A Comparative Analysis of Feature Selection for Loan Prediction Model

Karthikeyan S.M.
B.E., M.Tech
Asst. Prof., Dept. of CS&E
Adichunchanagiri Institute of Technology
Chikmagalur-577101, Karnataka, India

Pushpa Ravikumar, PhD
B.E., M.Tech
Professor & Head, Dept. of CS&E
Chikmagalur-577101, Karnataka, India

ABSTRACT
Enhancement in the banking region very huge customers are applying for different types of loans which is available in the all bank. But the bank has its own boundary assets which grant the permission for limited people. Loan approval is a very long and important step in bank organization. Banking sector need more precise predicting model for better accuracy. Predicting the credit customer is the very difficult task in bank sector. The predicting system should approve and rejects the loan application system. Loans are the core business for banks. Customer dataset is taken for identifying the key customer. The data mining technique are used for predicting the loans which containing high dimensional data. It contains some redundant and inappropriate attributes in the dataset. Machine learning techniques helps to predicting outcomes from huge amount of data. In this methodology it helps to focus on attributes and feature selection for identifying loans approval customer. In this proposed work two machine learning algorithms, Random Forest (RF) and Boruta Algorthim are applied to predict the key customer of the loan approval. This experimental result concludes that accuracy of Boruta Algorthim is better as compared to Random Forest algorithm. The social network analysis technique is also used to predict and to identify the key customer for further loan analysis.

General Terms
Loan Prediction, Machine Learning

Keywords
Feature Selection, Random Forest, Boruta, Social Network Analysis.

1. INTRODUCTION
Loan Prediction is extremely useful for representative of banks just as for the customer moreover. The point of this proposed work is to give speedy, quick and simple approach to pick the key customer. It can give unique focal points to the bank. The Loan Prediction system can consequently compute the heaviness of each underline participating in advance preparing and on new test information same highlights are handled as for their related weight[1]. A period breaking point can be set for the customer to check whether his/her credit can be endorsed or not. Loan Prediction System permits leaping to explicit application with the goal that it very well may be make sure on need premise. This proposed work is only for the overseeing authority of Bank/account organization, whole cycle of forecast is done secretly no banker would have the option to modify the pre-processing. Resulting against the specific loan id can be ship to different branch of their banks with the aim of a proper move on a application. There are many loans available In a bank[1]. Figure 1 shows the process of a loan approval.

- Secured Loans
- Unsecured Loans
- Home Loans
- Property Loans
- Non confirming Loans

Bank credit risk evaluation is generally utilized at banks the world over. As credit hazard assessment is exceptionally vital, an assortment of procedures are utilized for hazard level figuring. Also, credit risk is one of the fundamental elements of the banking network [2].

Dispersion of a loans is the center business part of pretty much each banks. The guideline segment the bank's asset is clearly came from the advantage secured from the credits scattered by the banks. The prime target in financial climate is to contribute their resources in safe hands where it is. Today numerous banks/monetary organization special treatment credit after a relapse cycle of check and approval yet at the same time there is no guarantee whether the picked customer is the meriting right customer out of all customer[2]. Through this proposed work we can foresee whether that specific customer is protected and the entire cycle of approval of highlights is robotized.

Data mining is an exceptionally energetic and important zone of research with the main aim of acquiring a lot and set of information gathered [3]. In the current time data mining is main stream in a banking area in light of the fact that there are proficient investigation techniques for distinguishing obscure and helpful data in banking information. Due to enormous information accessible the principle center is around information base positioning and assurance to settle on key choices.

Social network investigation oversees grouping of enormous proportion of clients they are related by a lot of specific associations. Social network attributes have its multi-social and over and over change organizations. Normally, social community is altered when congregation of people get together and structure some sort of new connection between one another through social associations[4]. Since bundle of the investigation techniques notable in the field of data mining are static in nature with that the data about the hour of collaborations happen isn't thought of. During the arrangement of such network, issues can happens for example essentials that concentrate numerous connections extremely confined people or optional customers from the organization customers are the main connection between two unique gatherings, range of customers in separated focuses. This can mess interchanges up which thus will make casualties the main component that moves through social network

information [5].

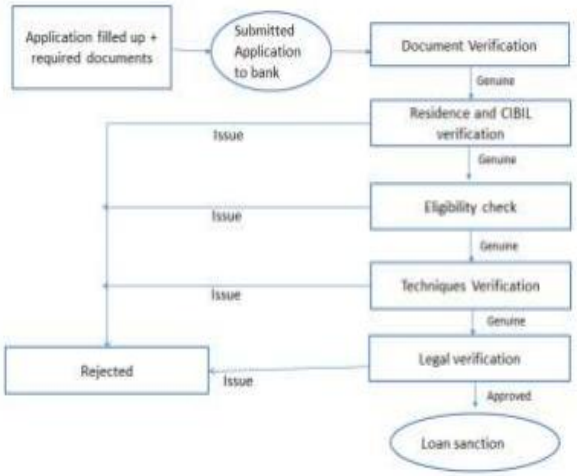


Fig 1: Process of Loan approval

2. FEATURE SELECTION APPROACH

Feature Selection is a significant part of replica structure which each expert must learn. The subsequent stage in the wake of setting up the dataset is removing the necessary highlights which are possible pointers of customer advance endorsement status[6]. For the most part include designing requires space information to distinguish the substantial arrangement of customers own and derived attributes. Specifically to advance predict there are numerous class of qualities including 1)customer individual details, 2) customer present history, 3) Financial appraisal, 4) past and current credit subtlety. There are a few ascribes under these classifications impact the advance forecast in some manner. Brute force strategy is a standard technique for doing exploratory examination of highlights and choosing the top K-attributes which are predominant for predicting the loan. Insights concerning detail highlights utilized in this exploration are clarified in this section.

All things considered, it helps in building predictive models liberated from corresponded factors, predispositions and undesirable commotion. A few factors are regularly discovered to be associated and upset accomplishing higher model precision.

2.1 Filter approach

This approach of feature selection highlight choice uses single factor examination method. It analyzes the prescient intensity of every individual variable individually. The informational collection containing bigger number of qualities should utilize channel approach rather than beast power subset draws near [7]. The prescient variable determination relies upon the general qualities of the preparation informational collection. The chose factors are autonomous of different models. Contrasted with covering approach, this methodology is quicker, computationally straightforward and adaptable. This strategy utilizes numerical assessment work like Addition Proportion, relationship based procedures, chi-Square, data pick up and so forth that are situated in inside attributes of preparing informational index. Since numerical assessment work is utilized we can know precisely why a given trait is chosen or not chose. Figure 2 shows the filtering approach for loan data [8].

Fig 2: Filtering approach for loan dataset

2.2 Wrapper method

Wrapper techniques look through the ideal subset of highlights by utilizing indicator or prepared calculations. This methodology utilize a various blends or subsets of properties from the dataset and finds the best subset of highlights. When using covering approach the customer essentially should consider (1) giving space of all potential blends of subsets, and (2) How to assess the indicator execution[9]. The best subset choice calculation can be arbitrary timberland, forward or in reverse determination and so forth the covering approach thinks about the calculation as a discovery, which takes care of all the attributes on the double and the calculation restores the subset of significant at-accolades[10]. The calculation considers the collaboration between the at-recognitions it gives the aftereffect of subset by thinking about the connection between the properties.

2.3 Embedded method

This approach takes the upsides of both the covering and channel approach. This methodology distinguishes the best highlights by utilizing trait subset and the presentation of the model itself[11]. In this line of attack the subset highlight determination and the prescient model structure can't be isolated. The usually utilized implanted element determination is regularization technique. It performs investigation by measurable methodology [12]. Regular approach presents extra requirements for expanding execution of the expectation. Instances of regularization calculations are the lasso, elastic net.

3. RELATED WORK

S. Vimala, K.C [1] describes the prediction of a loan risk model by the combination of a naïve bayes and support vector machine. By using classification techniques the naïve bayes is simple and robust is stated and also it uses the probability theory of classification.

X. Francis Jency [2] proposed the nature of a loan applications. From the best possible examination it is arranged for long haul credit and momentary advance was portrayed by lion's share of the diagrams. The predictive model helps to analyze the different constraint of a loan.

Pidikiti Supriya [3] utilized the analysis process is done for data cleaning and preprocessing. The missing value and the prediction model is built for loan data. The application gives

the information about the credit for passing the loan.

Kumar Arun [4] proposed that their application works properly and meet their bank requirement. Their system can be easily connected to any other systems. The prediction can be integrated with the automated processing system.

Bamshad et [5] has proposed another calculation subject to web use mining called Profile Assortments. In that calculation gathering is done on data base concerning relative sort of exchanges furthermore online visit packaging is applied to imagine the close to pages in each exchange.

Yoon Ho Cho [6] has utilized decision tree enlistment procedure, collusion rule mining figuring and information warehousing types of progress to manage the issue of sparsity and flexibility in organization arranged detaching system. Due to that new crossbreed philosophy has improved the practicality of the organization orchestrated disengaging technique by utilizing web use mining. Creators have utilized web logs as a data base to locate the reformist models utilizing apriori figuring. Other than to accumulate the customers producer utilize decision tree choice strategy.

Olfa Nasraoui [7] proposed downy surmise thinking system on smart web proposition structure. They have eliminated the customer profile using used web use mining and moreover they apply packing method for social occasion the customer information on customer data base. For gathering they have used different leveled independent packing method. Besides, for proposition they have used Soft assessment thinking methods.

Magdalini [8] have proposed a grouping strategy to convey improved and energetic proposals to the end client. For gathering they have used semantically comprehensible packs. Similarly for suggestion they use region mysticism which relies upon the expressions eliminated from the web substance.

Baoyao Zhou [9] have utilized continuous model burrowing procedure for envisioning the accompanying site page. In second step they utilized model base separating technique, which stores the consecutive web get to plans, and additionally strong for customer design arranging and proposition rules.

4. PROPOSED METHODOLOGY

4.1 Dataset

Data collection is the route toward social event and assessing information on centered variables in a set up proficient procedure which is unordered and saved in extraordinary arrangements having irregularity with lacking and invalid.

Attribute	Description
Job	Occupation of the Applicant
age	Age of the Applicant
Income	Monthly Income of the Applicant
Education	Education Qualification of the Applicant
Marital Status	Marital Status of the Applicant
Existing Loan	Whether the Applicant have an existing EMI or not.

Fig 3: Loan dataset of the customer with different attributes and their description

The customer loan data of different banks is considered for investigation. The credit information comprises of 15000 customer data on different sorts of loans. In figure 3 shows it incorporates customer subtleties and earlier history of customer exchange n different purposes.

4.2 Data preprocessing

Data preprocessing can frequently have a huge impact on assumption execution of a administered ML calculation. The end of commotion occasions is one of the most troublesome issues in inductive ML. Typically the eliminated examples have unreasonably veering off occasions that have an excessive number of invalid component esteem. These unreasonably veering off highlights are additionally alluded to as exceptions.

Furthermore, a typical way to deal with get used to the infeasibility of gaining from enormous informational collections is to choose a solitary example from the huge informational collection. Missing information taking care of is another issue regularly managed in the information planning steps. Figure 4 shows the methodology for a proposed work.

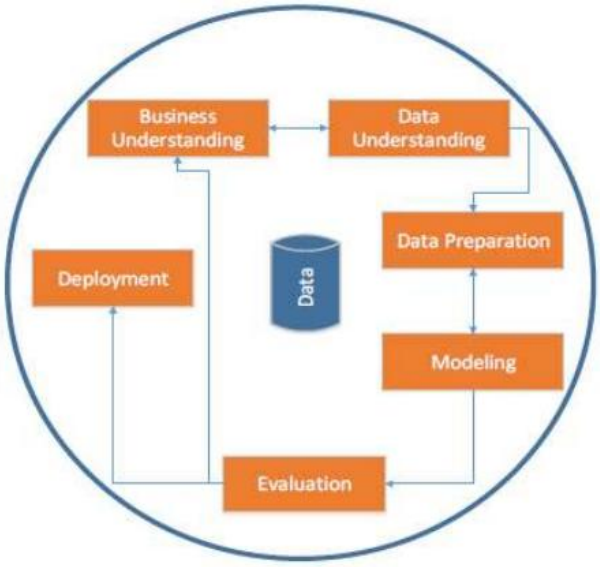


Fig 4: Methodology for proposed work

The preprocessing steps carried out for the data set are.

1. Data cleaning process

The data cleaning process is applied to loan data

- Attributes with practically invalid qualities were erased.
- Attributes with interesting qualities or just one value were erased.
- Unstructured characteristics with long free content were eliminated at this stage.
- Empty columns were taken out.
- "advance status" trait was utilized to remove the "advance status" class quality, and "credit prerequisite" property that shows if the borrower meets credit strategy necessities.
- "emp length" trait was changed from downright qualities (1 year) to mathematical qualities (1), 1 was appointed to "< 1 year" esteem as records have comparable conduct of "1 year" in importance to the class property and 0 was allocated to "n/a".

4.3 Feature selection

Feature selection is a significant part of mock-up structure which each expert must learn. The subsequent stage in the wake of setting up the dataset is extricating the necessary highlights which are possible pointers of client advance endorsement status. By and large, include designing requires area information to distinguish the legitimate arrangement of customer own and determined traits. Specifically to advance forecast, which are numerous classification of qualities including 1) customer individual subtleties, 2) customer present history, 3) financial assessment, 4) past and current advance subtleties. There are a few ascribes under these classifications impact the advance expectation in some manner. Animal power strategy is a standard technique for doing exploratory investigation of highlights and choosing the top K-ascribes which are predominant for the credit forecast. Insights regarding rundown of highlights utilized in this exploration are clarified in this part.

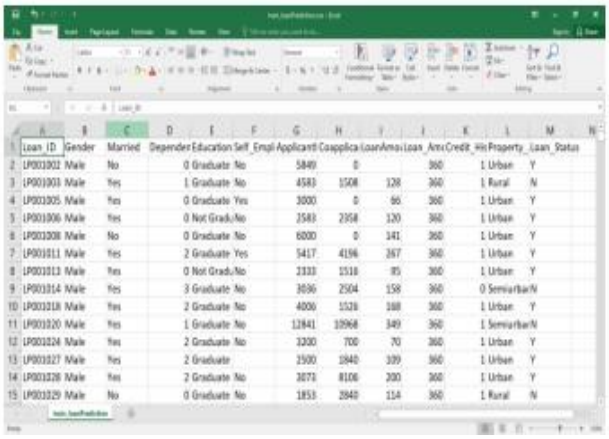
All things considered, it helps in building prescient models liberated from corresponded factors, predispositions and undesirable commotion. A few factors are regularly discovered to be connected and obstruct accomplishing higher model exactness.

Algorithm

1. Loan dataset D with p labeled class or variables where v={a1 ,a2 ,….ap}
2. Variables selection search – set j=1 select a distinct sub- set of variables Sj where 1<= Sj <=p.
1. Induce learning algorithm.
2. Evaluate the resulting model.

Selected attributes.

5. RESULTS



Loan_ID	Gender	Married	Dependents	Education	Self_Empl	Applicantl	Coapplica	LoanAmou	Loan_Amo	Credit_Hi	Property_	Loan_Status
LP001002	Male	No	0	Graduate	No	5849	0		360	1	Urban	Y
LP001003	Male	Yes	1	Graduate	No	4583	1508	128	360	1	Rural	N
LP001005	Male	Yes	0	Graduate	Yes	3000	0	66	360	1	Urban	Y
LP001006	Male	Yes	0	Not Gradu	No	2583	2358	120	360	1	Urban	Y
LP001008	Male	No	0	Graduate	No	6000	0	141	360	1	Urban	Y
LP001011	Male	Yes	2	Graduate	Yes	5417	4196	267	360	1	Urban	Y
LP001013	Male	Yes	0	Not Gradu	No	2333	1516	95	360	1	Urban	Y
LP001014	Male	Yes	3	Graduate	No	3036	2504	158	360	0	Semiurbar	N
LP001018	Male	Yes	2	Graduate	No	4006	1526	168	360	1	Urban	Y
LP001020	Male	Yes	1	Graduate	No	12841	10968	349	360	1	Semiurbar	N
LP001024	Male	Yes	2	Graduate	No	3200	700	70	360	1	Urban	Y
LP001027	Male	Yes	2	Graduate		2500	1840	109	360	1	Urban	Y
LP001028	Male	Yes	2	Graduate	No	3073	8106	200	360	1	Urban	Y
LP001029	Male	No	0	Graduate	No	1853	2840	114	360	1	Rural	N

Fig 5: Loan dataset of customer

Fig 5 shows the details of a customer. The attributes are

1. Loan_id
2. cid
3. age
4. location
5. gender
6. salaried
7. self_employed
8. dependent
9. cibil_score
10. net_monthly_income
11. work_experience_salaried
12. business_stability_self_employed
13. cureent_emi
14. total_emi_amount
15. businessman
16. annual_profit
17. annual_turover
18. no_years_itr_available
19. loan_amount_needed
20. finalized_property
21. property_market_value
22. property_budget
23. emi
24. tenure
25. rate_of_interest
26. loan_ammount_eligible

5.1 Random forest

Random forest is a flexible AI strategy fit for performing both relapse and order errands. It likewise embraces dimensional decrease techniques, treats missing qualities, anomaly esteem and other fundamental strides of information investigation, and does a genuinely great job. It is a sort of gathering learning strategy, where a gathering of feeble models consolidate to frame an amazing model. Arbitrary Backwoods develops numerous trees instead of a solitary tree in automobile model. To aggregate another article reliant on attributes, each tree provides a request and we express the tree "votes" for that class. The boondocks picks the portrayal having the most votes (over all the trees in the forest) and if there should be an event of backslide, it takes the typical of yields by different trees. Fig 6 shows the distinctive inc hub immaculateness hubs.

	IncNodePurity
loanid	3.8858132
cid	3.7649051
age	4.5438948
location	0.6252998
gender	1.1449178
salaried	0.4350325
self_employed	0.3407812
dependent	1.2331777
cibil_score	3.1613638
net_monthly_income	11.6698513
work_experience_salaried	2.3767038
business_stability_self_employed	3.8347265
cureent_emi	0.7346660
total_emi_amount	1.5802220
businessman	0.3706532
annual_profit	1.6595523
annual_tur0ver	1.7556627
no_years_itr_available	1.7681630
loan_amount_needed	4.5137179
finalized_property	0.1432651
property_market_value	1.6813556
property_budget	0.7216566
emi	10.5613767
tenure	2.0860951
rate_of_interset	0.0000000
loan_ammount_eligible	13.8666869

Fig 6 : Different inc_node_purity values

Fig 7 gives the loan status of a customer of a individual. When the customer apply for loan the algorthim gives an credit history of a customer.

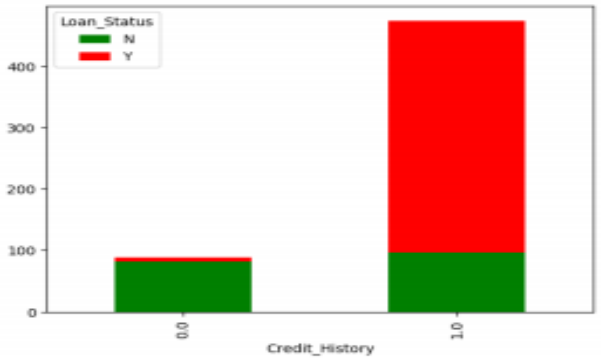


Fig 7: Loan status of the credit history

The variable significant plot shows in preview 4 is gained by building up a couple of trees, by then we can use clear limits importance(loan data) which addresses the mean development in center perfection. It very well may be seen that recursive element disposal calculation has chosen credited, cid, age, salaried, cibil score, net month to month pay loan amount needed, advance sum qualified and so forth as the significant component among the 46 highlights in the dataset.

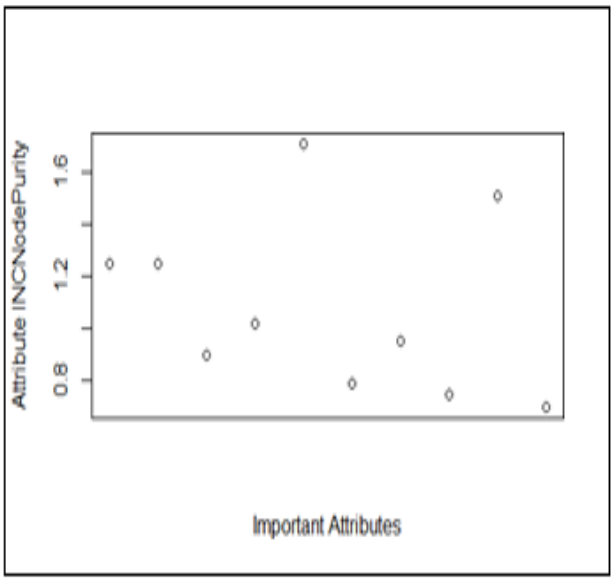


Fig 8: Importance of attribute graph

5.2 Bourta algorthim

The figure 9 shows the variable criticalness result for the media transmission dataset D. The computation performed 99 cycles for 31 exceptional credits, with that 21 attributes is changed as huge qualities, 7 attributes are contingent, 3 credits are superfluous.

```
Boruta performed 99 iterations in 56.69529 secs.
 21 attributes confirmed important: age, annual_inc_coapplicant,
annual_profit, annual_turnover, business_stability_self_employed and 16
more;
 3 attributes confirmed unimportant: finalized_property, location,
max_loan_property_value;
 7 tentative attributes left: applicant_type, businessman, dependent,
max_emi_income_value, no_years_itr_available and 2 more;
```

Fig 9: Output of variable selection using Boruta

Boruta algorthim returns three distinct variables for credits they are affirmed, dismissed and speculative, which is the eventual outcome of highlight choice. The figure 6 shows conditional and affirmed ascribes for the credit dataset D. Traits that are significantly better then the speculative qualities are viewed as affirmed. The provisional qualities have significance thusly near affirmed ascribes yet the calculation not ready to take choice with the default number of boruta calculation run. Figure 10 shows the tentative and confirmed attributes

```
Console ~/
max_loan_property_value            1.9487288  1.909340 -0.05
max_emi_income_value               2.6552977  2.654248 -0.24
emi                               19.6459814 19.562586 17.57
tenure                             7.0242975  7.085707  4.34
loan_ammount_eligible             21.3704236 21.359257 19.59
                                    normHits  decision
loanid                            0.98989899 Confirmed
cid                               0.97979798 Confirmed
age                               0.87878788 Confirmed
location                          0.02020202  Rejected
gender                            0.87878788 Confirmed
salaried                          0.67676768 Confirmed
self_employed                     0.55555556 Tentative
dependent                         0.55555556 Tentative
cibil_score                       0.98989899 Confirmed
net_monthly_income                1.00000000 Confirmed
work_experience_salaried          1.00000000 Confirmed
business_stability_self_employed  1.00000000 Confirmed
cureent_emi                       1.00000000 Confirmed
total_emi_amount                  1.00000000 Confirmed
professional                      0.58585859 Tentative
businessman                       0.59595960 Tentative
gross_annual_receipts             0.53535354 Tentative
annual_profit                     1.00000000 Confirmed
annual_turnover                   1.00000000 Confirmed
no_years_itr_available            0.05050505  Rejected
loan_amount_needed                1.00000000 Confirmed
applicant_type                    0.68686869 Confirmed
annual_inc_coapplicant            0.97979798 Confirmed
finalized_property                0.00000000  Rejected
property_market_value             0.91919192 Confirmed
property_budget                   0.35353535 Tentative
max_loan_property_value           0.11111111  Rejected
max_emi_income_value              0.44444444 Tentative
emi                               1.00000000 Confirmed
tenure                            1.00000000 Confirmed
loan_ammount_eligible             1.00000000 Confirmed
```

Fig 10: Tentative and confirmed attributes

Figure 11 shows the marital status of a customer in a loan dataset. This helps to predict the status of the particular customer in red color which is showing yes is married and blue is unmarried.

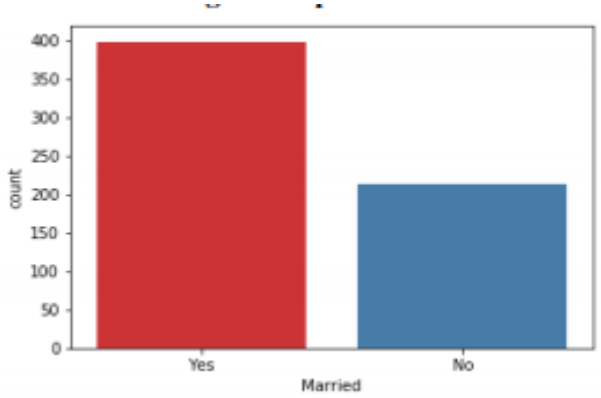


Fig 11 : Marital status of a customer in load dataset

The figure 12 shows the characteristic significance structure of the credit forecast dataset. The diagram speaks to the crate plot which demonstrates the significance of every factor in the telecom dataset. The different tones in the diagram speak to the enormity of the factors. Red box plots speak to insignificant scores of the traits. The yellow box plot speaks to the normal score of the quality. The green box plots compares to the scores of con-solidified properties in the given informational collection individually. As per variable significance chart, by utilizing covering technique ascribes came out as basic factors among the absolute of 31 qualities of media transmission dataset. These 21 ascribes are considered to have colossal impact on tribute forecast model.

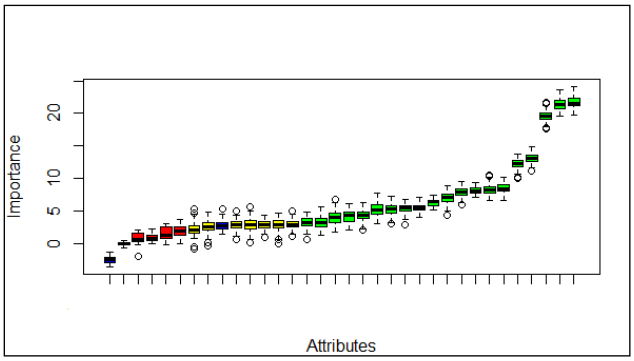


Fig 12: Box plot graph of attribute importance

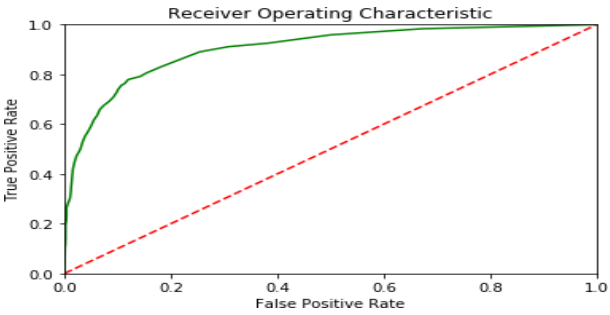


Fig 13: Accuracy graph for preprocessed data for loan

Figure 13 shows the exactness computation for pre-prepared information. Precision for the fore seeing the buying conduct of a customer by a disarray grid with the exactness 91.36. ROC twists are normally used in equal request to consider the yield of a classifier. The x-pivot indicates the bogus positive rate with the worth 0.1 to 1.0 and Y-hub determines the genuine positive rate.

Figure 14 shows the exactness computation for highlight separated information. Precision for fore seeing the buying conduct of a client by a disarray grid with the exactness 92.18. ROC twists are usually used in equal request to examine the yield of a classifier. The x-pivot determines the bogus positive rate with the worth 0.1 to 1.0 and Y-hub indicates the genuine positive rate. As a result it tends to be resolved that the exactness between the two models nearly stays as before.

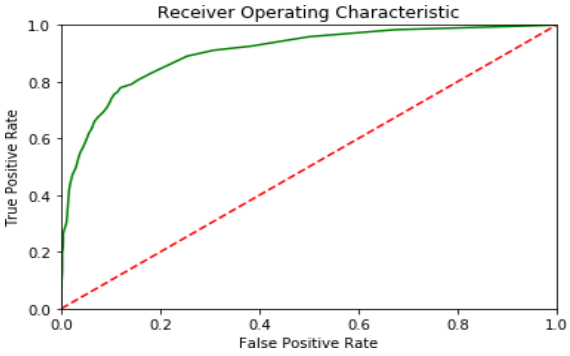


Fig 14 : Feature extraction graph for loan data

6. CONCULSION

The proposed work is focused on data mining techniques and r package is used to perform the comparative study on feature extraction of a loan dataset. The selection edge exertion is done based on different features classification. The R tool gives the visualization with the help of different graphs. The result includes age, net_monthly_income, loan

_amount_needed,, loan_eligible_amount etc as the most important attributes. The result is taken from the attribute selection approaches it can be further used with domain intelligence to obtain additional specific attributes which helps to predict efficiently. The final attribute selection can be used to build the social loan network for identifying the key customer. The proposed framework is useful for significant attribute selection and to build the accurate loan prediction model.

7. ACKNOWLEDGMENTS

I am highly thankful to Dr. Pushpa Ravikumar B.E.,M.Tech.,Ph.D.,LMISTE, Professor & Head of the department CS&E, Adichunchanagiri Institute of Technology, Chikkamagaluru-577102 for her insightful guidance and support all through the paper. At last I would like to thank all those who directly or indirectly assisted in the successful completion of the paper.

8. REFERENCES

[1] S. Vimala, K.C. Sharmili, ―Prediction of Loan Risk using NB and Support Vector Machine‖, International Conference on Advancements in Computing Technologies (ICACT 2018), vol. 4, no. 2, pp. 110-113, 2018.

[2] X. Francis Jency, V.P.Sumathi, Janani Shiva Sri, ―An Exploratory Data Analysis for Loan Prediction Based on Nature of the Clients‖, International Journal of Recent Technology and Engineering (IJRTE), Vol. 7, No. 48, pp. 176-179, 2018.

[3] Pidikiti Supriya, Myneedi Pavani, Nagarapu Saisushma, Namburi Vimala Kumari, K. Vikas, ―Loan Prediction by using Machine Learning Models‖, International Journal of Engineering and Techniques, Vol. 5, Issue 2, pp. 144-148, Mar-Apr 2019.

[4] Kumar Arun, Garg Ishan, Kaur Sanmeet, ―Loan Approval Prediction based on Machine Learning Approach‖, IOSR Journal of Computer Engineering (IOSR-JCE), Vol. 18, Issue 3, pp. 79-81, Ver. I (May-Jun. 2016).

[5] Bamshad Mobasher, Honghua Dai, Tao Luo, Miki Nakagawa. 2002. Improving the Effectiveness of Collaborative Filtering on Anonymous Web Usage Data.

[6] Yoon Ho Cho, Jae Kyeong Kim, Soung Hie Kim. 2002. A personalized recommender system based on web usage mining and decision tree induction. Expert Systems with Applications 23, 329–342.

[7] Olfa Nasraoui and Chris Petenes. 2003. An Intelligent Web Recommendation Engine Based on Fuzzy Approximate Reasoning. Proceedings of the IEEE International Conference on Fuzzy Systems - Special Track on Fuzzy Logic and the Internet.

[8] Magdalini Eirinaki, Charalampos Lampos, Stratos Paulakis, Michalis Vazirgiannis. 2004. Web Personalization Integrating Content Semantics and Navigational Patterns. WIDM'04, November 12-13. Washington, DC, USA. Copyright 2004 ACM 1-58113-978-0/04/0011.

[9] Feng Hsu Wanga, Hsiu-Mei Shao. 2004. Effective personalized recommendation based on time-framed navigation clustering and association mining. Expert Systems with Applications. 27, 365–377.

[10] Baoyao Zhou, Siu Cheung Hui and Kuiyu Chang. 2004. An Intelligent Recommender System using Sequential Web Access Patterns. Cybernetics and Intelligent Systems. IEEE Conference on Cybernetics and Intelligent Systems.

[11] Sumathi, C., P., Padmaja Valli, R., and Santhanam, T. Automatic Recommendation of Web Pages in Web Usage Mining. (IJCSE) International Journal on Computer Science and Engineering 02(09),3046-3052.

[12] Haibo Liu, Hongjie Xing, Fang Zhang. 2012. Web Personalized Recommendation Algorithm Incorporated with User Interest Change. Journal of Computational Information Systems 8(4), 1383-1390.

[13] Florent Garcin, Christos Dimitrakakis, Boi Faltings,2013. Personalized New Recommendation with Context Trees. ACM Journal.