the comma separated list of scene list value in this feature used in the feature engineering process and aids in understanding the customer usage behavior pattern.

3.2.4 Feature Engineering :

Process involved in feature engineering provide a major role in this research by developing the structure for the data from the raw data source collected from the data source for utilization in the model to aids the research aim i.e., to develop prediction model to predict the user for premium subscription. The following step need to process as a part of feature engineering.

Transformation on DataType: Feature labeled first open and enrolled date need to change their datatype as datetime object in order to calculate the time difference to calculate the time frame for the measure typical customer time frame to get enrolled in premium subscription after installation.

Defining Timeframe : To validate the predictive model effectively, establishing a timeframe for user conversion taken as 50 hours by analysing with distribution of time difference between enrolled date and first open date variable with the frequency of user from Fig. 4. inorder to maintain balance in predictive analysis. So user who have enrolled after 50 hours were reclassified as non converters.

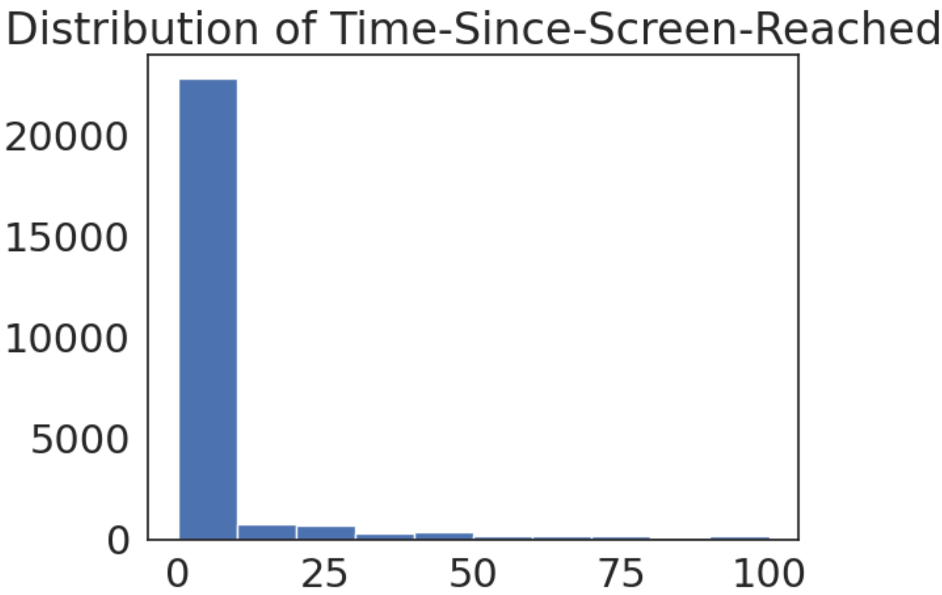


Figure 4: Distribution on time duration with user frequency

Truncation of Features : After defining the time frame for research scope excessive feature such as difference, enrolled data, first open were removed to simplify the research work.

External Data Integration: The data source has a external file 'topscreens.csv' which has commonly used screen values of the application. This can be used to create a separate feature called top screen which has intersecting values of each user screen list with the top screen. And the rest of the value form the screen list of each user make into the 'other screenlist' feature.

Funnel Creation : Further simplifying the process of capturing the customer behavior pattern, certain screen list feature values are recognised as sequence. so creating a funnel to group those screen list values i.e., saving, credit card and credit monitoring for complexity optimisation.

Modified Dataset : The preprocessed dataset modified after these process stored as 'feature engineered data.csv' for the remaining phase of model development and analysis.

4 Design Specification

This research implementing neural network algorithm to predict customer for premium subscriptions in a FinTech application. This neural network model designed for binary classification for predicting the customer for premium subscription service based on their behavior within the application. This neural network architecture is a multi layered model specifically for the research task i.e, binary classification. It consist of arrangement of sequential layers, designed to process the data on each layer and transmit the data to the forward layer, finally leading to the output layer i.e., prediction on analysing user behavior in fintech application, as illustrated in the flowchart in Fig. 5.

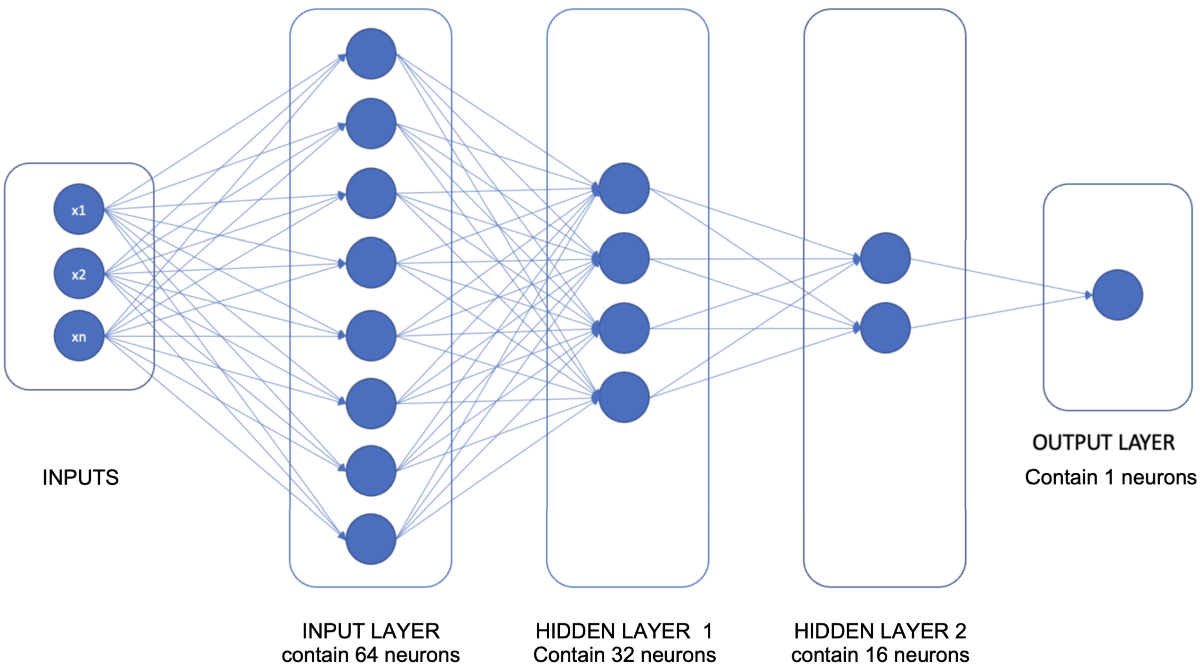


Figure 5: Model Design

- **Input Layer:** This is the starting point of the model where pre-processed data received i.e, Input Layer. This layer defined by its dimension as X train.shape[1], which represents the number of features present in the training dataset. The key role of this layer is to sets the stage for the network to build framework to handle the specific size and shape of the data to this layer.

- **Hidden Layer 1:** This beginning layer compromise of 64 neurons and act as primary unit for processing in neural network and incorporate ReLU activation function to bring non linearity nature to the model.This primary layer is responsible for capturing the complex pattern and enables the learning of the model and initiate the neural network model training.
- **Hidden Layer 2:** This hidden layer compromises 32 neurons and incorporate ReLU activation function in this layer too. which responsible to develop the feature transformation and generalisation performed by the hidden layer 1 for further analysis on the improved input data.
- **Hidden Layer 3:**This hidden layer compromises 16 neurons and maintaining the usage of ReLU activation function. This way of processing the data into the hidden layer consist of gradual decrement in the number of neurons promotes the enhancement and integration of analysed information within the model, which is responsible for better accuracy in the output.
- **Output Layer:** This final layer in the model i.e, output layer incorporates sigmoid activation function which is especially suited for binary classification task. This function output values ranges from 0 to 1 which indicated the users probability towards the respective output classes i.e, subscriber or non subscriber.

Model Compilation: Once the layers of model defined and get modeled, then the model gets compilation. In this process it is necessary to employ optimiser and loss function. In this research for optimiser, 'adam' optimiser employed inorder to handle the sparse variation and also it dynamically increase the learning rate of the model during the training process. And for the loss function 'binary crossentropy' is employed because of its capacity in evaluating the performance of the model i.e., binary classification probability value ranges from 0 to 1.

5 Experiment and Evaluation

5.1 Case Study 1: Neural Network

After designing the model architecture for the base machine learning model by using a Neural Network Algorithm to predict customers for premium subscriptions experiments were conducted by using this model as a base and the models were developed to enhance the performance for the task i.e., binary classification.

Model Structure: Initialisation of sequential model integrated with hidden layers, Initial input layer contains 64 neuron units configured with Rectified Linear Unit(ReLU) as an activation function to match the similar feature from the training data followed by the first hidden layer and second hidden layer equipped with 32 and 16 units respectively configure with ReLU on both layers for activation function. Follower by the output layer configured with Sigmoid Function for its nature of applicability in binary classification task.

Model Compilation and Training:

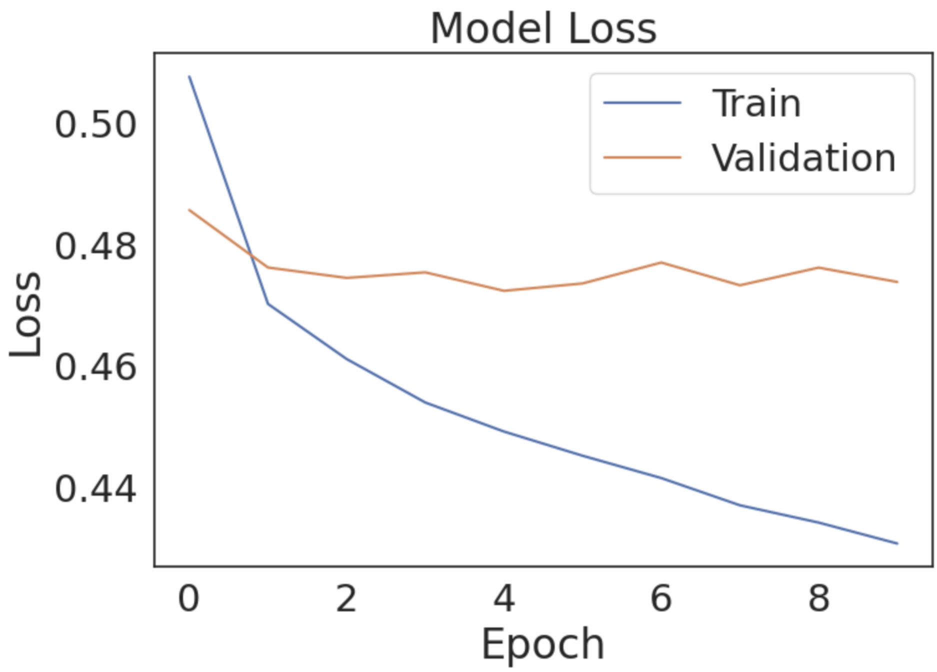


Figure 6: Learning Curve Analysis

The initial model developed was compiled with 'Adam' optimizer algorithm and binary crossentropy for handling loss function. Followed by the model compilation with a base consideration of 10 epoch cycles for training the neural network model for calculating the significant value for the epoch with the validation split of 20 % for analyze the performance. After the model is trained analyze the learning curve of the model plotted against training and validation loss to observe whether the performance of the model towards overfit or underfit as in Fig. 6. observed that the training loss went below the validation loss and not get into a constant state results in the probability for the overfitting condition towards the model. To overcome this issue early stopping technique was employed to stop the training of the model whenever the condition fallback to prevent overfitting of the model by the condition described in Fig. 7. as monitoring the validation loss metric to stop if it increases on the training epochs and also if its on the same value carry over three epoch will stop that training of the model to prevent the overfitting condition.

```
# Early Stopping
early_stopping = EarlyStopping(monitor='val_loss', min_delta=0.001, patience=3, verbose=1, mode='min')
```

Figure 7: Early Stopping Technique

Training results The model training shows the improvement in the initial epoch by reflecting lower validation loss and then a short spike during the fourth epoch and passing the condition set for the early stopping parameter and stooped the training of the model with the validation accuracy around 76.92 % over the fourth epoch.

Model Evaluation Thus the neural network model shows the accuracy after testing on the test dataset with an accuracy of 77.59 %. Confusion matrix plots to determine the True Positive, True Negative, False Positive and False Negative coefficients in Fig.

8. reflect the efficiency of the model showing a balance between True Positive and True Negative and also shows relatively high values for False Positive and False Negative which shows the high cost consumption for showing those values.

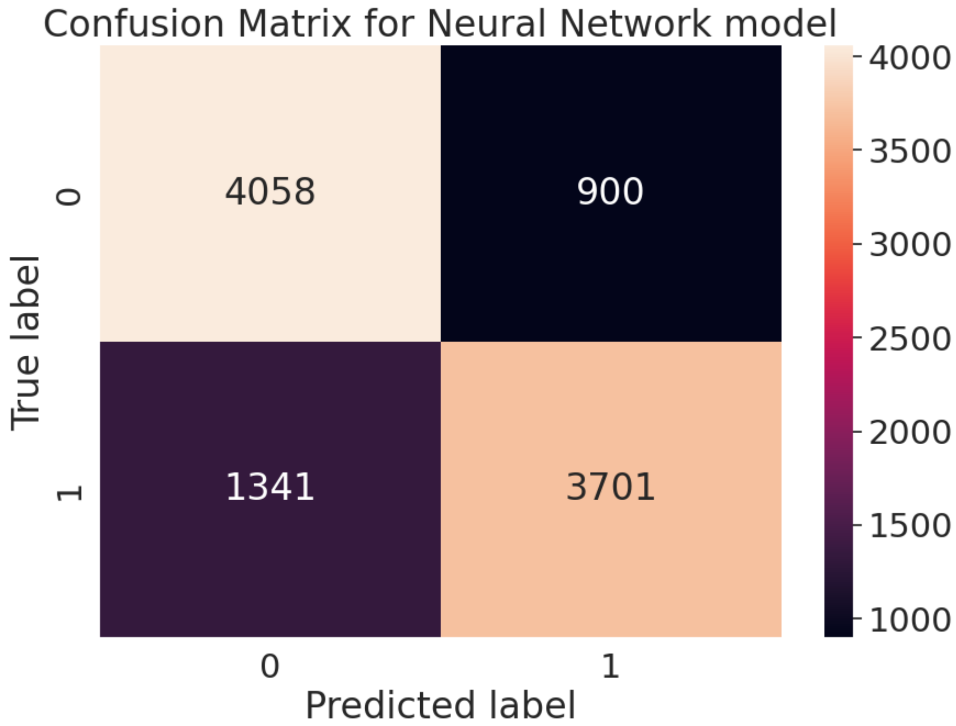


Figure 8: Confusion matrix for Neural Network model

This model served as a foundation for the subsequent experiments to process the predictive analysis task driven by the research problem.

5.2 Case Study 2: Neural Network with Regularisation

Implementing Regularisation: This second Case study was performed by incorporating the L2 regularisation technique because of the nature of directing to unique and more responsive model stability. It prevents the model from overfitting by providing a penalty to the loss function over the training process which is proportional to the square of the magnitude of the weight by applying through all the layers of the Neural network model which is initially built.

Model Compilation and Training: Thus the Network model applying regularisation technique on each layer of lambda value 0f 0.01 to enhance model generalization and prevent model overfitting complied with Adam optimizer with binary crossentropy for handling loss function. Followed by the model training process over 10 epochs with the validation split of 20 % from train data showing the training loss value on the higher side during the initial epoch due to the penalty by the regularisation technique and got stabled over the epochs.

Model Evaluation: Followed by the model training evaluating with the trained model against the test data shows the accuracy of 77.25 % and the validation loss value consistently decreases towards till the last epoch reflects model doesn't experience overfit. The performance of this model will be observed through other metrics such as confusion matrix, precision, recall and F1 score will reflect the efficiency of this model in binary classification i.e., predicting the customer for premium subscription.

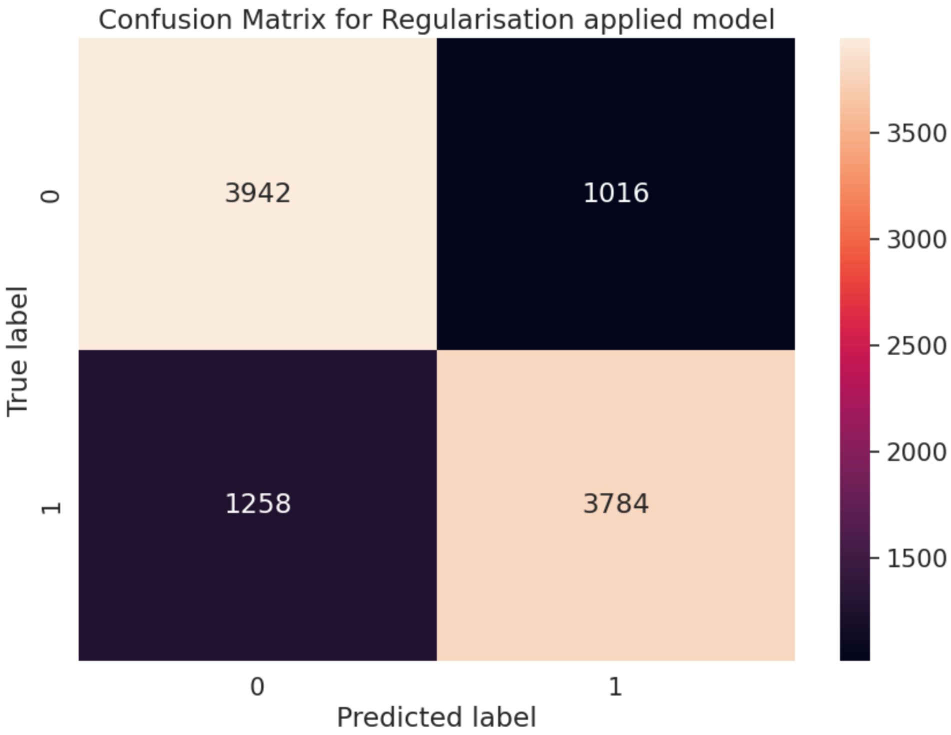


Figure 9: Confusion matrix for Regularisation applied model

Confusion matrix plots to determine the True Positive, True Negative, False Positive and False Negative coefficients in Fig.9. showing similar values to the Neural Network model from the above case study but with higher values for False Positive and False Negative coefficients showing it provides more cost for those coefficients.

This case study presents the significance of applying regularisation techniques in the prediction model to enhance model generalization by providing penalization to the loss function of the model.

5.3 Case Study 3: Neural Network with K-Fold Cross validation

Implementing K-Fold Cross validation: This case study was built on the similar model structure of Neural Network on the first case study with the three hidden layers with ReLU as activation function and the output layer with Sigmoid function for binary classification. StratifiedKFold library function provided by the sci-kit learn tool kit to divide the data source into k-fold time. In this research data source divided the value assigned to k-fold i.e., 5. Thus the dataset was divided into 5 equal parts and each individual part acted as a test data set while the other fold acted as a training dataset. It works in the way that each fold passes into the test phase while other folds were trained at that time and ensure the cross validation integrity.

Model Training and Validation: Models will be created for each fold and also maintain the learning from one model to not influence the other models. And also integrated with an early stopping technique similar to the first case study to prevent overfitting. And parameter with the value of one for verbose to get a clear log of the training process of the model.

Model Evaluation: Followed by the model training. The efficiency of the model is evaluated by the trained model against the test data where each data fold was evaluated against the test folds performance metrics were generated for each fold and the metrics namely accuracy, precision, F1 and recall were taken mean over metric values of each folds and observed the accuracy of 77.78 %.

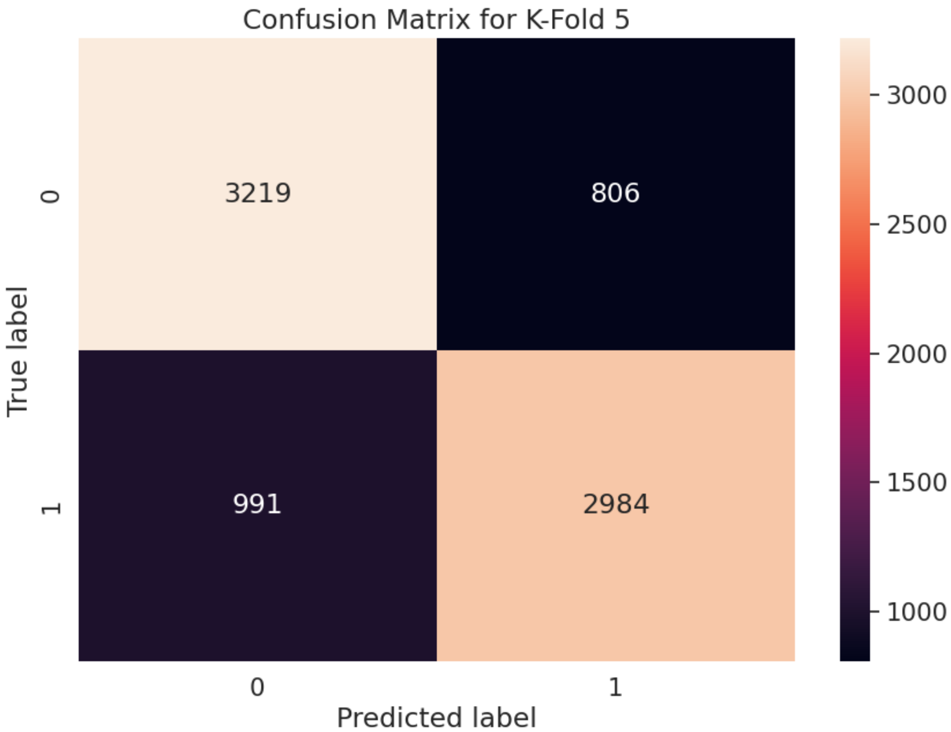


Figure 10: Confusion matrix for K-fold cross validation model

The confusion matrix plotted to calculate the True Positive, True Negative, False Positive and False Negative coefficient values in Fig. 10. where the respective model shows the less value between True Positive and True Negative and also shows less values for False Positive and False Negative shows the better efficiency on cost for those values.

5.4 Case Study 4: Neural Network with Ensemble Learning

This fourth case study experiment was built on the model using a neural network algorithm which is similar to the first case study but incorporates ensemble learning with that model to seek enhancement in effectiveness in the predictive analysis. And to build an effective system for predicting customers for premium subscriptions which is the research aim.

Modeling Strategy: The strategy planned in this case study experiment is to combine the multiple models trained and average their performance for the prediction. In this study, this research planned to incorporate three neural network models trained separately and the ensemble approach is to access the perspective of those multiple models and average their evaluation to enhance the overall efficiency of the prediction.

Model Training and Evaluation: Each model has initialized by the sequential model of architecture consisting of three hidden layers which consist of 64, 32 and 16 units on each respective layer with ReLU for the activation function and followed by the output layer having the sigmoid function for its applicability of usage in binary classification. Then each model was compiled and trained one after another integrated with the early stopping technique to stop the training of the of the respective model if its condition passes by the validation loss of the respective model. In this research case study model for the first neural network model training stopped at the cycle of epoch 6 where the validation loss value spiked up compared to the before epoch and passed the condition of early stopping technique and stopped training in order to prevent overfitting. This was followed by the second model stopped at epoch 7 and the third respective model's validation loss spiked up over the seventh epoch and stopped from the training. The performance metrics were calculated by taking the average over the three models and the metrics i.e., accuracy, precision, F1 score and recall were calculated with an accuracy of 78.22 %.

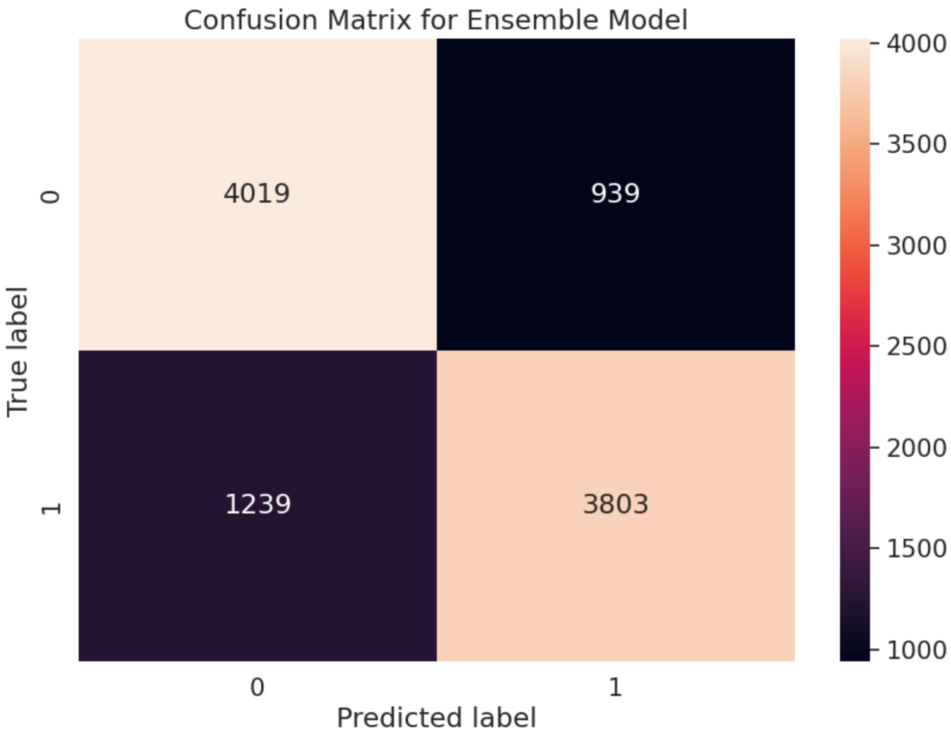


Figure 11: Confusion matrix for ensemble method case study

The Confusion matrix plotted to calculate the True Positive, True Negative, False Positive and False Negative coefficient values in Fig. 11. in which this case study balances high values for True Positive and True Negative. This case study concludes and the next section will compare the case study result and more efficient model in the aspect of this research problem

5.5 Discussion

Table 1: Summary of Model Results

Method	**Accuracy**	**F1 Score**	**Precision**	**Recall**
Neural Network	77.59%	76.76%	80.43%	73.40%
Regularisation	77.25%	76.89%	78.83%	75.04%
K-Fold Cross-Validation	77.78%	76.93%	79.44%	74.59%
Ensemble Method	78.22%	77.73%	80.19%	75.42%

In this research aspect, the primary aim is to predict the customers who were willing to enroll in the premium subscription of the product i.e., fintech mobile application. Case study experiment conducted by applying deep learning knowledge in the modelling and four studies were conducted. Among that case study that incorporates ensemble learning on the neural network model resulted as superior performance in the aspect of much effective model with the accuracy of 78.22 % of classifying the subscriber and non subscriber customers. By considering accuracy directly reflects the effectiveness in real time scenario were error in classifying the False Positive and False Negative make the value for business critical and questionable.

And this Ensemble Method case study has 77.73 % in F1 score metrics which is more valuable metrics than the accuracy in the scenario of handling imbalenced dataset. And precision metrics value of 80.19 % which reflects the positively predicted were actually true. This will reduce the risk of predicting incorrect as correct in this research this lead to spending marketing resources on the wrong customer who were not willing to subscribe the premium product of the company. And this ensemble case study model has significantly higher recall percentage i.e., 75.42 which play a vital role in increase revenue for the company from the mobile application were observed from the Table 1.

Table 2: Confusion Matrices

Method	**Confusion Matrix (TP, FP, FN, TN)**
Neural Network	3701, 900, 1341, 4058
Regularisation	3784, 1016, 1258, 3942
K-Fold Cross-Validation	2984, 806, 991, 3219
Ensemble Method	3803, 939, 1239, 4019

From the Confusion matrix analysis in Table 2. Ensemble method model has better distribution in True Positive of 3803 and True Negative of 4019 when considered as combined even with False Positive of 939 and False Negative of 1239 which addresses this model's reliability and lower in risk of miss classification.These Findings will play key role business understanding where understanding the customer is crucial task in behavior analysis i.e., research question. Analysing the customer behavior to predict them whether they were will buy the premium product of mobile application of fin tech company in the period of usage of their trail version. Thus analysing and classifying the customer behavior whether they willing to pay for premium product will help in decision making and allocating resources on marketing to target those customers.

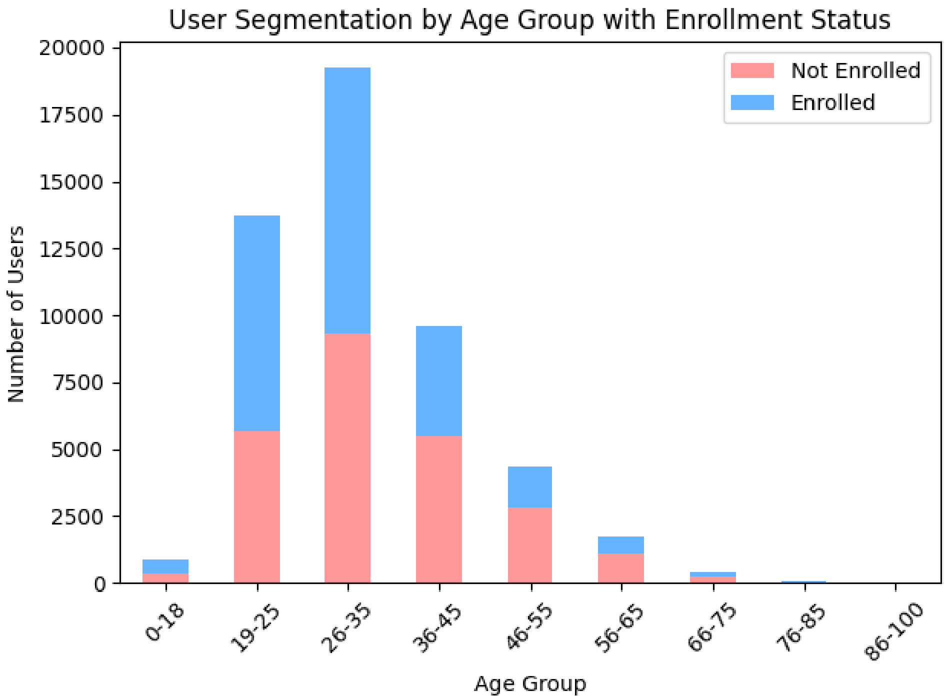


Figure 12: User segmentation based on Age

Fig. 12. provides a visual representation of segmentation of User age with enrolled and not enrolled customers. which show the customer around the age group between 19 and 25 has higher enrollment ratio.

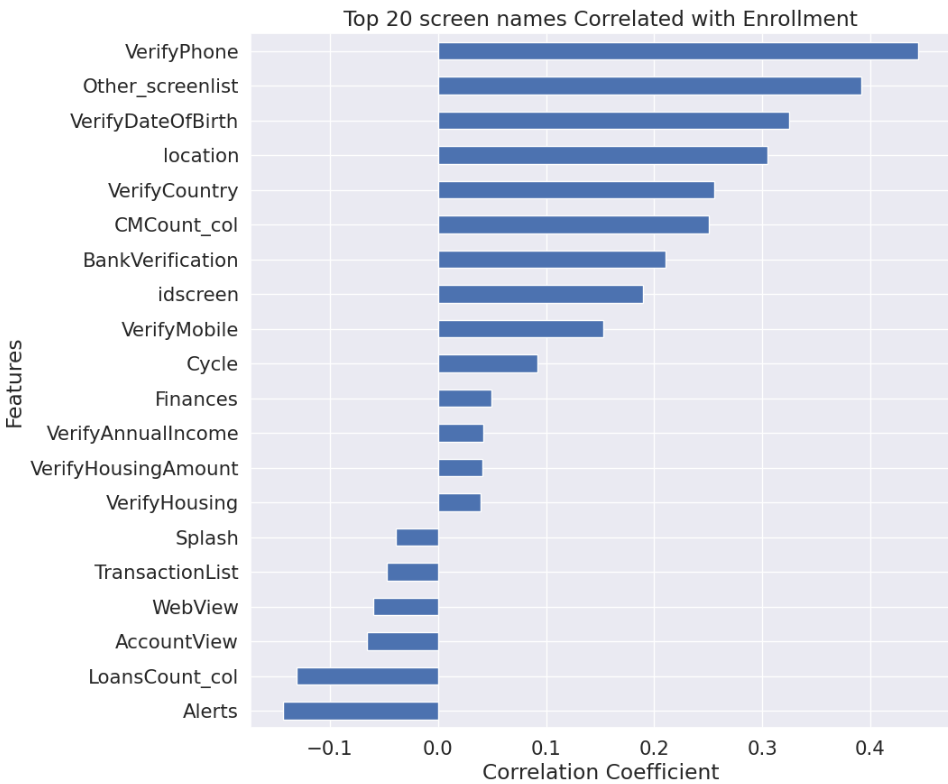


Figure 13: Top 20 screen correlation with enrolled feature

Fig. 13. represent the top 20 screen list feature correlation with enrollment feature in both positive and negative correlation relationship. This insights will assists the to focus on those feature to increase the user engagement by optimising those feature towards development.

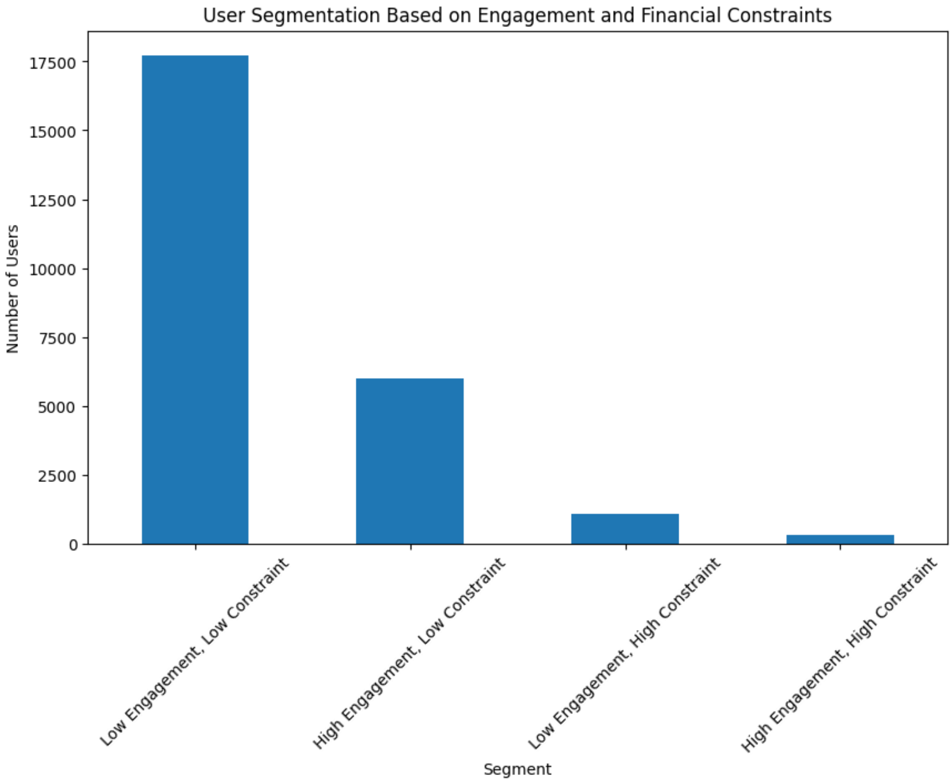


Figure 14: Customer Engagement based on financial constraints

Fig. 14. represent the Customer segmentation who were not enrolled can be segmented based on the engagement towards product and financial constrains. engagement constraint can be derived from app feature like 'BankVerification','VerifyPhone' and finance contrains can derived from feature like 'LoansCount col' reflects information about number of loans , 'CCCount col' reflects information about number of credit card , 'used premium feature' reflect usage of premium featrue during the trail period. using these constrains customer engagement graph driven based on financial constrains. With this customer segmentation graph the fintech company can tailored the offers to high engagement user but with high financial constrains like flexible payment method to attract more customer towards the product subscription which inturn increase the revenue generation.

6 Conclusion and Future Work

Thus through this research case study experiment by building four experiment model developed by using ensemble method to predict the customers for product subscription based on the customer behavior analysis within the fintech mobile application with the accuracy 78.22 % and also shoe better evaluation metrics comparing to other model cases. Through this Predictive analysis, models can be used to make target promotion the customer will aids in minimising the resources spending on marketing.

This research has some limitation in spite of deploying effective approach. The accuracy of the model is relies on the aspect of data gathered which might not cover some personal data that represent the possible interaction of the user. Furthermore with the interaction measurement might not fully reflect to predict the complex user decision. Ethical consideration need to made in promotional process to main user privacy on the acquired information. Due to this frequent evolving Fin-Tech industry, to maintain and improve the model accuracy frequent model update need to be made to stay update. Future works including the exploration of this domain into real time analysis to promote enhanced personalised marketing to make more user engagement. And extending research in various sector which involves behavior analysis as key resources for business growth by leveraging machine learning algorithms.

References

Akter, S., Islam, M. K., Hossain, M. N., Rahman, M. and Boshra, S. J. (2022). People thoughts prediction using machine learning on online shopping in bangladesh during covid-19 pandemic, *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, IEEE, pp. 1–7.

Amuda, K. A. and Adeyemo, A. B. (2019). Customers churn prediction in financial institution using artificial neural network, *arXiv preprint arXiv:1912.11346* .

Baratzadeh, F. and Hasheminejad, S. M. (2022). Customer behavior analysis to improve detection of fraudulent transactions using deep learning, *Journal of AI and Data Mining* **10**(1): 87–101.

Cirqueira, D., Hofer, M., Nedbal, D., Helfert, M. and Bezbradica, M. (2019). Customer purchase behavior prediction in e-commerce: A conceptual framework and research agenda, *International workshop on new frontiers in mining complex patterns*, Springer, pp. 119–136.

Ghochani, M., Afzalian, M., Gheitasi, S. and Gheitasi, S. (2013). Simulation of customer behavior using artificial neural network techniques, *International Journal of Information, Business and Management* **5**(2): 59.

Jiang, H., He, M., Xi, Y. and Zeng, J. (2021). Machine-learning-based user position prediction and behavior analysis for location services, *Information* **12**(5): 180.

Karvana, K. G. M., Yazid, S., Syalim, A. and Mursanto, P. (2019). Customer churn analysis and prediction using data mining models in banking industry, *2019 international workshop on big data and information security (IWBIS)*, IEEE, pp. 33–38.

Khodabandehlou, S. and Zivari Rahman, M. (2017). Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior, *Journal of Systems and Information Technology* **19**(1/2): 65–93.

Li, J., Pan, S., Huang, L. et al. (2019). A machine learning based method for customer behavior prediction, *Tehnički vjesnik* **26**(6): 1670–1676.

Liu, C.-J., Huang, T.-S., Ho, P.-T., Huang, J.-C. and Hsieh, C.-T. (2020). Machine learning-based e-commerce platform repurchase customer prediction model, *Plos one* **15**(12): e0243105.

Maheswari, K. and Priya, P. P. A. (2017). Predicting customer behavior in online shopping using svm classifier, *2017 IEEE international conference on intelligent techniques in control, optimization and signal processing (INCOS)*, IEEE, pp. 1–5.

Meshrarm, D. and Bekouni, A. (2021). Prediction and classification of customer behavior in electronic marketing using data mining and neural network (a case study of network marketing companies), *Applied Innovations in Industrial Management* **1**(2): 1–7.

Salehinejad, H. and Rahnamayan, S. (2016). Customer shopping pattern prediction: A recurrent neural network approach, *2016 IEEE symposium series on computational intelligence (SSCI)*, IEEE, pp. 1–6.

SCHUTTE, J. (n.d.). *ANALYZING CUSTOMER BEHAVIOR USING NEURAL NETWORKS*, PhD thesis, tilburg university.

Sudharsan, R. and Ganesh, E. (2022). A swish rnn based customer churn prediction for the telecom industry with a novel feature selection strategy, *Connection Science* **34**(1): 1855–1876.

Tanuwijaya, S., Alamsyah, A. and Ariyanti, M. (2021). Mobile customer behaviour predictive analysis for targeting netflix potential customer, *2021 9th International Conference on Information and Communication Technology (ICoICT)*, IEEE, pp. 348–352.

VLN, R. K. and Deeplakshmi, P. (2021). Dynamic churn prediction using machine learning algorithms-predict your customer through customer behaviour, *2021 International Conference on Computer Communication and Informatics (ICCCI)*, IEEE, pp. 1–6.

Wen, Y.-T., Yeh, P.-W., Tsai, T.-H., Peng, W.-C. and Shuai, H.-H. (2018). Customer purchase behavior prediction from payment datasets, *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pp. 628–636.

Zhao, L., Zuo, Y., Yada, K. and Liu, M. (2021). Application of long short-term memory based neural network for classification of customer behavior, *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE, pp. 994–999.