Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2790 – 2794
Received 05 March 2021; Accepted 01 April 2021.

Loan Prediction Using Logistic Regression in Machine Learning

S. Sreesouthry[1], A. Ayubkhan[2], M. Mohamed Rizwan[3], D. Lokesh[4], K. Prithivi Raj[5]

[1]Assistant Professor, Department of ECE, Sona College of Technology, Salem, Tamil Nadu, India.
E-mail: southrysouthry@gmail.com

[2]Assistant Professor, Department of ECE, Sona College of Technology, Salem, Tamil Nadu, India.
E-mail: ayubkhan.slm@gmail.com

[3]UG Student, Electronics and Communication Engineering, Sona College of Technology, Salem,Tamil Nadu, India.
E-mail: nawzir13nov99@gmail.com

[4]UG Student, Electronics and Communication Engineering, Sona College of Technology, salem,Tamil Nadu, India.
E-mail: lokeshdharma20@gmail.com

[5]UG Student, Electronics and Communication Engineering, Sona College of Technology, Salem, Tamil Nadu, India.
E-mail: rprithivi03@gmail.com

ABSTRACT

For several problems, the banking industry still wants a more scrutinized predictivemodelling framework. For the banking industry, forecasting credit defaulters is a daunting challenge. One of the quality measures of the loan is the loan status, it doesn't show everything immediately, butit is the first step of the loan lending process. The loan status is used for creating s credit scoring model. In order to identify defaulters', end valid clients, the credit scoring model is used for reliable review of credit data. This paper's target is to build a credit scoring model for credit data. In order to develop the financial credit scoring model, various machine learning methods are used. We propose a machine learning classifier-based analysis model for credit data in this paper. We use the Min-Max normalization and Linear Regression combination. Using the program package Jupyter notebook, the target is implemented. This recommended model offers the best precision of critical details. In commercial banks, it is used to forecast the loan status using machine learning classifier.

KEYWORDS

Credit Scoring, Logistic Regression, Loan Status, LoanLending Process, Min-MaxNormalization.

Introduction

In commercial loan lending, one of the most significant topics to tackle the banking sector is the scoring of borrowers' creditworthiness. Credit risk is defined as the risk of borrowers failing to satisfy their borrowing obligations. The method of credit rating is used to predict the risk of credit and to fraudulent activities. This credit rating schemes are used to make decisions in the light of borrowers' information. Lenders want to minimise the chance of loss of each lending decision in order to make loan decisions and understand the return that compensates for the risk.
Overall, the success and loss of the banking industry is based on its credit risk.

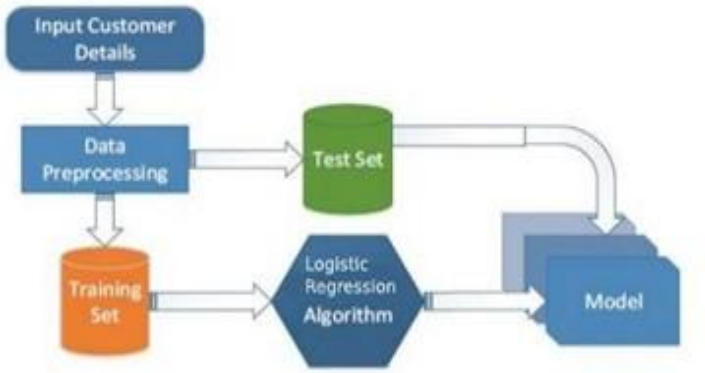
The credit balance may not be properly obtained, so the bank would forfeit it. Bank benefit is, thus, associated with their credit risk. A vital challenge and a difficult job to handle and analyse is credit risk. It is possible to split credit scoring activities into two categories, such as application scoring and behavioural scoring. Application Scoring is for the credit applicant to be graded into categories of 'good' and 'bad' risk. The role of behavioural scoring is to identify new clients based on their payment history and personal data.

Banks, housing finance companies and some NBFC companies are struggling with Loans of different kinds, such as mortgage loans, personal loans, Company loans etc. in all parts of the world. These ones, these Companies in rural, semi-urban and urban areas exist. Fields. These businesses after applying loan by customer. Validates customers' eligibility for or not to receive the loan. This, the paper offers a solution for this method to be automated by using algorithms for machine learning. So, the consumer is going to Fill up an application for an online loan.The method for discovering valuable information.

From massive databases is data mining. It consists of mining grouping, clustering and association rules. The primary

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2790 – 2794
Received 05 March 2021; Accepted 01 April 2021.

feature of the data mining process is classification. Nowadays, there are several classification strategies available.As follows, the remainder of the article is organised. The basic concepts of Machine learning, LogisticsRegression, Data Collection, Pre-Processing, Data, Featured Engineering, Conclusion.



Machine Learning

In general, machine learning is a computer science area that allows machine, the ability to learn without being programmed directly. Three key classes, such as supervised learning, unsupervised learning and semi-supervised learning, are categorised into machine learning tasks. Supervised machine learning techniques can be used with documented datasets of class labels and for unknown class label datasets, unsupervised learning techniques can be used. Semi-supervised learning is a machine learning approach that, during testing, mixes a small amount of labelled data with a significant amount of unlabelled data.

Semi-supervised learning falls between unsupervised (with no training data labelled) and supervised (with only training data labelled) learning. Huge quantities of data are available everywhere nowadays. Therefore, in order to obtain some useful information and to create an algorithm based on this analysis, it is very important to analyse this data. Via data mining and machine learning, this can be done. Machine learning is an important part of artificial intelligence and is used to create algorithms based on patterns in data and historical data relationships. In different areas, such as bioinformatics, machine learning is used. Through Machine learning and Logistic Regression, the given data is processed, and loan prediction is done.

Logistics Regression

Prediction of the bank's granting of the loan to the clients is done by using this method. Classification is used to build the model and then use Logistic Regression. For model growth, the sigmoid feature is used. Pre-processing is the model's key field in which it absorbs more time and then the study of exploratory results, which is Function Engineering and then model selection followed. Feeding the model to two different datasets, and then to that precedes the model, Logistic Regression.

Data Collection

Data was gathered from Kaggle, one of the most providers of data sources for the purpose of learning, and hence the data is collected from the Kaggle, which had two sets of details, one of which was for the preparation and the supplementary tests. The dataset for training is the model in which datasets are further divided into datasets was used to train the model train and the minor dataset.

Data Preprocessing

The technique of data mining has been used in pre-processing for transforming raw data collected in an online form into raw data useful and effective formats. There is a need for it to be converted to useful format as it may have some

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2790 – 2794
Received 05 March 2021; Accepted 01 April 2021.

irrelevant, incomplete formats noisy data and information. Data in order to deal with this issue a cleaning process was used. The data reduction approaches are used before data mining to deal with massive quantities of info. Then the interpretation of data will become it is simpler and it's going to get precise.

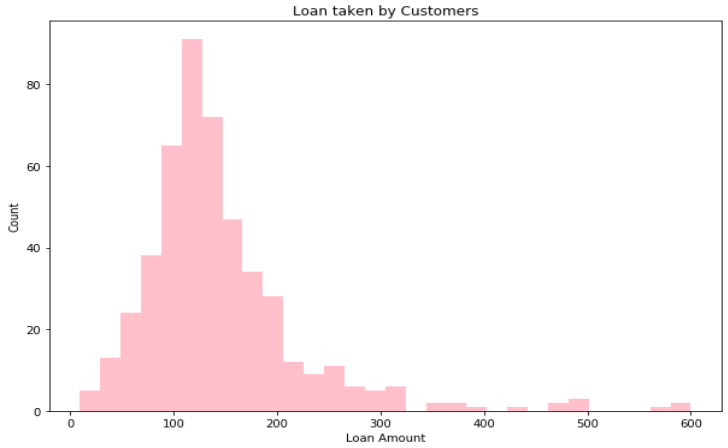
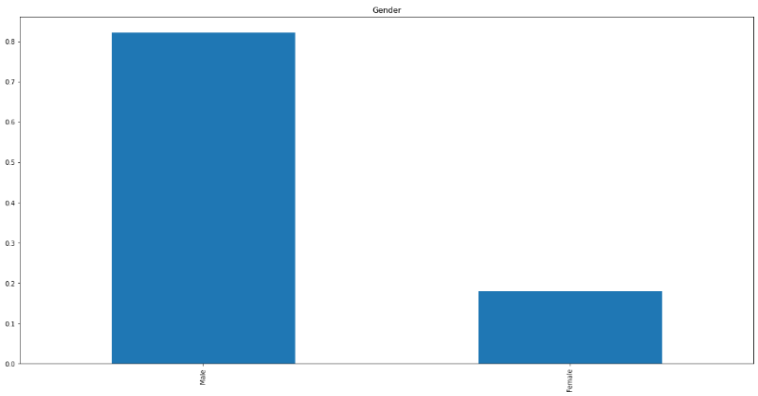
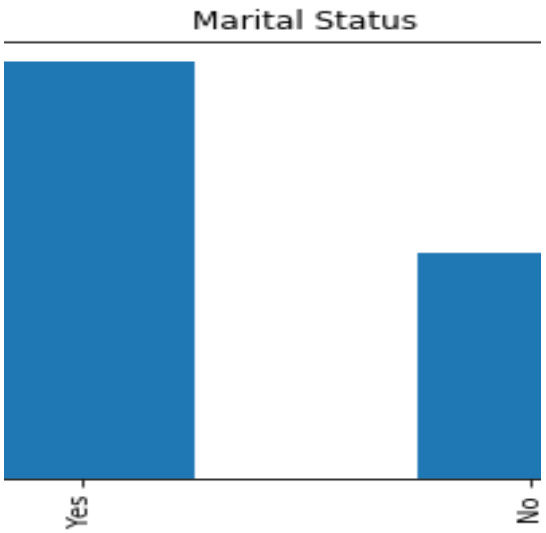


Fig. 5.1. Loan Density based on Loan Amount





Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2790 – 2794
Received 05 March 2021; Accepted 01 April 2021.

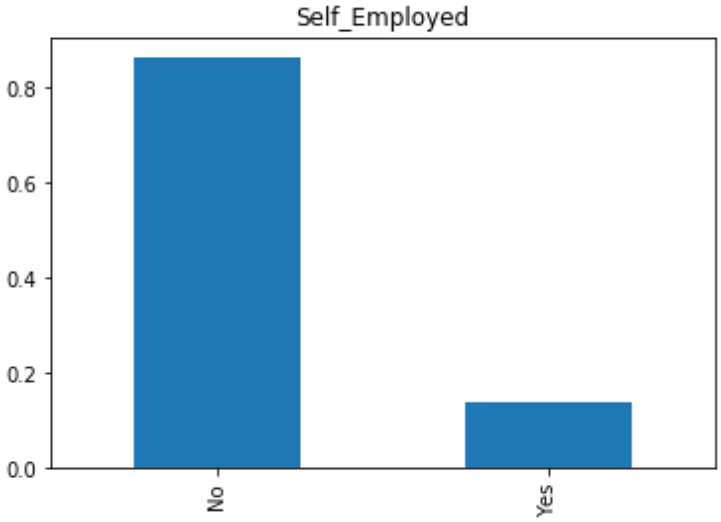


Fig. 5.2. Loan Status based on other characteristics

Data

ID=480
Gender=480
Marital Status=480
Dependents=480
Graduation=480
Self Employed=480
Income=480
Co-applicant Income=480
Loan Amount=480
Loan Term=480
Credit History=480
Location=480

Featured Engineering

A proper input dataset in feature engineering, which is as per machine learning algorithm specifications, compatible is ok, prepared. Our model library of Pandas and Numpy has been imported in order to run. So, the quality of machine learning enhances the model.

```
import pandas as pd
import numpy as np
```

Result and Conclusion

The method of forecasting begins with cleaning and Data collection, imputation of missing values, experimental information Data set analysis and then model construction to test the data set. Model and research on data from experiments. On a data set, the best-case scenario. On the initial data collection, the accuracy obtained is 0.77. After study, subsequent conclusions are drawn. Such applicants with the lowest credit score will not receive a loan. Approval, owing to a greater risk of not paying back the loan. Such applicants who have high levels, most of the time, low income and lower loan quantity requirements are more likely to get accepted, which makes sense, their borrowings. Some other characteristics, such as gender and marital status does not consider.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 2790 – 2794
Received 05 March 2021; Accepted 01 April 2021.

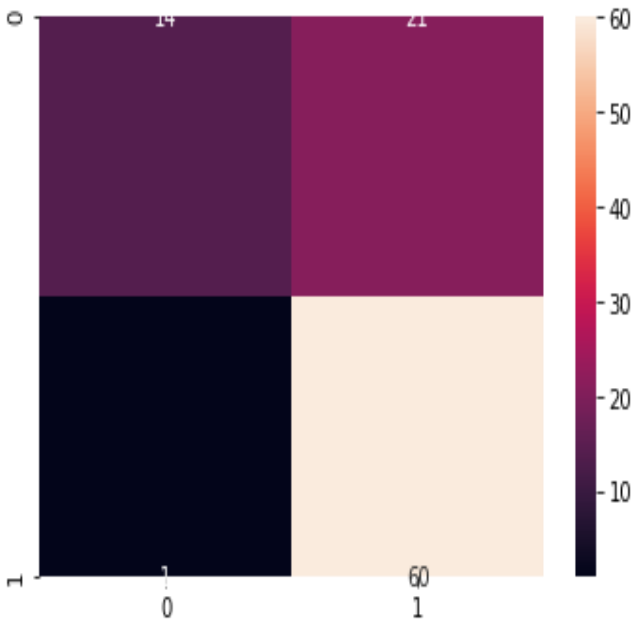


Fig. 8.1. Heat Map

References

[1] Abdelmoula, A.K. (2015). Bank credit risk analysis with k-nearest-neighbor classifier: Case of Tunisian banks. *Accounting and Management Information Systems*, *14*(1), 79-106.

[2] Arutjothi, G., &Senthamarai, C. (2017). Comparison of Feature Selection Methods for Credit Risk Assessment. *International Journal of Computer Science*, *5*(5).

[3] Arutjothi, G., & Senthamarai, C. (2017). Credit risk evaluation using hybrid feature selection method. *Software Engineering and Technology*, *9*(2), 23-26.

[4] Attig, A., & Perner, P. (2011). The Problem of Normalization and a Normalized Similarity Measure by Online Data. *Tran. CBR*, *4*(1), 3-17.

[5] Babu, R., & Satish, A.R. (2013). Improved of k-nearest neighbor techniques in credit scoring. *International Journal for Development of Computer Science & Technology*, *1*(2), 1-4.

[6] Bach, M.P., Zoroja, J., Jaković, B., & Šarlija, N. (2017). Selection of variables for credit risk data mining models: preliminary research. *In 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1367-1372.

[7] Byanjankar, A., Heikkilä, M., & Mezei, J. (2015). Predicting credit risk in peer-to-peer lending: A neural network approach. *In IEEE Symposium Series on Computational Intelligence*, 719-725.

[8] Devi, C.D., & Chezian, R.M. (2016). A relative evaluation of the performance of ensemble learning in credit scoring. *In IEEE International Conference on Advances in Computer Applications (ICACA)*, 161-165.

[9] Arutjothi, G., Senth Amarai, C. (2016). Effective Ana lysis of Financial Data using Knowledge Discovery Database. *International Journal of Computational Intelligence and Informatics*, *6*(2).

[10] Goel, H., & Singh, G. (2010). Evaluation of Expectation Maximization based Clustering Approach for Reusability Prediction of Function based Software Systems. *International Journal of Computer Applications*, *8*(13), 13-20.