Analysis of Finding Potential Bprrowers Using Machine Learning Algorithm

Anjali Kaithwas, Kalyani Kanoje, Poornima Ingle, Kanchan Zode
Guided By ;- Ms. Anuja Ghasad
Nagpur Institute of Technology, Nagpur, Maharashtra, India
Rashtrasant Tukdoji Maharaj Nagpur University,Nagpur, Maharashtra, India

ABSTRACT
In this paper, we use six modules, in first module, we just read the data set. In the second module, we perform visualization. In the third module, we remove all unwanted data and also remove highly co-related data. In the fourth module, we manage highly unbalanced data for using SMOTE algorithm and NEAR MISS algorithm, in the fifth module, we predict the accuracy of data for using (four algorithm) Logistics regression, decision tree, random forest, support vector machine, in the sixth module, we perform the prediction using hyper parameter tuning grid search CV by using (logistic regression, decision tree, random forest, support vector machine), after them we compare all prediction values and pick up the highest accuracy algorithm.
Keyword: Marketing, Handling data, logistic regression, support vector machine, random forest, decision tree, prediction.

I. INTRODUCTION

Banking (online and mobile) has become one of the most strategic channels used by bank customers. The organization for economic co-operation and development (OECD) has identified some of the core properties and crosscutting effects of the digital transformation [1] as the most important business challenge currently underway. Furthermore, the OECD recognizes banking as one of the sectors where such transformation is more relevant in economic, organizational, and social terms. The initial adoption of digital services could be examined using standard parametric statistical methods, examining customers digital journey is more complex. Digitalization is a challenging endeavor where several factors drive digitalization decisions. The advantages of following a machine learning approach in order to examine the bank customers digitalization process. The use of machine and causal machine algorithm in our research context allows us to reveal the process that individuals follow to make their financial digitalization choices. Unlike prior studies, we are not focused on a single dimension of the digitalization process but on several dimensions.

Lending club, the world's largest online financial platform for borrowers. It conceptual based on the internet to build a marketplace which keeps costs lower and investment return greater than traditional commercial banks. It has created a novel mechanism that make borrowing simple and easy for everyone. lending as a typical form of the internet financial application, is a service which directly connects to the loan borrowers to complete the transaction procedure through online platform with or without the intermediaries of commercial banks.

Nowadays, with the maturity of big data technology, machine learning is used in risk management controls of the financial field. Support vector machine (SVM) is widely used for prediction of stock prices and its movements. machine learning can classify available data into groups, which are then defined by rules set by analysts. When the classification is complete, the analysts can calculate the probability of a fault. Predicting whether a transaction is fraudulent or legitimate. Improve prediction systems to calculate the possibility of fault. Predictive analytics is one of the most promising examples of machine learning. Its applicable for everything; from product development to real estate pricing.

II. LITRATURE REVIEW

Oliver werth,C. Schwarzbach, Davinia roadrihuez card, 2020 deals with the Digitalization in banking sector and gives the detailed information about how digital transformation affects almost every area in societies and has consequences for incumbent companies.

Roberto, 2019 deals with the digital transformation in a banking sector and gives the detailed information about how information technology (IT) heavily impacts various industries products processes, and business models.

III. PROPOSED METHODOLOGY

In our day to day life e totally depend on machine its not only helps us but also reduce our time and effort. Now a days banking is part of our life not only transfer money in second but also provide a services in the form of apps the use fill random data in apps but this data maintain by banking. The bank have lot much of data where they manage. Use of machine learning one of the key factor to maintain data.

We use module and see how to maintain data in module 1: we read data and merge two different data in module 2: using that merge data we visualized data in this data visualization we plot graph use grid pie chart and etc. in module 3: we drop unwanted data and also remove null value from our data set we divided data train and test using person correlation we remove data which is highly correlated in module 4: we use algorithm in train and testing data set, in that we use logistic regression where we predict the output accuracy using a confusion matrix we get true positive, false positive false negative, and true negative value. We also use decision tree for visualization, random forest gave us score of accuracy of data support vector machines use for prediction and gave use accuracy of data. Using classification report we get a precision, recall, F1 score, support. In module 5: we perform handing the unbalanced data using smote algorithm and near miss algorithm.
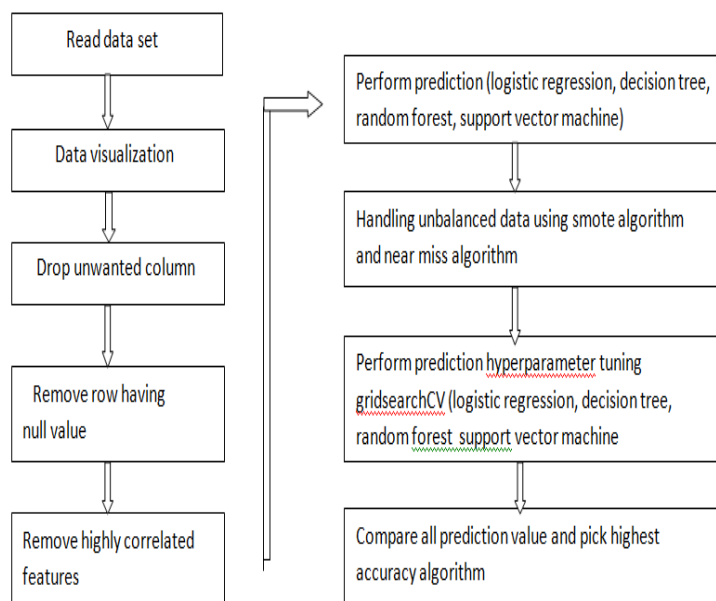
Fig :- analysis of finding potential borrowers using machine learning algorithm

Here we explain the flow diagram of analysis of finding potential borrowers using machine learning algorithms. First is read data set where we read the data which we will receive from the customer then after this stop we have to visualize the data from the data given by customer. Thirdly we will check all the data then the structure of order of the data it will delete the data which is non worth data for the real data then after it will remove the rows which are having null value. If in case of any data will be missing from any column it will automatically delete the data and the whole row and will be remove, so it will not show the ambiguous data and display the developer and convert information, if it will not remove the data then it will show the complex the process and make it hard to define the process and to be implement the project. Then it will remove the highly correlated data factor for the future data otherwise the data will be show wrong data and corrupt the data from all the data sheet.

Then it will perform the prediction with the Logistic regression, Decision tree, Random forest and Support Vector Machine in this, decision tree it will supervised learning technique that can be used for both classification and regression problems, then the logistic regression is a statistical analysis method to predict a binary outcome such as yes or no. random forest is a machine learning technique that's used to solve regression and classification problem it has highest true and false positive rates as the no. of explanatory variable increases in a dataset now the last is support vector machine is a linear model for classification and regression problem. When a number of feature are high compared to a number of data points in the dataset. Now, we jump to the next step that is handling unbalanced data using smote algorithm and near miss algorithms, now the smote

algorithms is a synthetic minority over sampling technique. It help to overcome the over fitting problem posed by random oversampling. Next is near miss algorithm is an algorithm that can help in balancing an imbalanced by randomly eliminating majority class examples. Next is near miss algorithms is an algorithm that can help in balancing an imbalanced dataset. It aims to balance class distribution by randomly eliminating majority class examples. Now at last it will compare all the prediction value and pick the highest accuracy algorithm and show the correct predicted value.
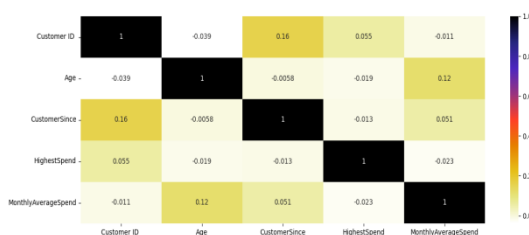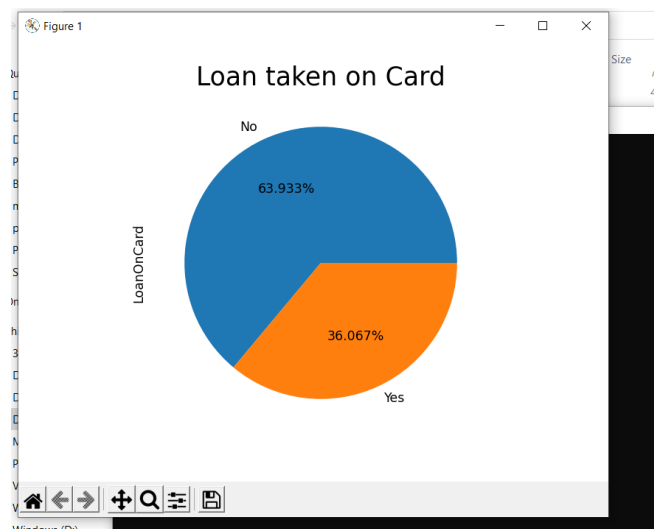
IV.   EXPERIMENTAL RESULT
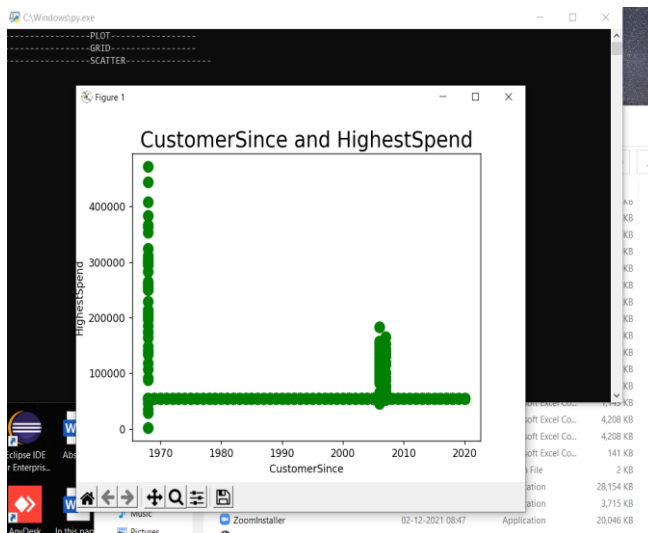
Here is the statistic for the Customer since and highest spend in the available data.
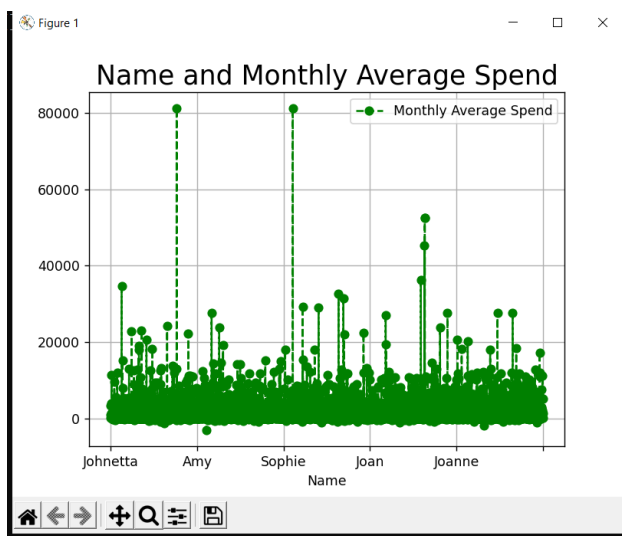
Here, the output of our project in the graph format.
In the above chart, we can see the statistical data for the Name and Monthly average spend.
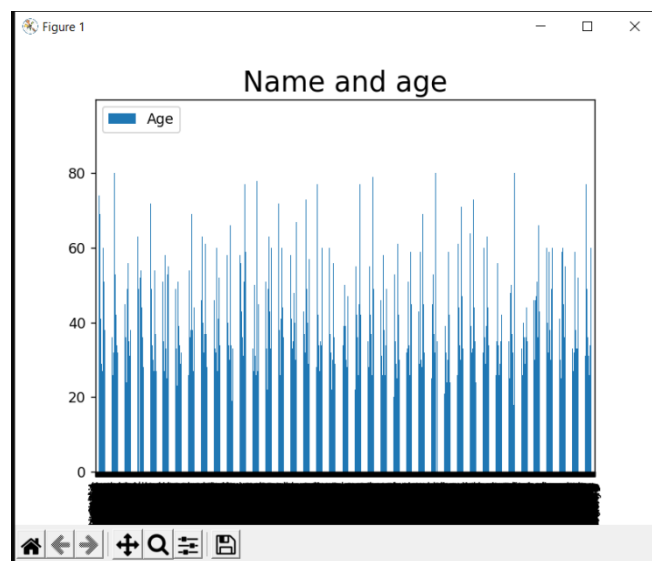
This is the statistic

This is the pie chart of our project output.

Here it has shows the percentage of the people who has applied for the loan and out of them only some of them are eligible for the loan.

Hence, Out of 100%, only 36.06% people are applicable for the loan and rest 63.93% are non applicable for the loan

This is the output of the module, it has show the data of the Name and Age

REFERENCE

[1]. Amel, D. & Rhodes, S.A. (1999). Strategic Groups in Banking. The Review of Economics and Statistics, Vol. 70, pp. 685 – 689.
[2]. Wewege, L. & Thomsett, M.C. (2020). The Digital Banking Revolution. How FinTech Companies are Transforming the Retail Banking Industry through Disruptive Financial Innovation, 3rd. Ed., De Gruyter.
[3]. Amit, R. & Zott, C. (2001). Value Creation in e-business, Strategie Management Journal, Vol. 22, No. 6/7, pp. 493 – 520.
[4]. Hadad, S., & Bratianu, C. (2019). Dematerialization of banking products and services in the digital era. Management & Marketing. Challenges for the Knowledge Society, 14(3), 318-337.
[5]. Schallmo, D. & C. A. Williams & L. Boardman (2017). Digital Transformation of Business Models Best Practice, Enablers, and Roadmap, International Journal of Innovation Management, Vol. 21, No. 8.
[6]. Wirtz, B. W. (2019). Digital Business Models. Concepts, Models, and the Alphabet Case Study, Springer International Publishing.
[7]. Porter, M.E, (2008). The Five Competitive Forces that Shape Strategy, Harvard Business Review, Jan., 86 (1), pp. 79 -83
[8]. Schallmo, D. & C. A. Williams & L. Boardman (2017). Digital Transformation of Business Models Best Practice, Enablers, and Roadmap, International Journal of Innovation Management, Vol. 21, No. 8.

[9]. EBA (2016). EBA Report on the Leverage Ratio Requirements under Article 511 of the CRR, EBA/Op/2016/13.
[10]. FSB(2019). FinTech and market structure in financial services: Market developments and potential financial stability implications.