IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 3, Ver. I (May-Jun. 2016), PP 79-81
www.iosrjournals.org

Loan Approval Prediction based on Machine Learning Approach

Kumar Arun, Garg Ishan, Kaur Sanmeet
(sh.arun.rana@gmail.com , CSED, Thapar University, India)
(ishangarg9292@gmail.com, CSED, Thapar University, India)
(sanmeetkbhatia@gmail.com,CSED,Thapar University,India)

Abstract: *With the enhancement in the banking sector lots of people are applying for bank loans but the bank has its limited assets which it has to grant to limited people only, so finding out to whom the loan can be granted which will be a safer option for the bank is a typical process. So in this paper we try to reduce this risk factor behind selecting the safe person so as to save lots of bank efforts and assets. This is done by mining the Big Data of the previous records of the people to whom the loan was granted before and on the basis of these records/experiences the machine was trained using the machine learning model which give the most accurate result. The main objective of this paper is to predict whether assigning the loan to particular person will be safe or not. This paper is divided into four sections (i)Data Collection (ii) Comparison of machine learning models on collected data (iii) Training of system on most promising model (iv) Testing*

Keywords: *Loan, Machine Learning, Training, Testing, Prediction.*

I. INTRODUCTION

Distribution of the loans is the **core business part** of almost every banks. The main portion the bank's assets is directly came from the profit earned from the loans distributed by the banks. The prime objective in banking environment is to invest their assets in safe hands where it is. Today many banks/financial companies approves loan after a regress process of verification and validation but still there is no surety whether the chosen applicant is the deserving right applicant out of all applicants. Through this system we can predict whether that particular applicant is safe or not and the whole process of validation of features is automated by machine learning technique. The disadvantage of this model is that it emphasize different weights to each factor but in real life sometime loan can be approved on the basis of single strong factor only, which is not possible through this system.

Loan Prediction is very helpful for employee of banks as well as for the applicant also. The aim of this Paper is to provide quick, immediate and easy way to choose the deserving applicants. It can provide special advantages to the bank. The Loan Prediction System can can automatically calculate the weight of each features taking part in loan processing and on new test data same features are processed with respect to their associated weight .A time limit can be set for the applicant to check whether his/her loan can be sanctioned or not. Loan Prediction System allows jumping to specific application so that it can be check on priority basis. This Paper is exclusively for the managing authority of Bank/finance company, whole process of prediction is done privately no stakeholders would be able to alter the processing. Result against particular Loan Id can be send to various department of banks so that they can take appropriate action on application. This helps all others department to carried out other formalities.

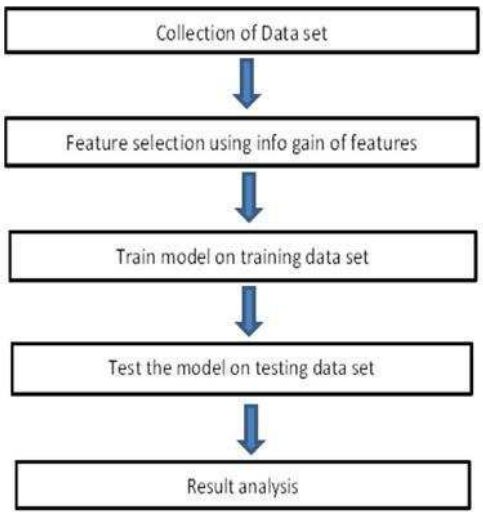
II. Data Set

The training data set is now supplied to machine learning model, on the basis of this data set the model is trained. Every new applicant details filled at the time of application form acts as a test data set. After the operation of testing, model predict whether the new applicant is a fit case for approval of the loan or not based upon the inference it conclude on the basis of the training data sets.

Variable Name	Description	Type
Loan_ID	Unique Loan ID	Integer
Gender	Male/ Female	Character
Maritial_Status	Applicant married (Y/N)	Character

Variable Name	Description	Type
Dependents	Number of dependents	Integer
Education_Qualification	Graduate/ Under Graduate	String
Self_Imployed	Self Imployed (Y/N)	Character
Applicant_Income	Applicant income	Integer
Co_Applicant_Income	Coapplicant income	Integer
Loan_Amount	Loan amount in thousands	Integer
Loan_Amount_Term	Term of loan in months	Integer
Credit_History	credit history meets guidelines	Integer
Property_Area	Urban/ Semi Urban/ Rural	String
Loan_Status	Loan Approved(Y/N)	Character

2.1 Loan Prediction Methodology



2.2 MACHINE LEARNING METHODS:

Six machine learning classification models have been used for prediction of android applications .The models are available in R open source software. R is licensed under GNU GPL. The brief details of each model is described below.

2.2.1 Decision Trees (C5.0):

The basic algorithm of decision tree [7] requires all attributes or features should be discretized. Feature selection is based on greatest information gain of features. The knowledge depicted in decision tree can represented in the form of IF-THEN rules. This model is an extension of C4.5 classification algorithms described by Quinlan.

2.2.2 Random Forest (RF):

Random forests [8] are a group learning system for characterization (and relapse) that work by building a large number of Decision trees at preparing time and yielding the class that is the mode of the classes yield by individual trees.

2.2.3 Support Vector Machine (SVM):

Support vector machines are administered learning models that uses association r learning algorithm which analyze features and identified pattern knowledge, utilized for application classification. SVM can productively perform a regression utilizing the kernel trick, verifiably mapping their inputs into high-dimensional feature spaces [9].

2.2.4 Linear Models (LM):

The Linear Model [10] is numerically indistinguishable to a various regression analysis yet burdens its suitability for both different qualitative and numerous quantitative variables.

2.2.5 Neural Network (Nnet):

Neural networks [14] are non-linear statistical data modeling tools. They are usually used to model complex relationships between inputs and outputs, to find patterns in data, or to capture the statistical structure in an unknown joint probability distribution between observed variables.

2.2.6 Adaboost (ADB):

Adaboost short for " Adaptive Boosting ". It is delicate to noisy information data and outliers. It is different from neural systems and SVM because Adaboost preparing methodology chooses just those peculiarities known to enhance the divining power of the model, decreasing dimensionality and conceivably enhancing execution time as potentially features don't have to be processed.[14]

III. Parameter setting for machine learning models

Model	Parameter Setting
Decision Trees	Min Split = 20, Max Depth = 30, Min Bucket = 7
Random Forest	Number of tree = 500,Number of variables=8
Support Vector Machine	Kernel Radial Basis
Linear Model	Multinomial
Ada boost	Min Split = 20, Max Depth = 30, Number of tree = 50
Neural network	Hidden layer nodes=10

Loan Prediction

IV. Conclusion

From a proper analysis of positive points and constraints on the component, it can be safely concluded that the product is a highly efficient component. This application is working properly and meeting to all Banker requirements. This component can be easily plugged in many other systems.

There have been numbers cases of computer glitches, errors in content and most important weight of features is fixed in automated prediction system, So in the near future the so –called software could be made more secure, reliable and dynamic weight adjustment .In near future this module of prediction can be integrate with the module of automated processing system. the system is trained on old training dataset in future software can be made such that new testing date should also take part in training data after some fix time.

References

[1]. Rattle data mining tool: available from http://rattle.togaware.com/rattle-download.html

[2]. Aafer Y, Du W &Yin H 2013, DroidAPIMiner: 'Mining API-Level Features for Robust Malware Detection in Android', in: Security and privacy in Communication Networks Springer, pp 86-103 .

[3]. Ekta Gandotra, Divya Bansal, Sanjeev Sofat 2014**, '**Malware Analysis and Classification: A Survey'available from http:// www.scirp.org/journal/jis

[4]. K. Hanumantha Rao, G. Srinivas, A. Damodhar, M. Vikas Krishna: Implementation of Anomaly Detection Technique Using Machine Learning Algorithms: Internatinal Journal of Computer Science and Telecommunications (Volume2, Issue3, June 2011).

[5]. *J. R. Quinlan. Induction of Decision Tree. Machine Learning, Vol. 1, No. 1. pp. 81-106., 1086.*

[6]. *Mean Decrease Accuracy https://dinsdalelab.sdsu.edu/metag.stats/code/randomforest.hml*

[7]. *J.R. Quinlan. Induction* of decision trees. MachinelearningSpringer, 1(1):81–106, 1086.

[8]. Andy Liaw and Matthew Wiener. Classification and Regression by randomForest. R News(http://CRAN.R-project.org/doc/Rnews/), 2(3):9–22, 2002.

[9]. S.S. Keerthi and E.G. Gilbert. Convergence of a generalizeSMO algorithm for SVM classifier design. Machine Learning, Springer, 46(1):351–360, 2002.

[10]. J.M. Chambers. Computational methods for data analysis. Applied Statistics, Wiley, 1(2):1–10, 1077.