BALTIC JOURNAL OF LAW & POLITICS
A Journal of Vytautas Magnus University
VOLUME 15, NUMBER 4 (2022)
ISSN 2029-0454

Cite: *Baltic Journal of Law & Politics* 15:4 (2022): 241-251
DOI: 10.2478/bjlp-2022-004025

Higher Accuracy on Loan Eligibility Prediction using Random Forest Algorithm over Decision Tree Algorithm

Narra Rahul Kumar
Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.

L. Rama Parvathy
Project Guide, Corresponding Author, Department of Data Science, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.

Received: August 8, 2022; reviews: 2; accepted: November 29, 2022.

Abstract

Aim: The main goal of this research is to create a proficient prediction of Supervised Machine learning algorithms for checking whether a person is eligible to get loan approval or not. **Material and Methods:** Random forest algorithm and Decision Tree algorithm are the two groups of algorithms that are applied in this study. The paper consists of around 981 rows and 13 columns used to train and test the Machine learning models to verify the loan status of persons. The experimental research had performed with N=10 iterations for each algorithm by taking a G-power of 80%. **Results:** The outcomes of this experimental research mean accuracy is over 89.94% in the Random forest algorithm and 86.69% in the Decision Tree algorithm. After performing the statistical analysis, independent sample tests show that the significant difference between the two algorithms is $p = 0.024$ where $p < 0.05$. It shows that the Random forest algorithm and Decision Tree algorithm are more stable. **Conclusion:** This research work aims to implement the innovative approach of skillful Machine learning algorithms for innovative loan eligibility prediction and also to improve accuracy in existing algorithms like the Random Forest algorithm and Decision Tree algorithm. By comparing all the analyses of experimental results. It is clearly shown that the Random forest algorithm has the highest accuracy over the Decision Tree algorithm for loan eligibility prediction.

Keywords

Decision Tree algorithm, Innovative Loan Eligibility, Machine learning, Random forest algorithm, Statistical analysis, Supervised Learning.

INTRODUCTION

The main purpose of this research is to develop a Supervised machine learning model with high accuracy to predict whether a borrower will repay the loan to the banks, financial institutions or not within the given loan term period (Aslam et al. 2019). In the present-days, one of the main reasons for the country's economy depends on the banks. The primary business of the banks is lending, and the profits of banks are dependent on the interest on Loans. It is not possible to check every customer's data and predict whether the customer is eligible for loan approval or not (Arutjothi and Senthamarai 2017). The Machine learning model reduces the task for the banks by predicting whether a customer

is eligible for a loan or not with less amount of time (Rath, Das, and Acharya 2021). It is a time-saving process for both the banks and also for the customers to check their loan status. Machine learning (ML) techniques are very useful in predicting outcomes for large amounts of data within a short period of time (Sheikh, Goel, and Kumar 2020).

There are around 2,280 articles published in Google Scholar and 433 articles in Science direct related to Loan approval prediction using Machine learning techniques. Among the articles and publications, the refereed papers are A study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning ("Study on a Prediction of P2P Network Loan Default Based on the Machine Learning LightGBM and XGboost Algorithms according to Different High Dimensional Data Cleaning" 2018)Xiaojun et al. 2018("Study on a Prediction of P2P Network Loan Default Based on the Machine Learning LightGBM and XGboost Algorithms according to Different High Dimensional Data Cleaning" 2018) was cited 111 times. An empirical comparison of Machine learning methods on bank client credit assessments ("Website," n.d.) was cited 58 times. Credit Risk Analysis Using Machine and Deep Learning Models cited 54 times (Addo, Guegan, and Hassani 2018). Integration of unsupervised and supervised machine learning algorithms for credit risk assessment cited 23 times ("Integration of Unsupervised and Supervised Machine Learning Algorithms for Credit Risk Assessment" 2019)WangBao, NingLianju, and KongYue 2019("Integration of Unsupervised and Supervised Machine Learning Algorithms for Credit Risk Assessment" 2019). From all the above paper's prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning was the best study ("Study on a Prediction of P2P Network Loan Default Based on the Machine Learning LightGBM and XGboost Algorithms according to Different High Dimensional Data Cleaning" 2018)Xiaojun et al. 2018("Study on a Prediction of P2P Network Loan Default Based on the Machine Learning LightGBM and XGboost Algorithms according to Different High Dimensional Data Cleaning" 2018) as the authors were clearly explained about the Supervised machine learning techniques and data cleaning approaches for loan prediction machine learning model.

Previously our team has a rich experience in working on various research projects across multiple disciplines (Venu and Appavu 2021; Gudipaneni et al. 2020; Sivasamy, Venugopal, and Espinoza-González 2020; Sathish et al. 2020; Reddy et al. 2020; Sathish and Karthick 2020; Benin et al. 2020; Nalini, Selvaraj, and Kumar 2020).Getting fewer accuracy values than the satisfactory results for the Machine learning models as referred to in the previous papers makes it work on this innovative Loan eligibility prediction paper. It takes more time to organize and process data in a correct manner and also to improve accuracy in the Random forest and the Decision Tree. Our team in the department has much experience in research on Machine learning models, so it's helpful to come up with an innovative idea in machine learning approaches for developing efficient algorithms with higher accuracy in the innovative loan eligibility prediction. The aim is to improve more accuracy in the Machine learning algorithms and to show that the Random forest algorithm has higher accuracy over the Decision tree algorithm.

MATERIALS AND METHODS

This study for loan eligibility prediction was done in the Software engineering laboratory, Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences. The datasets for this research study were downloaded from Kaggle which is an open-source platform for the vast number of datasets for the research and their study. There were two groups of samples considered in the project study. Group-1 belongs to the Random Forest algorithm and Group-2 Belongs to the Decision Tree algorithm. The same set of sample sizes has two iterations. Iteration-1 for the loan approved customers and Iteration-2 for the loan not approved customers. The sample size for each group was calculated by using previous study results in clinical.com by keeping g power as 80 %, threshold 0.05 and confidence interval as 95%(Yu and Pan 2016; Liu et al. 2019). The

type I error is a part of alpha error which is taken as 0.05 and it shows the testing procedure and the difference between the two algorithms of the Random Forest algorithm and the Decision Tree algorithm with an enrollment ratio is approximately I.

Random Forest Algorithm

Random Forest is one of the most important algorithms in Supervised Machine learning. It can be used for both classification and regression purposes. The algorithms made with high dimensionality can be capable of handling large datasets. The Random forest algorithm has more no of trees which helps prevent overfitting the model. It can handle missing values easily. Random forests are very flexible and possess higher accuracy values. A Random forest is a predictive tool, not a descriptive tool. Normalization is not required as it uses a rule-based approach. The regression problems can be solved using Mean Square Error(MSE) (1) and classification problems can be solved using the Gini Index formula (2) used to decide how many nodes are on the decision tree branch.
MSE Formula:

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (fi - yi)^2 \qquad (1)$$

From the above formula, The N is the number of data points. *fi* is the value returned by the model and Yi is the actual value of data point *i*.
Gini Index Formula:

$$Gin = 1 - \sum_{i=1}^{c} (p_i)^2 \qquad (2)$$

In the above formula, *Pi* represents the Relative frequency of the class you are observing in the dataset and *C* represents the number of classes.

Pseudocode for Random Forest Algorithm

Input: Training dataset
Output: Classifier accuracy
A training set S:= (x1, y1), . . . ,(xn, yn), features F, and number of trees in forest B.

```
function RandomForest(S , F)
        H ← ∅ 3
        for i ∈ 1, . . . , B do
                S (i) ← A bootstrap sample from S
                hi ← RandomizedTreeLearn(S (i) , F)
                H ← H ∪ {hi}
        end for
        return H
end function
function RandomizedTreeLearn(S , F)
        At each node:
                f ← very small subset of F
                Split on best feature in f
        return The learned tree
end function
```

Decision Tree Algorithm

Decision algorithm is one of the most widely used supervised machine learning algorithms. It can be used for both classification and regression problems but the decision tree algorithm is usually used to solve the classification difficulties. The test results of the decision tree are performed based on the features found in the given data set. The decision tree always starts with the root node and ends with the decisions made by leaves. The output of the decision tree always executes either yes or no. The decision tree algorithm always consists of Root Nodes, Decision Nodes, and Terminal Nodes. Decision trees learn from data to approximate a sine curve with a set of if-then-else decision rules. The deeper the tree, the more complex the decision rules and the fitter the model. It is known as a

choice tree in light of the fact that, like a tree, it begins with the root hub, which develops further branches and builds a tree-like design.

Decision Tree Algorithm Equation:

Entropy:

$$E(S)=\sum_{i=1}^{c} - P_i log_2 p_i \quad (3)$$

From the above equation (3)

we say that S -> current state

pi->probability of an event i of state S or Percentage of class i in a node of State S.

Information Gain:

$$\text{Information Gain} = \text{Entropy(before)} - \sum_{j=1}^{K} Entropy(j, after) \quad (4)$$

From the equation (4) the word "before" is the dataset before the split, k is the number of subsets generated by the split, and (j, after) is subset j after the split.

Pseudocode for Decision tree

Input: Training dataset

Output: Classifier predicted Accuracy

An attribute- valued dataset DT

Read & import dataset
Replace missing values
Preprocess the dataset
split it to train and test
Import the data into DT algorithm
Tree1={}
If DT is "pure" OR other stopping criteria had met then
    terminate
endif
for all attribute, b ε DT do
    Compute information-theoretic criteria if we split on b
end for
b_{best} = Best attribute according to above-computed criteria
Tree = Create a decision node that tests b_{best} the root.
DT_v= Induced sub-datasets from DT based on b_{best}
for all D_v do
    $Tree_v$= C4.5(DT_v)
    Attach $Tree_v$ to the corresponding branch of Tree
end for
return Tree

The experimental procedure is completed with the help of the Jupyter Notebook which is a tool present in Anaconda distribution. The study for loan eligibility prediction is carried out with a Hardware configuration of Windows 11 OS, i5 10th Generation, X64 bit processor CPU, and 512GB SSD drive with 8GB RAM. The different packages and libraries like Numpy, Pandas, Matplotlib, Sklearn, and Seaborn exist with the Python programming language. These packages are used in this Random forest and Decision Tree algorithms. Once the data set is cleaned and filled with the missed values the dataset is split into both train data and test data. Then data was imported into the machine learning classifier and built a machine learning model for the Group-1 Random forest algorithm and Group-2 Decision Tree algorithms. The machine learning model is finally trained and evaluated and found accurate results with the help of their metrics with different test sizes. After performing the same process with 10 sample test sizes the accuracy values were forwarded into the SPSS IBM tool and performed different iterations. By comparing both algorithms we got a Random forest algorithm that has higher accuracy values over the Decision Tree algorithm.

The data set for this research is collected from Kaggle which is an open-source platform for getting machine learning datasets (**Singh** 2020). There are 981 rows and 13 columns are obtained by combining the train and test datasets that are used in the

algorithms. In the datasets, different dependent and independent variables are considered to perform Machine learning experimental procedures for Innovative loan eligibility prediction.

Statistical Analysis

The IBM SPSS is software for editing and analyzing all sorts of data. IBM Spss version 25 is the statistical software tool used for the loan eligibility prediction data analysis. Table 1 and Table 2 are clearly shown the dependent and independent variables used for training and testing the data sets. An independent variable T-test was carried out to compare the parameters on both groups. The IBM Statistical tool can analyze the data and help to identify the Mean, Standard deviation, standard mean error, and also some of the Independent t-tests like Mean difference, Standard error differences between two algorithms. Before sending results into the Spss tool the data sets are standardized and then the data is converted into arrays ("Integration of Unsupervised and Supervised Machine Learning Algorithms for Credit Risk Assessment" 2019)WangBao, NingLianju, and KongYue 2019("Integration of Unsupervised and Supervised Machine Learning Algorithms for Credit Risk Assessment" 2019). Finally, the obtained iteration values for the Random Forest algorithm and Decision Tree algorithm values are sent into SPSS, and all the observations of statistical analysis were found in this experimental procedure.

RESULTS

Table 3 shows the different accuracies for the Random Forest algorithm and Decision Tree Algorithm with different test sizes. The accuracy for algorithms fluctuates with different test sizes in decimals. So, we had performed each algorithm with 10 iterations, and the mean accuracy values were found with help of the SPSS tool. By performing the Group Statics, the Mean accuracy score for the Random forest algorithm 89.94% is higher than the Decision Tree algorithm 86.69% as mentioned in Table 4.

Table 4 shows the metrics of Group statistics for the Mean, Standard Deviation, and Standard Error Mean for both the algorithms. The Mean accuracies for Random Forest and Decision Tree Algorithm are 89.94% and 86.69%. The standard deviation for Random Forest Algorithm is 0.51130 and for the Decision Tree algorithm is 1.68581. Table 4 clearly shows that the standard Error Mean for Random Forest Algorithm is .16169 and for the Decision Tree Algorithm is 0.53310. Table 5 shows the Independent samples test by differentiating both the algorithms. The Sig value is .024 where $P < 0.05$ and the results were satisfactory. The standard Error Difference for the Random Forest and Decision Tree algorithms is 0.55708. The Mean difference for both the Random Forest algorithm and Decision Tree algorithm is 3.25200.

Figure 1 shows the Innovative Loan Prediction architecture diagram. The architecture diagram shows the various steps that are performed for getting better accuracies and a machine learning model. The steps that are involved in the Innovative loan approval machine model are Data Collection, Exploratory Data Analysis, Data visualization, Handling Missing values, Splitting the training and testing data, Training the Machine learning model, and Finding Metrics Values, and finally testing the machine learning model.

Figure 2 shows the Simple Mean accuracy bar for the Random forest algorithm and Decision Tree algorithm. The bar chart describes the comparison of the mean accuracy of the Random Forest 89.94% with the Decision Tree 86.69%. The bar charts of Random Forest and Decision Tree in the graphs are plotted on the X-axis by taking accuracy values on the Y-axis. The standard error mean for Random Forest Algorithm is 0.16169 and the Decision Tree algorithm is 0.53310. In Fig. 2. graph, the Error Bars are mentioned with 95% CI and $\pm$ 1 SD. The graph clearly shows that the Random forest has better and higher accuracy results over the Decision Tree algorithm.

DISCUSSION

By performing all the statistical analysis Results the Random Forest has higher accuracy than the Decision Tree algorithm for loan approval prediction. The opposite findings that are observed in authors had used the different machine learning techniques and got a Maximum accuracy score of 82% in Random Forest Algorithm and 72% accuracy in the Decision Tree Algorithm(Udaya Bhanu and Narayana 2021). And the author (Madaan et al. 2021) had achieved an accuracy of 80% for the Random Forest with the help of machine learning. Whereas after performing the statistical analysis, in this paper we got higher accuracies for the Random forest algorithm with 89.94% and Decision Tree algorithm 86.69%. The significance value of 0.024 showed the performed statistical analysis hypothesis holds good in this research. This paper had achieved more accuracy than the above-mentioned (Udaya Bhanu and Narayana 2021) and (Madaan et al. 2021) research.

The author's (Zhu et al. 2019) paper clearly concludes that Random Forest is the best algorithm in Machine learning for loan prediction with an accuracy score of 98% in his paper. The authors were used different data sets that consisted of 115,000 loan data of users with 102 attributes whereas in this experimental research fewer data set values are used and achieved a mean accuracy of 89.94% for the Random Forest Algorithm which is lesser than the referred (Zhu et al. 2019) paper. In this paper, we proved that the Random Forest algorithm has higher accuracy than the Decision Tree algorithm.

Although the results of implementation were good in both Random Forest and Decision Tree algorithms there exist some limitations. The algorithms were trained with the fewer data set values it is not possible to predict exact loan approval status for the prediction of a higher amount of loan approval status. So, we need to include more variables in the data and make the model more efficient for algorithms. The algorithm is not developed with the User interface, only developers can process the user's data. The accuracy of Random Forest and Decision Tree should also increase by making data in a more efficient manner that can be developed in Further Processes.

In the future scope, the accuracy will be improved for both Random Forest and Decision Machine learning models for innovative loan eligibility prediction in our study by performing different methods while training the model. We will collect data from different sources with different data variables provided by the banks and financial institutions and then try to improve more accuracy and the predictions for loan eligibility are increased and more accurate. In the future, we will develop this model with the user interface and deployed it in the cloud. So, it is easy to verify and to identify whether the person is Eligible for Loan approval or not by customers.

CONCLUSION

The experimental research for loan eligibility prediction is implemented using Python programming and the IBM SPSS tool. These tools show that the Random forest algorithm and Decision Tree have enhanced the accuracy results of the algorithms. From the above study, we found that the Random forest algorithm has around 89.94% and the Decision Tree algorithm has around 86.69% Accuracy. Hence we had proven that the Random forest algorithm has higher accuracy over the Decision Tree algorithm for innovative loan eligibility prediction.

DECLARATIONS

Conflict of Interest

The authors of this paper declare no conflict of interest.

Author Contribution
Author NRK was involved in data collection, data analysis, manuscript preparation. Author LRP was involved in the conceptualization, data validation, and critical analysis of the manuscript.

Acknowledgments
The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding
The authors would like to thank the following organizations for providing financial support that enabled us to complete the study.

1. Inmitto Solutions Pvt. Ltd.
2. Saveetha University
3. Saveetha Institute of Medical And Technical Sciences
4. Saveetha School of Engineering

REFERENCES

Addo, Peter Martey, Dominique Guegan, and Bertrand Hassani. 2018. "Credit Risk Analysis Using Machine and Deep Learning Models." *Risks* 6 (2): 38.

Arutjothi, G., and C. Senthamarai. 2017. "Prediction of Loan Status in Commercial Bank Using Machine Learning Classifier." In *2017 International Conference on Intelligent Sustainable Systems (ICISS)*. IEEE. https://doi.org/10.1109/iss1.2017.8389442.

Aslam, Uzair, Hafiz Ilyas Tariq Aziz, Asim Sohail, and Nowshath Kadhar Batcha. 2019. "An Empirical Study on Loan Default Prediction Models." *Journal of Computational and Theoretical Nanoscience* 16 (8): 3483–88.

Benin, S. R., S. Kannan, Renjin J. Bright, and A. Jacob Moses. 2020. "A Review on Mechanical Characterization of Polymer Matrix Composites & Its Effects Reinforced with Various Natural Fibres." *Materials Today: Proceedings* 33 (January): 798–805.

Gudipaneni, Ravi Kumar, Mohammad Khursheed Alam, Santosh R. Patil, and Mohmed Isaqali Karobari. 2020. "Measurement of the Maximum Occlusal Bite Force and Its Relation to the Caries Spectrum of First Permanent Molars in Early Permanent Dentition." *The Journal of Clinical Pediatric Dentistry* 44 (6): 423–28.

"Integration of Unsupervised and Supervised Machine Learning Algorithms for Credit Risk Assessment." 2019. *Expert Systems with Applications* 128 (August): 301–15.

Liu, Kaiyang, Jun Peng, Jingrong Wang, Boyang Yu, Zhuofan Liao, Zhiwu Huang, and Jianping Pan. 2019. "A Learning-Based Data Placement Framework for Low Latency in Data Center Networks." *IEEE Transactions on Cloud Computing*. https://doi.org/10.1109/tcc.2019.2940953.

Madaan, Mehul, Aniket Kumar, Chirag Keshri, Rachna Jain, and Preeti Nagrath. 2021. "Loan Default Prediction Using Decision Trees and Random Forest: A Comparative Study." *IOP Conference Series: Materials Science and Engineering*. https://doi.org/10.1088/1757-899x/1022/1/012042.

Nalini, Devarajan, Jayaraman Selvaraj, and Ganesan Senthil Kumar. 2020. "Herbal Nutraceuticals: Safe and Potent Therapeutics to Battle Tumor Hypoxia." *Journal of Cancer Research and Clinical Oncology* 146 (1): 1–18.

Rath, Golak Bihari, Debasish Das, and Biswaranjan Acharya. 2021. "Modern Approach for Loan Sanctioning in Banks Using Machine Learning." In *Algorithms for Intelligent Systems*, 179–88. Singapore: Springer Singapore.

Reddy, Poornima, Jogikalmat Krithikadatta, Valarmathi Srinivasan, Sandhya Raghu, and Natanasabapathy Velumurugan. 2020. "Dental Caries Profile and Associated Risk Factors Among Adolescent School Children in an Urban South-Indian City." *Oral Health & Preventive Dentistry* 18 (1): 379–86.

Sathish, T., and S. Karthick. 2020. "Gravity Die Casting Based Analysis of Aluminum Alloy

with AC4B Nano-Composite." *Materials Today: Proceedings* 33 (January): 2555–58.
Sathish, T., D. Bala Subramanian, R. Saravanan, and V. Dhinakaran. 2020. "Experimental Investigation of Temperature Variation on Flat Plate Collector by Using Silicon Carbide as a Nanofluid." In *PROCEEDINGS OF INTERNATIONAL CONFERENCE ON RECENT TRENDS IN MECHANICAL AND MATERIALS ENGINEERING: ICRTMME 2019*. AIP Publishing. https://doi.org/10.1063/5.0024965.
Sheikh, Mohammad Ahmad, Amit Kumar Goel, and Tapas Kumar. 2020. "An Approach for Prediction of Loan Approval Using Machine Learning Algorithm." In *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*. IEEE. https://doi.org/10.1109/icesc48915.2020.9155614.
Singh□Sonali. 2020. "Loan Approval Prediction." https://www.kaggle.com/sonalisingh1411/loan-approval-prediction.
Sivasamy, Ramesh, Potu Venugopal, and Rodrigo Espinoza-González. 2020. "Structure, Electronic Structure, Optical and Magnetic Studies of Double Perovskite Gd2MnFeO6 Nanoparticles: First Principle and Experimental Studies." *Materials Today Communications* 25 (December): 101603.
"Study on a Prediction of P2P Network Loan Default Based on the Machine Learning LightGBM and XGboost Algorithms according to Different High Dimensional Data Cleaning." 2018. *Electronic Commerce Research and Applications* 31 (September): 24–39.
Udaya Bhanu, L., and Dr S. Narayana. 2021. "Customer Loan Prediction Using Supervised Learning Technique." *Mathematical Sciences Research Journal. An International Journal of Rapid Publication* 11 (6): 403–7.
Venu, Harish, and Prabhu Appavu. 2021. "Experimental Studies on the Influence of Zirconium Nanoparticle on Biodiesel–diesel Fuel Blend in CI Engine." *International Journal of Ambient Energy* 42 (14): 1588–94.
"Website." n.d. https://doi.org/10.3390/su11030699.
Yu, Boyang, and Jianping Pan. 2016. "Sketch-Based Data Placement among Geo-Distributed Datacenters for Cloud Storages." *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*. https://doi.org/10.1109/infocom.2016.7524627.
Zhu, Lin, Dafeng Qiu, Daji Ergu, Cai Ying, and Kuiyi Liu. 2019. "A Study on Predicting Loan Default Based on the Random Forest Algorithm." *Procedia Computer Science*. https://doi.org/10.1016/j.procs.2019.12.017.
Munkhdalai, L.; Munkhdalai, T.; Namsrai, O.-E.; Lee, J.Y.; Ryu, K.H. An Empirical Comparison of Machine-Learning Methods on Bank Client Credit Assessments. Sustainability **2019**, 11, 699. https://doi.org/10.3390/su11030699.
Orlova, E.V. Decision-Making Techniques for Credit Resource Management Using Machine Learning and Optimization. Information **2020**, 11, 144. https://doi.org/10.3390/info11030144.

TABLES AND FIGURES

Table 1. The below table shows the Independent variables that are mentioned in the Loan eligibility prediction data set and taken for the Random Forest and Decision Tree algorithms.

Independent Variables	Data Description
Loan_ID	Unique Loan ID
Gender	Male/ Female
Married	Applicant married (Y/N)
Self_Employed	Self-employed (Y/N)
Education	Applicant Education (Graduate/ Undergraduate).

Table 2. The below table shows the dependent variables that are mentioned in the Loan eligibility prediction data set and taken for the Random Forest and Decision Tree algorithms.

Dependent Variables	Data Description
ApplicantIncome	Applicant income
CoapplicantIncome	Co-Applicant income
LoanAmount	Loan amount in thousands
Loan_Amount_Term	Term of the loan in months
Credit_History	credit history meets guidelines
Property_Area	Urban/ Semi-Urban/ Rural
Loan_Status	(Target) Loan approved (Y/N)

Table 3. The below table shows the 10 iterations of the Random Forest algorithm and Decision Tree algorithm with different test sizes to perform training and testing iterations and their extracted accuracies.

Test Size	RFA Accuracy	DT Accuracy
0.2	90.36	89.34
0.25	90.24	88.21
0.28	90.91	88.36
0.3	90.17	87.12
0.35	89.81	86.42
0.38	89.24	86.34
0.4	89.54	86.60
0.43	89.31	83.97
0.45	89.81	84.36
0.5	90.05	86.20

Table 4. From the group statistics, the standard deviation and Mean accuracy for the Random forest algorithm are 0.51130 and 89.9440, and also for the Decision Tree is 1.68581 and 86.6920.

Group Statistics					
	RF, DT	N	Mean	Std. Deviation	Std. Error Mean

Accuracy	**RF**	10	89.9440	.51130	.16169
	DT	10	86.6920	1.68581	.53310

Table 5. The below table shows the Independent samples of RF and DT, by comparing the accuracy of both algorithms the significance rate is 0.024, and the std error difference is 0.55708.

Independent Samples Test							
		Levene's Test for Equality of Variances		**t-test for Equality of Means**			
		F	**Sig.**	**Mean Difference**	**Std. Error Difference**	**95% Confidence Interval of the Difference**	
						Lower	**Upper**
Accuracy	**Equal variances assumed**	6.123	.024	3.25200	.55708	2.08162	4.42238
	Equal variances not assumed			3.25200	.55708	2.02082	4.48318

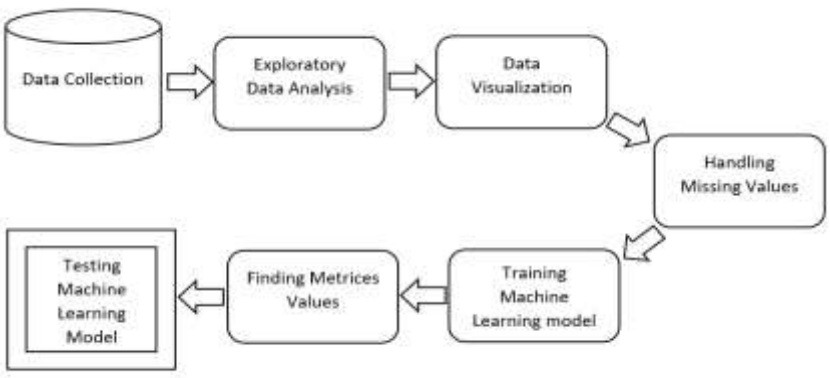

Fig. 1. Innovate Machine learning classifier architecture.

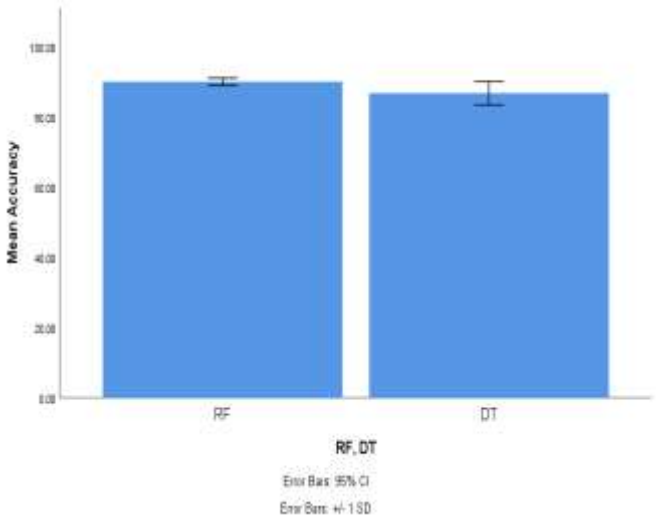

Fig. 2. The bar chart shows the comparison of the Mean Accuracy of analysis of Loan eligibility prediction using the Logistic Regression Algorithm (LOR) and Decision Tree Algorithm (DT). The Logistic Regression Provides higher accuracy and more compatible results. The parameters that are mentioned in the above graph are On the X-axis: LOR vs DT. Y-axis: Mean Accuracy is ±1 SD.