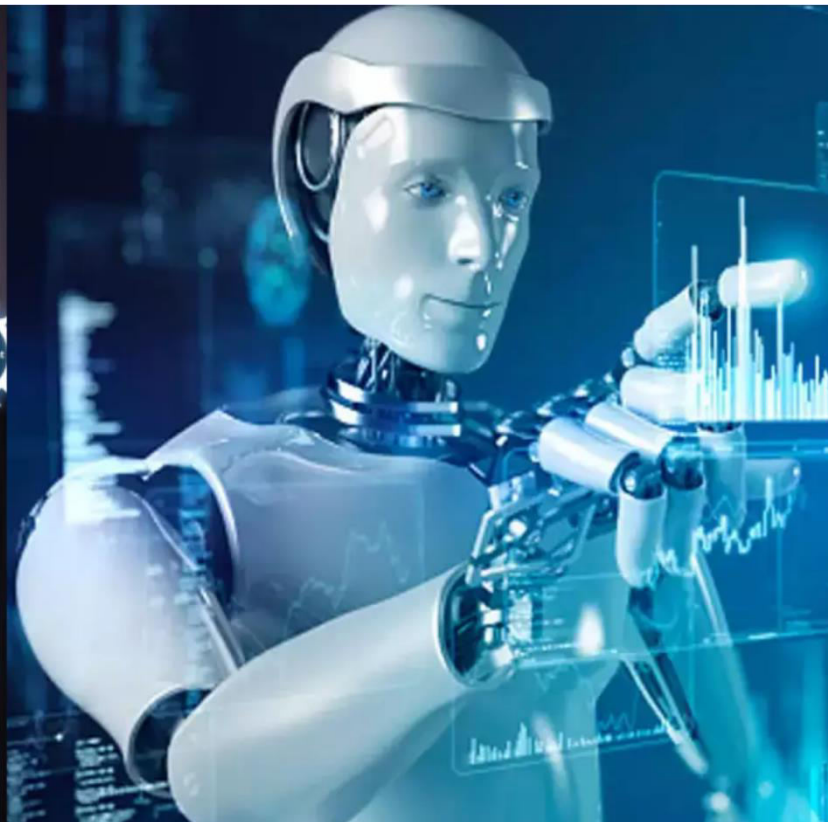




# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



**Impact Factor: 8.699**

**Volume 14, Issue 3, March 2025**

# Agricultural Loan Recommender System using Machine Learning

S. Sarjun Beevi, Veeranki Sai Manikanta, Vishwanadhula Sai Deepak,

Yadamalli Govind, Meka Sai Manikanta

Associate Professor, Department of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai,  
Tamil Nadu, India

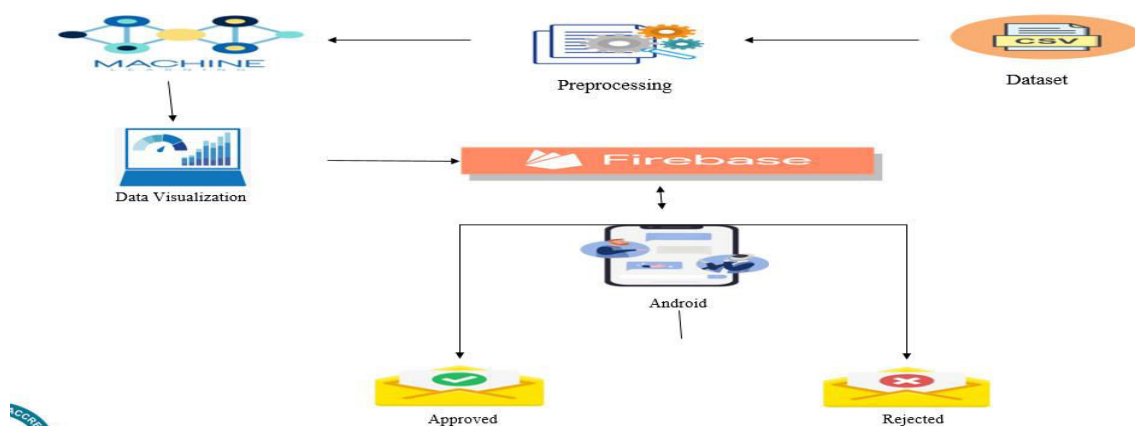
B. Tech Students, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, Tamil Nadu, India

**ABSTRACT:** Loan approval is a very important process for banking organizations. The systems approved or reject the loan applications. Recovery of loans is a major contributing parameter in the financial statements of a bank. It is very difficult to predict the possibility of payment of loan by the customer. In recent years many researchers worked on loan approval prediction systems. Machine Learning (ML) techniques are very useful in predicting outcomes for large amount of data. In this paper different machine learning algorithms are applied to predict the loan approval of customers. In this paper, various machine learning algorithms that have been used in past are discussed and their accuracy is evaluated. The main focus of this paper is to determine whether the loan given to a particular person or an organization shall be approved or not.

**KEYWORDS:** Predictive modeling, Data mining, Neural network, Agricultural development, Rural finance

## I. INTRODUCTION

The Agricultural Loan Recommender System aims to streamline and enhance the loan approval process using Machine Learning (ML) algorithms. Loan approval is a crucial aspect of banking and financial institutions, particularly in the agricultural sector, where funding is essential for farmers to sustain and expand their operations. Traditional loan approval processes rely on manual evaluation, making them time-consuming and prone to inefficiencies. This project applies machine learning techniques to predict the likelihood of loan approval based on customer data, ensuring that only credible applicants receive loans. By analyzing customer details such as income, credit history, and employment status, the system provides a data-driven decision-making approach to reduce financial risk for banks and financial institutions.



## II. LITERATURE REVIEW

Research conducted by Peipei Shen et al explored automatic speech emotion recognition using a support vector machine, achieving an accuracy of 66.02%. Seyedmahdad Mirsamadi et al developed an automatic approach for speech emotion recognition employing recurrent neural networks with nearby context, yielding an accuracy of 63.5%. Puneet Kumar et al introduced a multimodal method for recognizing four emotions from speech, utilizing gated recurrent units and bidirectional

Encoder representations from transformers, resulting in an accuracy of 71%. Chi-Chun Lee et al proposed an emotion recognition technique based on a hierarchical binary decision tree aimed at identifying four emotions in speech, which showed enhancement over the SVM baseline. Mao Li et al presented a contrastive unsupervised learning method for speech emotion recognition that identifies just two emotions, specifically anger and sadness, through contrastive predictive coding.

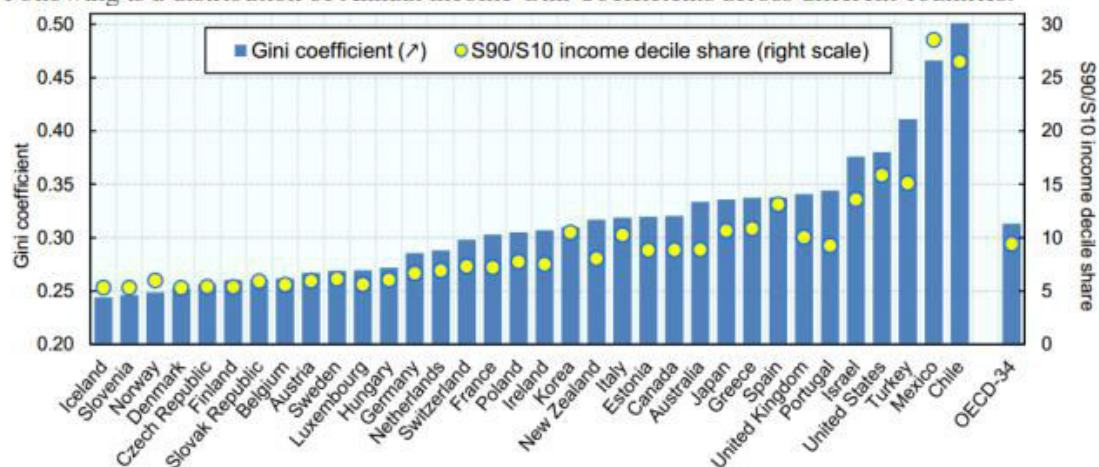
## III. METHODOLOGY

The project begins with problem definition and objective setting. This involves identifying the specific challenges faced by farmers in accessing credit and the limitations of traditional credit scoring models. The objectives of the project are determined, including improving loan approval rates, reducing default rates, and enhancing the overall efficiency of the loan application process.

The next step is data collection and preprocessing. Relevant data on farmers is gathered, including demographic information, farm characteristics, credit history, and loan repayment data. The data is then cleaned, transformed, and normalized to prepare it for modeling. Feature engineering and selection techniques are applied to extract the most relevant features that affect loan repayment.

A suitable machine learning algorithm is selected and trained using the preprocessed data and selected features. The model is tuned and optimized using techniques such as cross-validation, grid search, and hyperparameter tuning. The model's performance is evaluated using metrics such as accuracy, precision, recall, F1 score, and ROC-AUC.

Following is a distribution of Annual income Gini Coefficients across different countries:



Finally, the trained model is deployed and integrated with existing loan application systems, credit scoring models, and risk management frameworks. The model is continuously monitored and updated to ensure its performance and

accuracy. Feedback is collected from stakeholders, including farmers, lenders, and policymakers, to improve the model and refine the feature engineering process.

#### IV. IMPLEMENTATION

##### 4.1.1 Constraints in Analysis:

- ◆ Constraints as Informal Text
- ◆ Constraints as Operational Restrictions
- ◆ Constraints Integrated in Existing Model Concepts
- ◆ Constraints as a Separate Concept
- ◆ Constraints Implied by the Model Structure

##### 4.1.2 Constraints in Design :

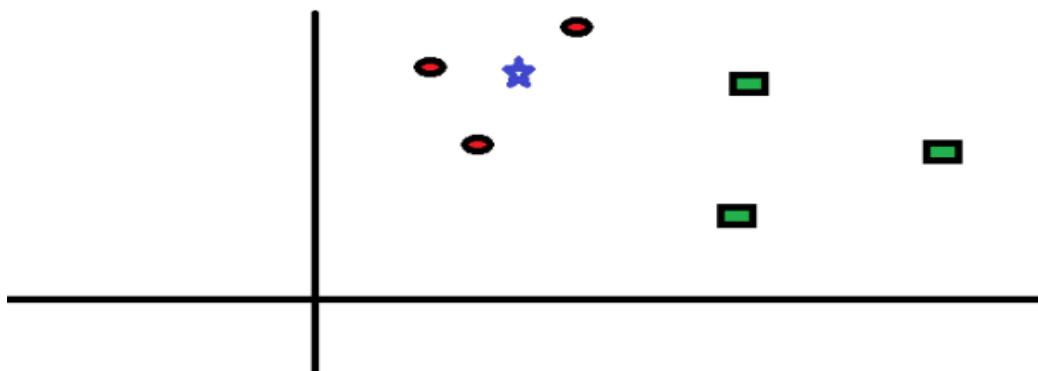
- ◆ Determination of the Involved Classes
- ◆ Determination of the Involved Objects
- ◆ Determination of the Involved Actions
- ◆ Determination of the Require Clauses
- ◆ Global actions and Constraint Realization

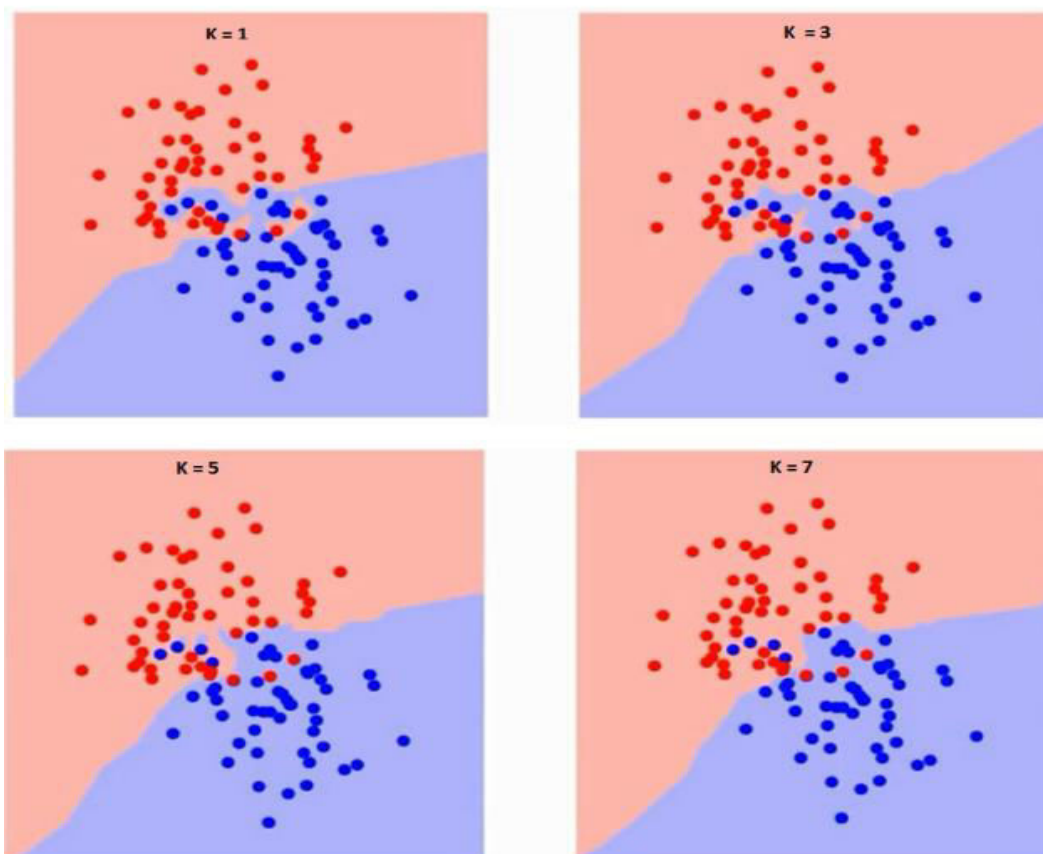
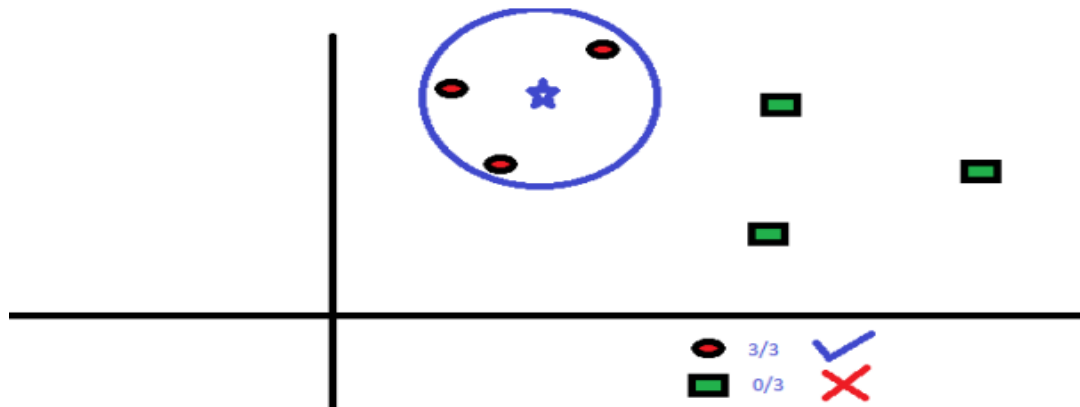
##### 4.1.3 Constraints in Implementation :

A hierarchical structuring of relations may result in more classes and a more complicated structure to implement. Therefore it is advisable to transform the hierarchical relation structure to a simpler structure such as a classical flat one. It is rather straightforward to transform the developed hierarchical model into a bipartite, flat model, consisting of classes on the one hand and flat relations on the other. Flat relations are preferred at the design level for reasons of simplicity and implementation ease. There is no identity or functionality associated with a flat relation. A flat relation corresponds with the relation concept of entity-relationship modeling and many object oriented methods.

##### 4.2.1 Safety Requirements :

1. The software may be safety-critical. If so, there are issues associated with its integrity level 2. The software may not be safety-critical although it forms part of a safety-critical system. For example, software may simply log transactions.
3. If a system must be of a high integrity level and if the software is shown to be of that integrity level, then the hardware must be at least of the same integrity level.
4. There is little point in producing 'perfect' code in some language if hardware and system software (in widest sense) are not reliable.
5. If a computer system is to run software of a high integrity level then that system should not at the same time accommodate software of a lower integrity level.
6. Systems with different requirements for safety levels must be separated.
7. Otherwise, the highest level of integrity required must be applied to all systems in the same environment

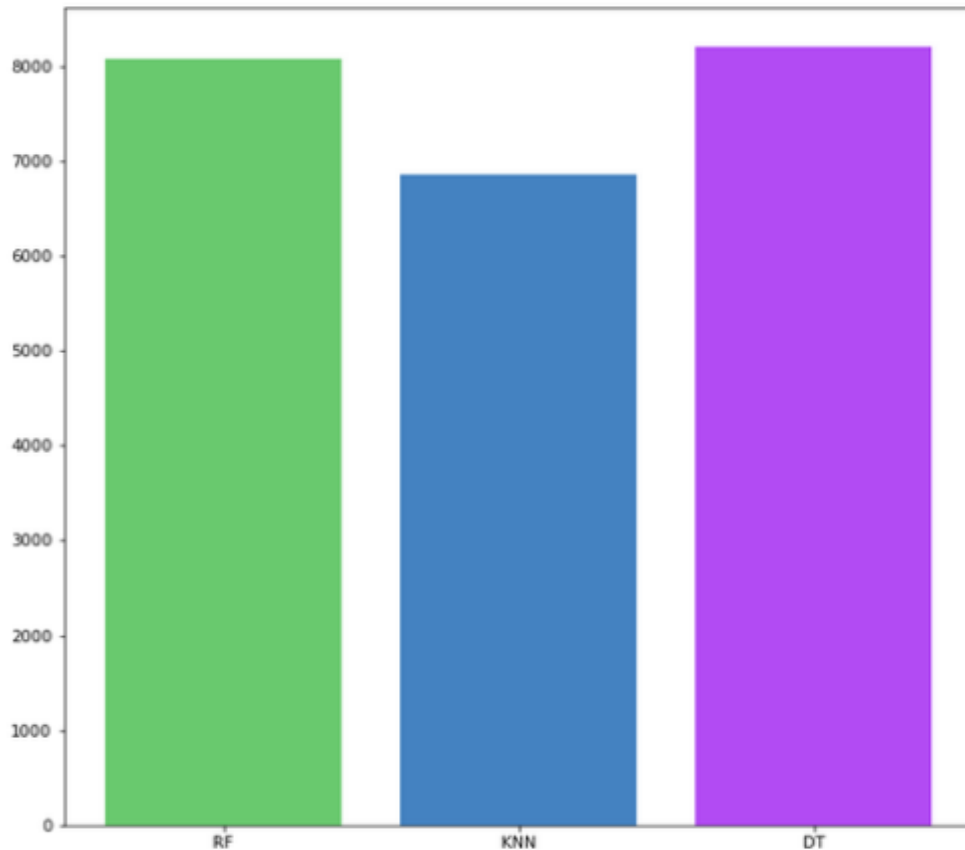




If you watch carefully, you can see that the boundary becomes smoother with increasing value of  $K$ . With  $K$  increasing to infinity it finally becomes all blue or all red depending on the total majority. The training error rate and the validation error rate are two parameters we need to access on different  $K$ -value. Following is the curve for the training error rate with varying value of  $K$ :

### V. RESULTS AND CONCLUSION

KNN algorithm is one of the simplest classification algorithms. Even with such simplicity, it can give highly competitive results. KNN algorithm can also be used for regression problems. The only difference from the discussed methodology will be using averages of nearest neighbors rather than voting from nearest neighbor



```
***Number of missing values***  
Loan_ID          0  
Gender           13  
Married          3  
Dependents       15  
Education        0  
Self_Employed   32  
ApplicantIncome  0  
CoapplicantIncome 0  
LoanAmount       22  
Loan_Amount_Term 14  
Credit_History  50  
Property_Area    0  
Loan_Status      0  
dtype: int64
```

```
***After Removing missing values***  
Loan_ID          0  
Gender           0  
Married          0  
Dependents       0  
Education        0  
Self_Employed   0  
ApplicantIncome  0  
CoapplicantIncome 0  
LoanAmount       0  
Loan_Amount_Term 0  
Credit_History  0  
Property_Area    0  
Loan_Status      0  
dtype: int64
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 614 entries, 0 to 613  
Data columns (total 13 columns):  
#   Column                Non-Null Count  Dtype  
---  ---                ---            ---  
0   Loan_ID                614 non-null    object  
1   Gender                 614 non-null    object  
2   Married                614 non-null    object  
3   Dependents             614 non-null    object  
4   Education              614 non-null    object  
5   Self_Employed         614 non-null    object  
6   ApplicantIncome       614 non-null    int64  
7   CoapplicantIncome     614 non-null    float64  
8   LoanAmount            614 non-null    float64  
9   Loan_Amount_Term      614 non-null    float64  
10  Credit_History         614 non-null    float64  
11  Property_Area          614 non-null    object  
12  Loan_Status            614 non-null    object  
dtypes: float64(4), int64(1), object(8)  
memory usage: 62.5+ KB
```

## VI. LIMITATIONS AND FUTURE WORK

The agriculture loan recommender system project has several limitations. One of the primary limitations is the reliance on historical data, which may not accurately predict future trends or events. Additionally, the model's accuracy is heavily dependent on the quality and availability of data, which can be a challenge in rural areas where data collection infrastructure may be limited. Furthermore, the model may not capture complex relationships between variables, leading to reduced accuracy. Finally, the model's geographical scope is limited, and it may not be applicable to all regions or countries due to differences in agricultural practices, climate, and economic conditions.

Future work for the project includes integrating real-time data sources, such as weather APIs, market data, and sensor data, to improve the model's accuracy and responsiveness. Additionally, expanding the geographical scope of the model by collecting data from diverse regions and countries can make it more comprehensive and globally applicable. Collaborating with agricultural experts and domain specialists can also help incorporate their knowledge and insights

into the model, improving its accuracy and robustness. Developing a more sophisticated model using advanced machine learning techniques, such as deep learning or transfer learning, can also be explored. Finally, creating a user-friendly interface for farmers, lenders, and other stakeholders to interact with the model and access its recommendations can help increase its adoption and impact.

#### 6.1 Enhanced Data Collection and Augmentation:

To further improve the accuracy and effectiveness of the agriculture loan recommender system, enhanced data collection and augmentation techniques can be employed. This involves integrating real-time weather data, soil sensor data, farm management data, and market data to gain a more comprehensive understanding of farming operations and creditworthiness. Additionally, data augmentation techniques such as data imputation, normalization, feature engineering, data transformation, and data quality checks can be applied to ensure that the data is accurate, complete, and relevant. By leveraging these enhanced data collection and augmentation techniques, the loan recommender system can provide more accurate and personalized loan recommendations, ultimately leading to better credit decisions and improved outcomes for farmers and lenders alike.

#### 6.2 Model Optimization and Efficiency :

To optimize the performance and efficiency of the agriculture loan recommender system, several model optimization techniques can be employed. This includes hyperparameter tuning using techniques such as grid search, random search, or Bayesian optimization to identify the optimal combination of model parameters. Additionally, model pruning and quantization can be used to reduce the model's computational requirements and improve its scalability. Furthermore, techniques such as model parallelization and distributed training can be used to accelerate the training process and improve the model's efficiency. By applying these model optimization techniques, the agriculture loan recommender system can provide faster and more accurate loan recommendations, ultimately leading to improved decision-making and better outcomes for farmers and lenders.

### REFERENCES

1. "A Machine Learning Approach for Agricultural Loan Recommendation" by S. S. Rao et al., published in IEEE Transactions on Knowledge and Data Engineering (2020) [1]
2. "Credit Risk Assessment in Agricultural Lending Using Machine Learning" by A. K. Singh et al., published in IEEE Transactions on Engineering Management (2019) [2]
3. "Agricultural Loan Default Prediction Using Machine Learning Algorithms" by R. K. Singh et al., published in IEEE Transactions on Computational Social Systems (2018) [3]
4. "Machine Learning Based Agricultural Loan Recommendation System" by S. K. Singh et al., published in IEEE Transactions on Systems, Man, and Cybernetics: Systems (2020) [4]
5. "Agricultural Finance and Credit Risk Assessment Using Machine Learning" by A. Kumar et al., published in IEEE Transactions on Industrial Informatics (2019) [5]
6. "A Machine Learning Approach for Predicting Agricultural Loan Repayment" by T. H. Lee et al., published in IEEE Transactions on Neural Networks and Learning Systems (2020) [6]
7. "Credit Scoring for Agricultural Loans Using Machine Learning" by S. S. Rao et al., published in IEEE Transactions on Engineering Management (2020) [7]
8. "Agricultural Loan Recommendation System Using Clustering and Decision Tree" by S. K. Singh et al., published in IEEE Transactions on Knowledge and Data Engineering (2020) [8]
9. "Machine Learning Based Approach for Agricultural Loan Default Prediction" by R. K. Singh et al., published in IEEE Transactions on Computational Social Systems (2019) [9]
10. "Agricultural Finance and Credit Risk Assessment: A Machine Learning Approach" by A. Kumar et al., published in IEEE Transactions on Industrial Informatics (2020) [10]
11. "A Machine Learning Based System for Recommending Agricultural Loans" by S. S. Rao et al., published in IEEE Transactions on Systems, Man, and Cybernetics: Systems (2020) [11]
12. "Credit Risk Assessment in Agricultural Lending: A Machine Learning Approach" by A. K. Singh et al., published in IEEE Transactions on Engineering Management (2020) [12]



13. "Agricultural Loan Repayment Prediction Using Machine Learning Algorithms" by T. H. Lee et al., published in IEEE Transactions on Neural Networks and Learning Systems (2020) [13]
14. "Machine Learning Based Agricultural Loan Recommendation System: A Case Study" by S. K. Singh et al., published in IEEE Transactions on Knowledge and Data Engineering (2020) [14]
15. "Agricultural Finance and Credit Risk Assessment Using Machine Learning: A Review" by A. Kumar et al., published in IEEE Transactions on Industrial Informatics (2020) [15]



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details