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 virtual 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.

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 ARCHITECTURE

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.

![System Architecture Diagram]

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>
<thead>
<tr>
<th>TABLE 1: Customer Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
</tr>
<tr>
<td>59 admin.</td>
</tr>
<tr>
<td>56 admin.</td>
</tr>
<tr>
<td>41 technician</td>
</tr>
<tr>
<td>55 services</td>
</tr>
<tr>
<td>54 admin.</td>
</tr>
</tbody>
</table>

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.

![Contribution of Loans Graph]

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.

![Distribution of Balance by Education Graph]

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

![Distribution of Ages by Occupation Graph]

The Fig.3.4 elaborates the distribution of Duration by job divided between each of the occupations based on the boxplot graph.
The Fig. 3.4 represents age distribution by mode of contact used by customer through the box plot on 3 attributes.

The Fig. 3.5 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.

V. RESULT

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

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>KNN</td>
<td>55%</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>66%</td>
</tr>
<tr>
<td>Gradient Boosting</td>
<td>67%</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>63%</td>
</tr>
<tr>
<td>MLP Classifier</td>
<td>70%</td>
</tr>
</tbody>
</table>

From all the above algorithms, XGBoost algorithm which was used for testing data was considered to be the most accurate with a score of 0.67218 out of 1. On this basis, the prediction 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.
A Unified Analysis of Bank Customer using Machine Learning

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 into consideration and improve the technology used for the same.

REFERENCE


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.