CREDIT CARD DEFAULTER PREDICTION USING MACHINE LEARNING

K.VENKATESH , MOTHUKURI KATYAYANI

Assistant Professor MCA , DEPT , Dantuluri Narayana Raju college , Bhimavaram , AndhraPradesh
Email id:-kornalavenkatesh@gmail.com
PG Student of MCA, Dantuluri Narayana Raju College, Bhimavaram, AndhraPradesh
Email id:-katyayani.m.44@gmail.com

ABSTRACT

In our project, we mainly focussed on credit card defaulters prediction in real world. The motive is to find any credit card defaulter. In order to find whether a customer is defaulted or not the model is trained from past credit card transactions with Billing and Payment history . The models that we have used are logistic regression.

This abstract introduces the concept of credit card defaulters prediction, a critical component of the modern financial landscape. It explores the mechanisms and technologies employed to predict defaulters in credit card company. Credit card defaulter prediction leverages advanced techniques such as machine learning and artificial intelligence to proactively predict and mitigate defaulters in real-time.

This paper delves into the challenges posed by credit card companies in predicting defaulters, the benefits of implementing credit card defaulter prediction systems, and the evolving landscape of this essential field. It underscores the significance of defaulter prediction in ensuring the accuracy, security, and trustworthiness of credit cardholders for both consumers and financial institutions. As digital commerce continues to expand, the need for effective credit card defaulter prediction mechanisms becomes increasingly vital in safeguarding the integrity of financial institutions in our digital world.

The widespread adoption of credit cards in today's digital economy has revolutionized the way we conduct financial transactions. However, with the convenience and easy disbursement of credit cards comes the ever-present risk of credit card defaulters. These defaulters can range from minor re-payments customers to major defaulters.

More defaulters within the credit card companies may lead financial inconvenience and instability to cardholders and operational challenges for financial institutions and may lead to bankruptcy.

1 INTRODUCTION

Nowadays the usage of credit cards has dramatically increased. As credit card becomes the most popular mode of payment for both online as well as regular purchase, the defaulters associated with re-payment of utilized bills are also rising. In this paper, we model the credit card holders payment history and billing details using a Random Forest to show how it can be used for the prediction of Defaulters. In both algorithms is initially trained with the billing and payment pattern of a cardholder. Based on the supplied information, our system will predict whether the customer will be default or not with a prediction percentage, which can be used by the bankers to anticipate whether the customer will pay his future bill or not. Based on the supplied information, our system will predict whether the customer will be default or not with a prediction percentage, which can be used by the bankers to anticipate whether the customer will pay his future bill or not. We present detailed experimental results to show the effectiveness of our approach and compare it with other techniques.

Credit card defaulter prediction is a vital component of the modern financial landscape, aimed identifying irregularities in credit card re-repayment and thus identifying the defaulters. Predicting the defaulters can help the credit card companies to avoid customers from accidental overcharges and sending timely reminders. Left unchecked, these defaulters can lead to financial inconveniences for consumers by paying penalties and operational disruptions for businesses.

In an era dominated by digital commerce and cashless transactions, credit cards have become an indispensable part of our daily lives. These plastic cards offer convenience, flexibility, and efficiency in making payments, whether it's for online shopping, dining out, or travel bookings. However, with the widespread use of credit cards, the rise of defaulters in credit card re-payment has also surged, presenting a significant challenge for both cardholders and financial institutions..

2.LITERATURESURVEYANDRELATEDWORK

2.1Defaulter Prediction in Credit Card System Using SVM & Decision Tree

With growing advancement in the electronic commerce field, Credit card bill re-payment defaulters spreading all over the world, causing major financial losses. In the current scenario, Major cause of financial losses is credit card defaulters; they not only affects credit card companies but leads to economy disruption. Decision tree, Genetic algorithm, Metal earning strategy, neural network, HMM is the presented methods used to predict credit card defaulters. In contemplating system for defaulters prediction, artificial intelligence concept of Support Vector Machine (SVM) & decision tree is being used to solve the problem. Thus by the implementation of this hybrid approach, financial losses can be reduced to greater extent.

2.2Machine Learning Based Approach to Credit Card Defaulter Prediction

Credit card re-payment defaulter are the customers who failed to re-pay the credit card bills fully on time. Credit card defaulters is a fast-growing issue through the emergence of online KYC services. In the real world, a highly accurate process in credit card defaulters prediction is needed since surge in the defaulters may causes financial loss. Therefore, our approach proposed the overall process of predicting the defaulters based on machine learning, supervised and unsupervised method to predict defaulters and process large amounts of financial data. Moreover, our approach performed sampling process and feature selection process for fast processing with large volumes of transaction data and to achieve high accuracy in credit card defaulters prediction. F-measure and ROC curve are used to validate our proposed model.

2.3The Use of Predictive Analytics Technology to Predict Credit Card Defaulters in Canada

This research paper focuses on the creation of a scorecard from relevant evaluation criteria, features, and capabilities of predictive analytics vendor solutions currently being used to predict credit card defaulters. The scorecard provides a side-by side comparison of five credit card predictive analytics vendor solutions adopted in Canada. From the ensuing research findings, a list of credit card defaulters PAT vendor solution challenges, risks, and limitations were outlined.

2.4 BLAST-SSAHA Hybridization for Credit Card Defaulter Prediction

This paper propose to use two-stage sequence alignment in which a profile Analyzer (PA) first determines the billing and payment pattern of 30,000 customers. The payment history traced by the profile analyzer are next passed on to a deviation analyzer (DA) for possible alignment with past default behavior. The final decision about the nature of a customer is taken on the basis of the observations by these two analyzers. In order to achieve online response time for both PA and DA, we suggest a new approach for combining two sequence alignment algorithms BLAST and SSAHA

3 EXISTING SYSTEM

It is vital that credit card companies are able to identify credit card defaulters so that customers are not charged for late payment and penalties for the bills that they did not pay. The credit card defaulter prediction problem includes modeling past credit card transactions with the data of the past six months billing and payment details. This model is then used to recognize whether a customer is defaulter or not

Every credit card customer is monitored as and when required. This monitoring can be done by financial institutions, credit card companies, or third-party defaulter prediction services.

Each customer is assigned a prediction score or risk score based on rule-based checks and machine learning predictions. Customers with high defaulter scores has to be notified to the credit card companies for further decision making.

In cases where a customer is highly suspicious or has a high likelihood of being an defaulter, the credit card companies should be notified for the cardholder's review. Cardholders may receive notifications asking them to pay the outstanding amounts to avoid penalties and further legal proceeding.

Disadvantages of Existing system

Overzealous credit card defaulters prediction algorithms can occasionally flag legitimate customers as defaulter, causing inconvenience to customers

The accuracy of defaulter prediction heavily relies on the quality and accuracy of the training data, which may be affected by data inconsistencies or errors.

Some advanced machine learning algorithms lack transparency, making it challenging to understand the reasoning behind defaulter prediction decisions.

In existing System, a research about a case study involving credit card defaulter prediction, where data normalization is applied before Cluster Analysis and with results obtained from the use of Cluster Analysis and Artificial Neural Networks on defaulter prediction has shown that by clustering attributes neuronal inputs can be minimized. This research was based on unsupervised learning. Significance of this paper was to find new methods for defaulter prediction and to increase the accuracy of results

The data set for this paper is based on real life transactional data by a large European company and personal details in data is kept confidential. Accuracy of an algorithm is around 50%. Significance of this paper was to find an algorithm and to reduce the cost measure. The result obtained was by 23% and the algorithm they find was Bays minimum risk. The Clustering doesn't produce the less accuracy when compared to Regression methods in scenarios like credit card defaulter prediction. Comparatively with other algorithms k-means produce less accurate scores in prediction in this kind of scenario

4 PROPOSED WORK AND ALGORITHM

The Credit card defaulter prediction system is initiated for predicting the credit card defaulters by feeding past billing and payment details. The transactions done by credit card holders are collected from the credit card company or third parties.

Our goal is to implement machine learning model in order to classify, to the highest possible degree of accuracy, credit card defaulters from a dataset gathered from Kaggle. After initial data exploration, we knew we would implement a logistic regression model for best accuracy reports.

Logistic regression, as it was a good candidate for binary classification. Python sklearn library was used to implement the project, We used Kaggle datasets for Credit card defaulters prediction, using pandas to data frame for class ==0 for no default and class==1 for default, mat plot lib for plotting the default and no defaulter user data, train_test_split for data extraction (Split arrays or matrices into random train and test subsets) and used Logistic Regression machine learning algorithm for defaulter prediction and print predicting score according to the algorithm. Finally Confusion matrix was plotted on true and predicted.

Advantages

The results obtained by the Logistic Regression Algorithm are best compared to any other Algorithms.

Defaulters Prediction: Predicting defaulters in credit card bill re-payment helps prevent erroneous activities, protecting both customers and businesses from financial losses.

Cost Savings: Identifying defaulters promptly reduces the financial impact of charge backs and unauthorized transactions on

businesses, leading to cost savings.

Customer Trust: Accurate defaulter prediction builds trust among customers, as they are less likely to face wrongful charges and negative experiences.

Machine learning algorithms can adapt and improve over time by learning from new transaction data, enhancing their defaulter prediction capabilities.

- The results obtained by the Logistic Regression Algorithm are best compared to any other Algorithms.
- The Accuracy obtained was almost equal to cent percent which proves using of Logistic algorithm gives best results.
- The plots that were plotted according to the proper data that is processed during the implementation
- The 'amount' feature is the transaction amount. Feature 'class' is the target class for the binary classification and it takes value 1 for positive case (defaulter) and 0 for negative case (no defaulter).
- Random forest ranks the importance of variables in a regression or classification problem in a natural way can be done by Random Forest.
- This system produce the efficient output as compare to the existing Kaggle algorithms

5 METHODOLOGIES

MODULES

DATASET

This paper utilizes the dataset provided by revolution analytics for the credit card defaulters prediction from Kaggle. Dataset has 30,000 transactions with last 6 months billing and payment history. The dataset is divided as 60%, 20% and, 20% in the Train, Valid and Test set, respectively

DATA PREPROCESSING

For efficient implementation of the classification algorithm, data pre-processing is performed before feature selection. Under-sampling is performed to make the dataset balanced to avoid the biasing of the classification algorithm towards the majority class. Feature Selection is implemented on a balanced dataset.

FEATURE SELECTION

Feature selection methods are used to remove unnecessary, irrelevant, and redundant attributes from a dataset that do not contribute to the accuracy of a predictive model or which might reduce the accuracy of the model. In this paper seven feature selection techniques namely Select-K-best, Feature Importance, Extra tress classifier, Person's correlation, Mutual Information, Step forward selection and Recursive feature elimination are used.

FEATURE IMPORTANCE

Feature importance is a class of techniques for assigning scores to input features to a predictive model that indicates the relative importance of each feature at the time of making a prediction. It reduces the number of input features. In this paper, feature importance is implemented using an extra tree classifier from the decision tree. Extra Trees is similar to Random Forest, it builds multiple trees and splits nodes using random subsets of features, but unlike Random Forest, Extra Tree samples without replacement and nodes are split on random

6 RESULTS AND DISCUSSION

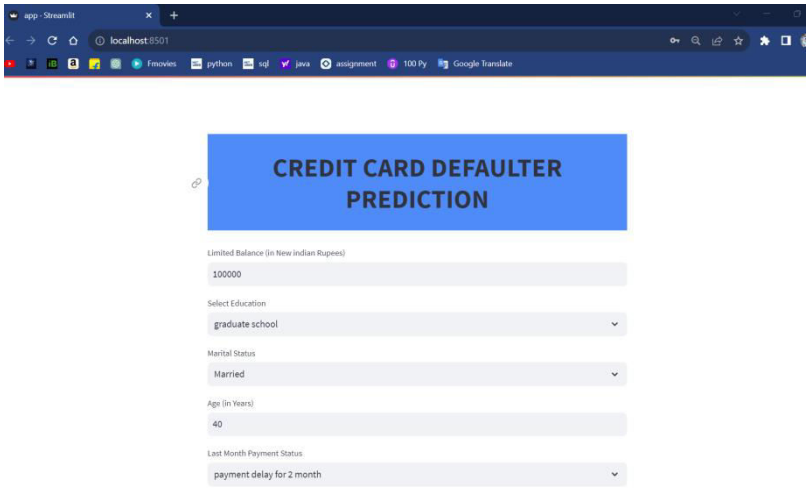

Fig1 :Credit card defaulter prediction.

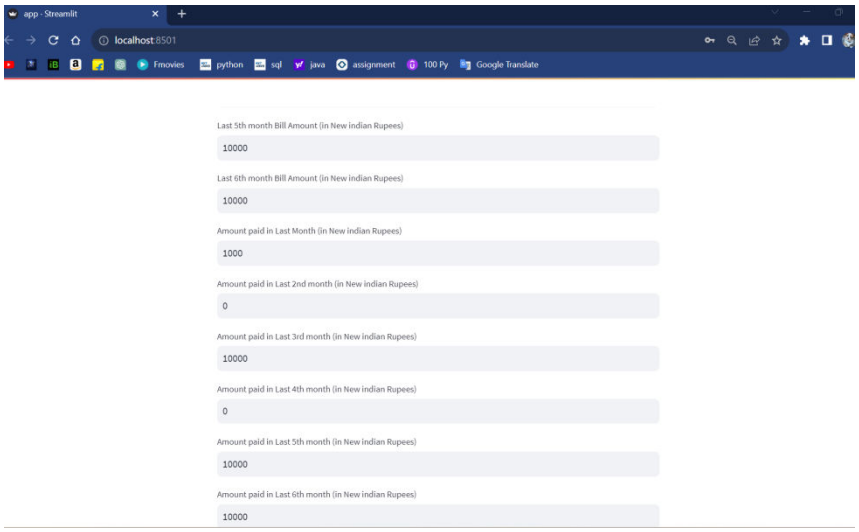

Fig 2: Submitting the form before prediction.

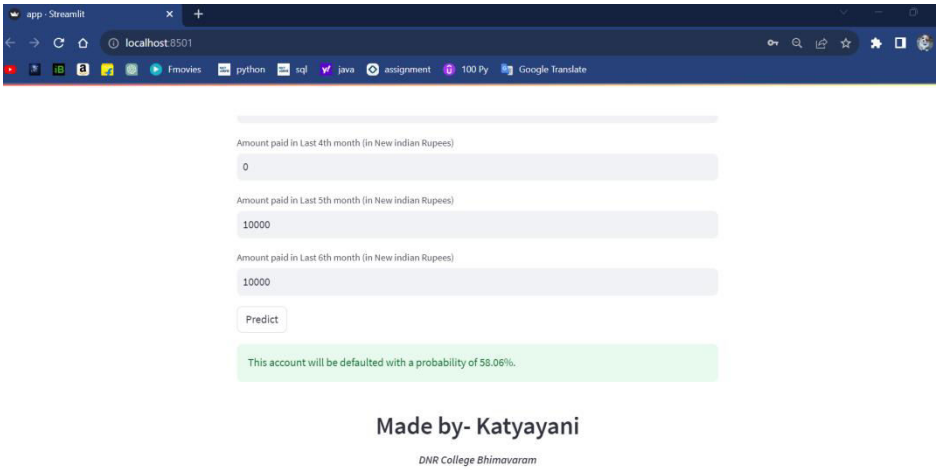

Fig 3: Predicted value

6.CONCLUSION AND FUTURE SCOPE

CONCLUSION

In conclusion, our Credit Card Defaulter Prediction System represents a significant advancement in the field of financial security and fraud prevention. This system offers users an efficient solution for identifying and the credit card defaulters and default percentage.

This machine learning defaulter prediction tutorial showed how to tackle the problem of credit card defaulter prediction using machine learning. It is fairly easy to come up with a simple model, implement it in Python and get great results for the Credit Card Defaulter Prediction task on Kaggle. Credit card defaulter prediction system using whale optimization algorithm and SMOTE (Synthetic minority optimization technique) aims in indentifying the defaulters based on the past bill re-payment transactions made by the card holder.

In this research, we have proposed a method to predict the credit card defaulters that is based on deep learning. We first compare it with machine learning algorithms such as k- Nearest Neighbor, Support vector machine etc. Finally we have used the neural network, even though tough to train the model which would fit fine to model for predicting an credit card defaulter. In our model, by using an artificial neural network (ANN) which gives accuracy approximately equal to 100% is best suited for credit card defaulter prediction .It gives accuracy more than that of the unsupervised learning algorithms

FUTURE SCOPE

- Continued advancements in machine learning will lead to more accurate and adaptive defaulters prediction models
- More KYC details and CIBIL credit score should be verified before issuing new cards.
- Our Credit Card Defaulters Prediction System is continuously improving to take. The system's future scope has several exciting avenues

From the above analysis, it is clear that many machine learning algorithms are used to predict the credit card defaulters but we can observe that the results are not satisfactory, so we would like to implement deep learning algorithms to predict credit card defaulters accurately.

7 REFERENCES

1.A. Taha, S.J. Malebary, An intelligent approach to credit card defaults prediction using an optimized light gradient boosting machine, IEEE Access 8 (2020) 25579–25587, doi:10.1109/ACCESS.2020.2971354.

2.S. Makki, Z. Assaghir, Y. Taher, R. Haque, M. Hacid, H. Zeineddine, An experimental study with imbalanced classification approaches for credit card defaulter prediction, IEEE Access 7 (2019) 93010–93022, doi:10.1109/ACCESS.2019.2927266.

3.C. Jiang, J. Song, G. Liu, L. Zheng, W. Luan, Credit card defaulters prediction: a novel approach using aggregation strategy and feedback mechanism, IEEE Internet Things J. 5 (5) (Oct. 2018) 3637–3647, doi:10.1109/JIOT.2018.2816007.

4.https://www.ijeat.org/wp-content/uploads/papers/v8i5/D6468048419.pdf

5.https://www.ijert.org/credit-card-defaulters-prediction-using-machine-learning

6.https://www.sciencedirect.com/science/article/pii/S2666285X21000066