International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

A Unified Analysis of Bank Customer using Machine Learning

R. Veeramani, P. B. Pavani Reddy, Nikhil Raj, Mehul Jain

Abstract: ***In order to discover information, virtual data evaluation involves trendy computer tools and display. The use of various digital research methodologies has become quite frequent throughout society. The aim of this proposed model is to provide a new vision of customer relationship administration through visual representation and interactive techniques. Special mentions of the investigation include that consumers have a better change on consumer truthfulness in the banking sector, specifically with the expenses towards a better direction, the bond between them becomes much firmer. Lastly, comments are presented on how to make the consumer satisfied and increase their trust. They are consistently hunting for new techniques to improve their business profits like, correct commodity marketing to best consumers, accurate use of own channels, visitation on a low level to bank branches for everyday money withdrawals and etc. A unified analysis of bank customer's helps to identify the person's judgment and decides whom to distribute the services of bank and how much credits can be offered up to a limit. It also helps the bank issuer's understand profile of their consumers.***
The model uses different algorithms (KNN, Logistic Regression XGB Classifier, Decision tree and MLP classifier) to train and test the data and then analyses the best algorithm to perform the task. As customer profiling in a bank is a highly requested service for such task, a unified analysis of the same consumer segmentation is specially customized for money minting industry.

Keywords: Credit limit, Customer profiling, Machine Learning, Segmentation.

I. INTRODUCTION

In today's cutting-edge time in financial sector, banks contain huge databases containing client's data and their account transactions of past. Therefore, the finance sectors need to separate these enormous databases into tiny groups to have the option to observe these client's actions for utilizing them in the most suitable manner to achieve profits and consumer loyalty to build benefit. To achieve this task, client division is used. Segmentation produces client profiles, which manages an account with a portrayal of their customers depending on a lot of factors. Segmentation of client is defined as characterizing the customer groups based on their behavior. These finance sectors face several challenges, such as default prediction, risk management, retention of customers. Customer segmentation for various purposes allows higher profitable capacity to be achieved and prevents any risks. Therefore, it is necessary to identify clients properly to find a fitting solution.

Machine Learning can teach computers how to learn, to teach themselves equations and concepts. It not only enhances the level of connection with existing clients, but also plays an important role in foreseeing customer behavior on the premise of a certain team of vents or trends that classify their prospective approach, and in planning to offer credit products to consumers. It takes the focus away from the client and modifications the role banking institutions perform in layout.

The techniques used for machine learning are needed to profile the actions of clients to the network.

II. RELATED WORK

It explains about exponentially increasing the adoption of virtual banks and also adopting the local banks for the virtualization. The motive behind this project is to present the theoretical figure of online banking tactics and also discover out the important aspects of visual banking strategies in gaining consumer's trust in Malaysian perspective utilizing Technology Acceptance Model (TAM), Theory of Reasoned Action (TRA) and Diffusion of Innovation (DOI) research to adapt online transaction methods to satisfy the consumers demand with the advanced progress in billing, paying methods. Conclusions from this study shows that the finance plans which is being used to test the individual variables in this research is important in making an effect on consumer trust amongst the finance sectors in this country. This is helpful for further future improvements in this virtual banking sector. [1]

The widespread use of smartphone has risen in the nation of Thailand which led many customers to perform financial transactions through their mobile devices. Moreover, a customer's loyalty plays a crucial role in developing a prolonged bond in web-based atmosphere amid financial sectors and consumers. Thus, this study was created to generate trust among the customers of Thailand. From 336 cases, a certain sample was selected particularly by using Purposive Sampling Method.

Manuscript received on March 15, 2020.
Revised Manuscript received on March 24, 2020.
Manuscript published on March 30, 2020.
* Correspondence Author
R.Veeramani*, Assistant professor of Information Technology in SRM IST, Chennai, Tamil Nadu, India. Email: gvmani.r@gmail.com
P.B. Pavani Reddy, Department of Information Technology, SRM IST, Chennai, Tamil Nadu, India. Email: pavaniireddy1411@gmail.com
Nikhil Raj, Department of Information Technology, SRM IST, Chennai, Tamil Nadu, India. Email: usernikhil001@gmail.com
Mehul Jain, Department of Information Technology, SRM IST, Chennai, Tamil Nadu, India. Email: jmehul1008@gmail.com

The research included trust, ability, good will, and safety. Multiple Linear Regression was taken as a data observation and quantitative method was acting as an application study through multiple opinions. Results were indicating that it left a negative impact of customer's loyalty to a question mark on the person's moral values. But it didn't make any difference for the "Thai Customer statistical significance at 0.05 levels". Therefore, banks must have an increased privacy of data for customers.[2]

III. SYSTEM ARCHITECTEURE

The architecture first explains about how the customer data is procured first and then will be preprocessed through tokenization and normalized for easy understanding. The data then goes by the Supervised machine algorithms that will train the data so that the machine is able to understand the process, Finally, the data will be tested through the XGBoost algorithm that bring out the predicted result.

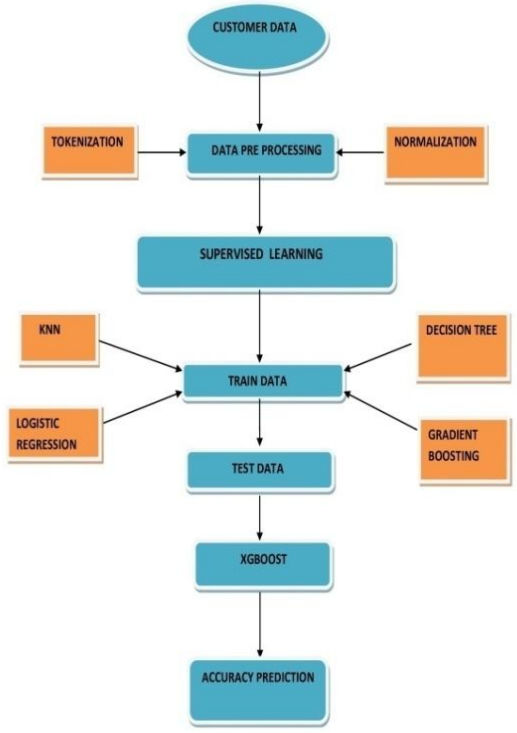

Fig.4.1 System Architecture

IV. METHODS AND IMPLEMENTATION

1: Assessment of Data

"Exploratory Data Analysis (EDA)" is a method that analyses data and gives introduction to many techniques; Maximization of data into small groups, to uncover the beneath structure, to identify primary variables, detection of anomalies and outliers and testing present assumptions. EDA is dependent on graphical representation of data and visual information. EDA represents the fluctuations which are unpredictable and the common model doesn't have this feature.

The below Table 1 represents the customer data base of 11,162 people and the sample is displayed here.

TABLE 1: Customer Data

age	job	marital	education	balance	housing	loan	contact	deposit
59	admin.	married	secondary	2343	yes	no	unknown	yes
56	admin.	married	secondary	45	no	no	unknown	yes
41	technician	married	secondary	1270	yes	no	unknown	yes
55	services	married	secondary	2476	yes	no	unknown	yes
54	admin.	married	tertiary	184	no	no	unknown	yes

The Fig.3.1 shows about contribution of loans wherein the red bar indicates the people who haven't deposited and the green bar indicates people who have deposited.

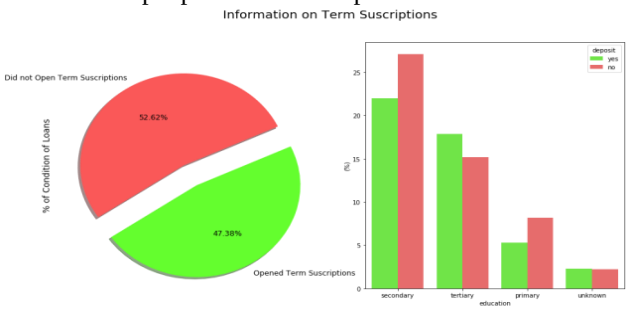

Fig.3.1 % of Contribution of Loans.

The Fig.3.2 shows the Distribution of Balance by Education on four factors; secondary, tertiary, primary and unknown in the form of boxplot graph.

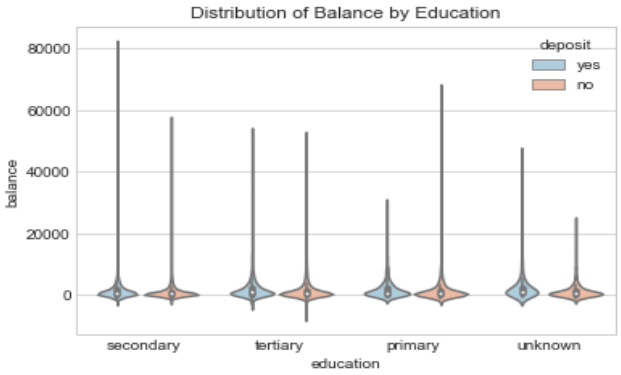

Fig.3.2 Distribution of Balance by Education.

The Fig.3.3 shows that the dots represent ages and the corresponding boxes are the dense region where maximum number of dots lie.

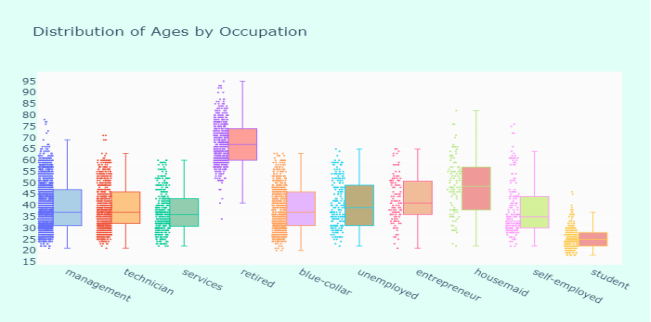

Fig.3.3 Distribution of Ages by Occupation.

The Fig.3.4 elaborates the distribution of Duration by job divided between each of the occupations based on the boxplot graph.

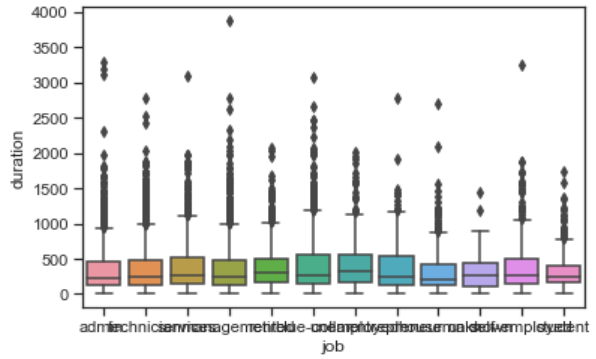

Fig.3.4 Distribution of Duration by Job

The Fig.3.5 represents age distribution by mode of contact used by customer through the box plot on 3 attributes.

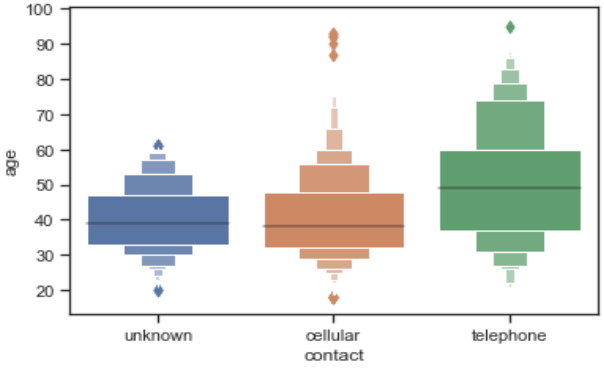

Fig.3.5 Age Distribution by Contact Medium

The Fig.3.6 displays the Distribution of Job by age through bar plot wherein the y axis represents the occupation and x-axis describes the age. The highest distribution is shown by the 'retired' category.

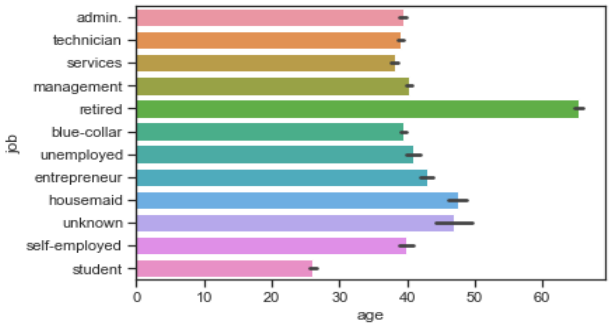

Fig.3.6 Distribution of Job by Age.

2: Text Pre-processing

Text pre-processing is a part of Natural Language Processing algorithm of Machine Learning that simplifies texts into more readable manner. There are 3 parts to this process. Tokenization refers to division of paragraphs into sentences and sentences into words or phrases and will highlight keywords in a paragraph and make them standout. Normalization converts texts into lower case formats for easy understanding. Noise removals erase any extra spaces in texts to avoid any errors.

3: Prediction

The following algorithms are used for training the data:

KNN Classifier: This algorithm classifies similarities between data points by using data (e.g. distance function) and then the data is placed to the class which has the closest neighbors around it.

Logistic Regression: It is a supervised learning algorithm that takes values of output 'y' from the input 'x'. It uses the sigmoid function and can derive the result as 0 or 1. The formula used for this function is: $f(x) = \frac{1}{1+e^{-(x)}}$

Gradient Boosting Classifier: It is a machine learning boosting algorithm which produces prediction based on decision trees. It trains a weak model and improvises it.

Decision Tree Classifier: It is a powerful tool used for used for classification and prediction the outcomes. It has a tree like structure which has a leaf node and other internal nodes which give a result based on the latter ones. The results are loosely based on either yes or no function.

MLP Classifier: It is a part of Neural network used for classification and is known as Multi-Layer Perceptron and has three layers namely; input, output and hidden layer which makes a complex model and is used in training datasets. This algorithm is used to test the data that has been trained. It also helps to predict the outcome of the data.

XGBoost: This algorithm is one of the Machine Learning boosting algorithms and is an advanced version of gradient boosting. The feature of Extreme Gradient Boosting is that it increases speed and efficiency of the model. Since the computational data is very slow to process. This algorithm is mainly used on train dataset of decision trees.

V. RESULT

In conclusion, the data was divided into 2 parts: Train data a nd Test data. 75% of data was used to train and the remainin g 25% was for testing.

Algorithm	Accuracy
KNN	55%
Logistic Regression	66%
Gradient Boosting	67%
Decision Tree	63%
MLP Classifier	70%

From all the above algorithms, XGBoost algorithm which wa s used for testing data was considered to be the most accurate with a score of 0.67218 out of 1. On this basis, the predicti on the customer details will be accurate up to 67%.

VI. CONCLUSION AND FUTURE WORK

This paper basically explains about the customer profiling at a bank's transaction level for consumers. The procedure to execute this is explained above with the implementations and the results can also be derived from the said exercise. Through these results one can easily identify how banks help profile customer's data and also fulfill its goals in real time scenario. In the current age, Indian banks have various branches at many places where only customer data is stored. In future prospects, demographic information can also be taken in consideration and improve the technology used for the same.

REFERENCE

1. Jacqueline Tham, Aye Aye Khin "A Conceptual Model of Virtual Banking Strategies in Achieving Customers "Loyalty of Malaysian Perceptive 2016.
2. Chanokrhada Wichittakul, Kriengsin Prasongsukarn "Factors Affecting The Level Of Trust In Mobile Banking: A Case Study Of Customer Perception Toward Commercial Mobile Banking Adoption" 2018.
3. R.M.Rani, Dr.M. Pushpalatha, "Generation of Frequent sensor epochs using efficient Parallel Distributed mining algorithm in large IOT", Computer Communications, Volume 148, 15 December 2019, Pages 107-114
4. R.Mythili, Revathi Venkataraman, T.SaiRaj, "An attribute-based lightweight cloud data access control using hyper graph structure", The Journal of Supercomputing(JoS) ,Published online: 02 Jan 2020 DOI: 10.1007/s11227-019-03119-7
5. S.Sivamohan, Liza.M.K, R.Veeramani, Krishnaveni.S, Jothi.B, "Data Mining Techniques for DDOS Attack in Cloud Computing", IJCTA Interntional Scoience Press, Pg: 149-156
6. S Pandiaraj, Aishwarya, Surbhi, Alisha Minj, Priyanshu Singh, "Enabling Cloud Database Security Using Third Party Auditor", International Journal of Engineering and Advanced Technology (IJEAT), Volume-8 Issue-4, April, 2019
7. R.Veeramani, Dr.R.Madhan Mohan, "Iot Based Speech Recognition Controlled Car using Arduino", International Journal of Engineering and Advanced Technology,Volume-9 Issue-1, October 2019
8. T.H. Feiroz khan, N.Noor Alleema, Narendra Yadav, Sameer Mishra, AnshumanShahi "Text Document Clustering using K-Means and Dbscan by using Machine Learning", International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249 – 8958, Volume-9 Issue-1, October 2019
9. S.Babeetha, B. Muruganantham, S. Ganesh Kumar, A. Murugan, "An enhanced kernel weighted collaborative recommended system to alleviate sparsity", International Journal of Electrical and Computer Engineering (IJECE), Volume 10, February 2020, Page No. 447-454
10. avitha.R, K.Malathi, "Recognition and Classification of Diabetic Retinopathy utilizing Digital Fundus Image with Hybrid Algorithms", October 2019,International Journal of Engineering & Advanced Technology(IJEAT), Volume 9, Issue 1, 109-122
11. T.Chandraleka, Jayaraj R, " Hand Gesture Robot Car using ADXL 335 ", International Journal of Engineering and Advanced Technology (IJEAT)', Volume-8 Issue-4, Nov 2019
12. H.Sangeetha, S.Abinayaa, "Smart Irrigation Systems using Sensors and GSM" in 'International Journal of Recent Technology and Engineering (IJRTE)', Volume-8 Issue-1, May 2019. Page No.:884-886
13. B.Sathya Bama,,Y.BevishJinila, "Attacks in Wireless sensor networks-A Research" ,International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-8, Issue- 9S2, July 2019
14. Vellingiri, J., S. Kaliraj, S. Satheeshkumar and T. Parthiban , "A Novel Approach for User Navigation Pattern Discovery and Analysis for Web Usage Mining", Journal of Computer Science 2015, vol 11 (2): Page no 372.382

AUTHORS PROFILE

R. Veeramani, Working as an Assistant Professor at Dept. of IT, SRM IST, Chennai, Tamil Nadu, India. His research interests are in Ad-hoc Network, IOT, Cloud Computing and Artificial Intelligent. In addition, he is member of Indian Science Congress Association.

P.B Pavani Reddy is currently pursuing Bachelor of Technology in Information Technology from SRM Institute of Science and Technology, Chennai, Tamil Nadu, India.

Nikhil Raj is currently pursuing Bachelor of Technology in Information Technology from SRM Institute of Science and Technology, Chennai, Tamil Nadu, India.

Mehul Jain is currently pursuing Bachelor of Technology in Information Technology from SRM Institute of Science and Technology, Chennai, Tamil Nadu, India.