



CUSTOMER ANALYSIS FOR BANKS USING MACHINE LEARNING

¹Prakhar Singh, ²Devendra Singh, ³Pavraj Thapliyal, ⁴Vedant Garg

¹ B.Tech Student, Department of Computer Science & Engineering, SRM Institute of Science & Technology, Ghaziabad,

²B.Tech Student, Department of Computer Science & Engineering, SRM Institute of Science & Technology, Ghaziabad,

³B.Tech Student, Department of Computer Science & Engineering, SRM Institute of Science & Technology, Ghaziabad,

⁴B.Tech Student, Department of Computer Science & Engineering, SRM Institute of Science & Technology, Ghaziabad,

Abstract: In order to generate good revenue, every business needs to make proper data-oriented decisions. These decisions lead to the emergence of successful business strategies, which themselves are data driven. This is even more relevant in the world of banking, where the stakes are quite high. The huge number of customers increase the risk of a massive default on credit cards if banks are not being careful. It becomes difficult for banks to cater to the individual needs of customers. Therefore, we propose the usage of these machine learning practices to predict the next step, sort your data, and apply multiple algorithms to your dataset to find your target audience. Using these practices, we will be able to predict whether a customer will default their next credit card bill or will be able to repay their loan and organising the customers so that bank can manage their resources to cater to the right group of people who will buy their different schemes.

Index Terms - Dataset, Classifications, Machine Learning.

I. INTRODUCTION

The customer analysis application will support all platforms with the single codebase using python. With this application banks can easily predict clients who are likely to subscribe to their term deposit, whom should they give loan too, or predict if a customer will default their credit card payment, and help banks understand better their customers and how to best to communicate with them. This will increase the banks efficiency and effectiveness. The machine learning part will be done using scikit learn library in python, for the UI part this application will be using streamlit (python)/HTML/CSS, and for deployment Heroku and GitHub pages.

II. ANACONDA

Anaconda is a software tool which is generally made for the Python and R languages. This tool is basically the birthplace of Data Science. This distribution consists of various data-science packages suitable for Microsoft Windows and Linux OS. This tool consists of various Python interpreters, R Studio and Visualization tools as shown in Figure 1.

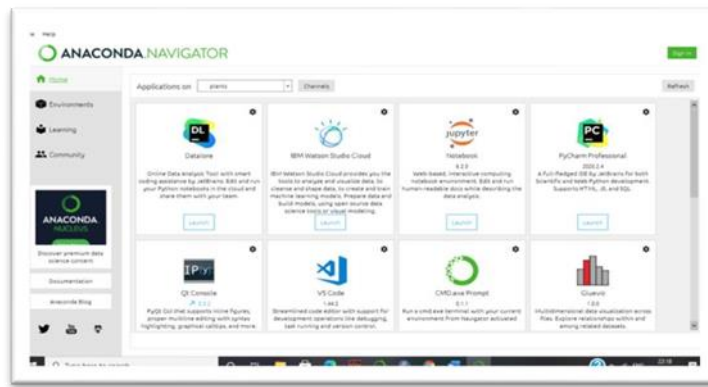


Fig 1: Anaconda Interface

III. JUPYTER NOTEBOOK

Jupyter Notebook is one of the best Python interpreter provided by the Anaconda which is used for performing machine learning and data science processes. It is available free and comfortable to use.

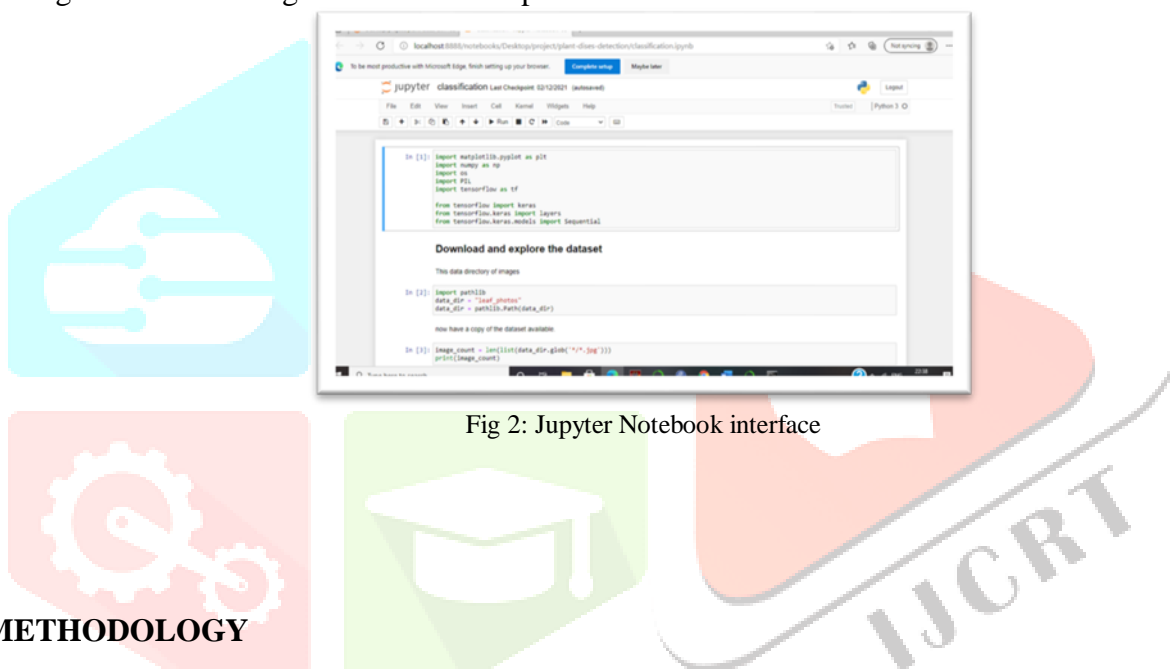


Fig 2: Jupyter Notebook interface

IV. METHODOLOGY

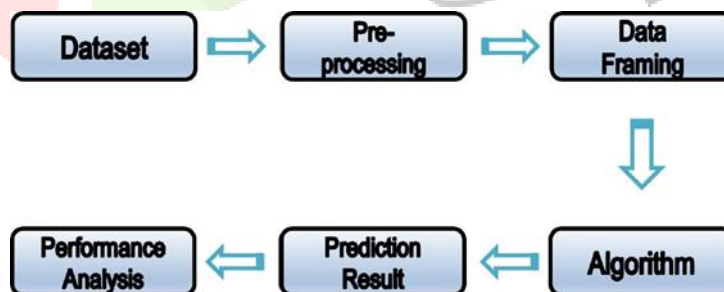


Fig 3: Methodology representation

The dataset is downloaded from Kaggle.

PREPROCESSING

The stage where some extra information is removed from the collected data. In this stage we work on data to remove the outliers and any other unnecessary data. A number of steps are involved in preprocessing. Some of them are: data cleaning, Feature scaling, Feature encoding, Feature selection and Dimensionality reduction.

DATA FRAMING

In this stage we select the columns that suits best to our model using different techniques.

ALGORITHM/RESULT/ANALYSIS

In this step we choose the algorithm which gives better accuracy. This is done by running different algorithms through our dataset using cross validation and selecting the best accuracy model then followed by hyper parameter tuning to increase the accuracy.

SETTING UP UI

After getting our model in desired format we will be setting up our user interface and will work on it. For UI part we are aiming to use streamlit and html/CSS.

V. EXISTING PROBLEM

It was the act of going through some of these research papers that kindled our imagination regarding the usage of machine learning models in the banking industry. As we have already seen in our abstract, the banking industry is plagued by a whole array of diverse problems and issues, where simply manual research and analysis proved woefully insufficient.

So, our approach to solving this problem was to first analyse gather all the data that we could about how this problem is being tackled right now by the various machine learning enthusiasts around the world. The result of this research is the Literature Survey that we have exhibited above.

However, we have also realized that most of the researcher work we interacted with dealt only with the most visible problems without delving deeper into the other issues that affect this industry. As a result, the majority of the academic work focused only on loan defaulters, whilst ignoring the massive defaults that banks suffer annually due to negligence shown towards smaller credit card defaulters.

Further, the papers also do not explore the broader systemic issues that can lead to financial insecurity and default in the first place, such as inadequate access to credit or economic inequality.

VI. PROPOSED SOLUTION

Machine learning is a good way to solve modern banking problems. It can help banks and companies to find out people who are very likely to not be able to pay back their loans. It can also help detect people who are doing malpractices related to banking. Banks can make much better decisions if they use these algorithms because they can evaluate very large sets of data and thus make much better decisions than human bankers. So, the financial losses of banks will be reduced.

A common way to use machine learning in banking is classification algorithms.

By classifying data into different categories, banks can develop models that can help them in predicting the true worth of a borrower and whether they should cater to such customers.

We also know that affinity grouping is another important application of machine learning in banking.

By analyzing customer's previous behavior, banks can group customers with similar interests, & financial backgrounds together.

This gives banks a chance to provide services that meet the unique needs and demands of each customer.

Profiling is also a well known application of machine learning in the banking industry. By analyzing customer's data, such algorithms can create profiles that are unique to each customer, which include their spending habits, and transaction history.

Such profiles help banks to identify customers who are more likely to default and take necessary measures to avoid such useless risks. Profiling also helps banks to create proper personalized marketing campaigns and offer targeted products and services to each customer.

These methods can allow us to not just make predictions about loan defaulters but also analyze the habits and behaviour of all customers of a banking institution, so as to provide a proper personalized banking experience to them.

VII. ARCHITECTURAL DESIGN

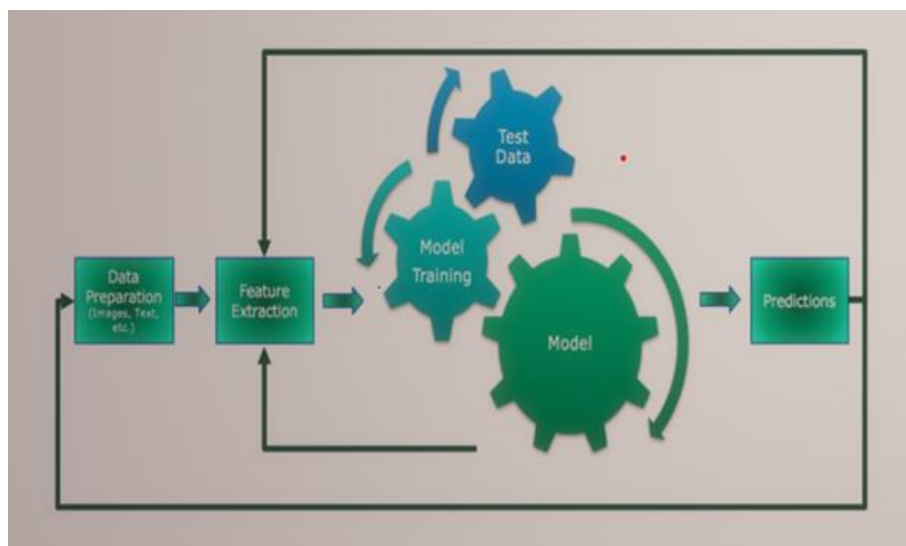


Fig 4: Architectural design

The System is design as such that first the model is created with the help of python then the UI of apps are created using streamlit and html/CSS and then these are deployed on Heroku and webpage is hosted on GitHub pages.

1. First our data is cleaned and best features are chosen.
2. Then our data is trained on the required algorithm.
3. After training and testing model is extracted.
4. The model is used in the apps for creating the ui.
5. After this the models are deployed and then connected to our main page to get all of them together.

VIII. RESULT

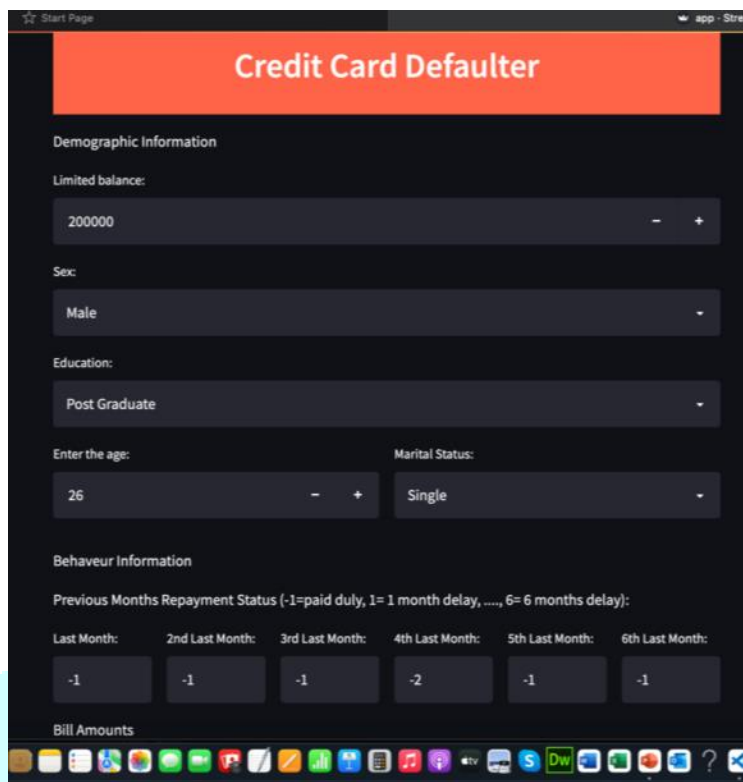


Fig 5: Credit card defaulter system final result

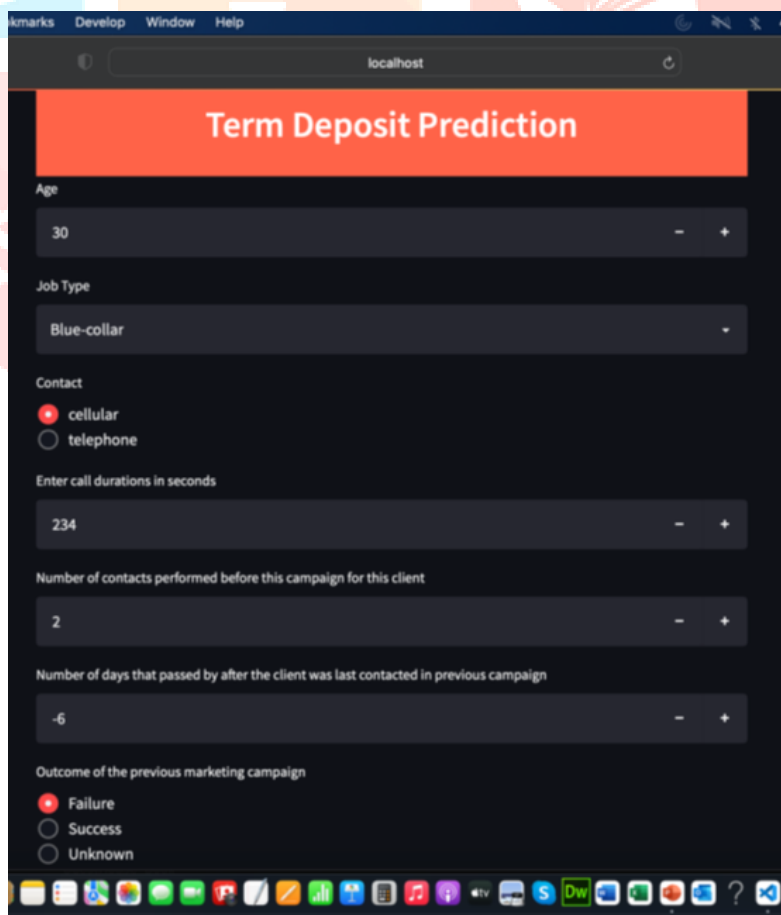


Fig 6: Term Deposit Prediction

The screenshot shows a web browser window with the URL 'localhost'. The page has a dark theme and a red header with the text 'Bank Loan Prediction'. Below the header, there are several input fields with numerical values and a 'Predict' button. The values are: Age: 34, Education: Graduate, Employed for how many years: 11.00, Debt to income ratio: 4.00, Credit debt: 2.00, and Other debt: 5.00. Below the 'Predict' button, a green message box displays the text: 'The person will not Default its our recommendation to give this person loan'.

Fig 7: Bank Loan prediction app

The screenshot shows a web browser window with the URL 'localhost'. The page has a dark theme and a red header with the text 'Customer Segmentation'. Below the header, there are several input fields with numerical values. The values are: Balance amount left in their account to make purchases: 1001000, How frequently the Balance is updated(Frequency between 0 and 1): 0.71, PURCHASES: 254, Maximum purchase amount done in one-go: 60000, Amount of purchase done in installment: 23, Cash in advance given by the user: 60000, and How frequently the Purchases are being made, score between(Frequency between 0 and 1): 0.70.

Fig 8: Customer Segmentation

IX. CONCLUSION

In our application of customer analysis, we achieved our objectives using machine learning. Our application predicts whether record applicants will be eligible for a loan, defaults to credit card payments, makes time deposits, and groups customers to be more productive when our project is completed.

The customer analysis app can assist organization in making the best judgement to increase their profits and minimize their risks. Customer analysis app will undoubtedly assist and help the banking sector to establish a much more effective system for working. Thus, this is necessary for us to create and test new strategies so that we can perform better than the performance of commonly used data mining models for these domains.

As a result, in the upcoming future, these so-called algorithms may be made much more efficient, robust, and reliable. These applications can be integrated into a system that process automatically in the upcoming future.

These systems, for now, are trained on an existing training dataset, but the algorithms may be implemented in upcoming future so that additionally tested data can also be allowed to be included in the training datasets.

IX. APPLICATIONS OF MACHINE LEARNING/ FUTURE SCOPE

Loan prediction-

The Loan Prediction System can tell if the customer will be able to repay or not thus can be used to make better decisions and minimize the risk of loss.

Term deposit predictor-

Term deposit predictor can help by letting know the chance of customer going to invest in it or not and thus saving the time and resources.

Credit card defaulter-

Credit card defaulter can be used to knowing which customer has a chance of defaulting on future payments and thus helps in taking actions accordingly.

Customer segmentation-

Customer segmentation helps determine customer relationships in each segment and maximize each customer's value to the business, thereby helping to develop better marketing strategies.

REFERENCES

- [1] Z. Feng and M. Feng, "Research on credit card scoring model based on AHP," *Finance Theory and Practice*, vol. 1, pp. 74–77, 2016.
- [2] R. Mei, Y. Xu, and G. Wang, "Study on analysis and influence factors of credit card default prediction model," *Statistics and Applications*, vol. 5, no. 3, pp. 263–275, 2016.
- [3] F. Thabtah, S. Hammoud, F. Kamalov, and A. Gonsalves, "Data imbalance in classification: experimental evaluation," *Information Sciences*, vol. 513, pp. 429–441, 2020.
- [4] Sheikh MA, Goel AK, Kumar T. An Approach for Prediction of Loan Approval using Machine Learning Algorithm. In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) 2020 Jul 2
- [5] Vaidya A. Predictive and probabilistic approach using logistic regression: application to prediction of loan approval. In 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT) 2017 Jul 3
- [6] Tejaswini J, Kavya TM, Ramya RD, Triveni PS, Maddumala VR. Accurate Loan Approval Prediction Based On Machine Learning Approach. *Journal of Engineering Science*. 2020 Apr;11(4):523-32.
- [7] Suebsing, A. and Vajiramedhin, C. (2013). Accuracy rate of predictive models in credit screening, *Applied Mathematical Sciences* 7(112): 5591–5597.
- [8] Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R. (1996). *Advances in knowledge discovery and data mining*, Vol. 21, AAAI press Menlo Park.
- [9] Sulekha Goyat. "The basis of market segmentation: a critical review of the literature. *European Journal of Business and Management* www.iiste.org. 2011. ISSN 2222-1905 (Paper) ISSN 2222-2839 (Online). Vol 3, No.9, 2011
- [10] Venkatesh, A. and Jacob, S. G. (2016). Prediction of credit-card defaulters: A comparative study on performance of classifiers, *International Journal of Computer Applications* 145(7).
- [11] Radhakrishnan, B., Shineraj, G. and Anver Muhammed, K. (2013). Application of data mining in marketing, *IJCSN International Journal of Computer Science and Network*, ISSN (Online) pp. 2277–5420.