****

**Assessment Report**

on

**“Predict Loan Default”**

submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2024-25

in

**CSE(AI)**

By

Name : Akshit Vats

Roll Number : 202401100300026

Name : Alok Kumar Mishra

Roll Number : 202401100300029

Name : Aman Singh Yadav

Roll Number : 202401100300031

Name : Arnav Bhardwaj

Roll Number : 202401100300062

Name : Anushka Shakya

Roll Number : 202401100300060

Section: A

**Under the supervision of**

“Bikki Kumar”

**KIET Group of Institutions, Ghaziabad**

**May, 2025**

**1. Introduction**

With the growing demand for intelligent agricultural solutions, machine learning provides effective tools to enhance crop productivity. This project focuses on designing a **Crop Recommendation System** that predicts the most suitable crop to cultivate based on soil and environmental parameters. The system aims to help farmers make informed decisions and optimize agricultural yields by leveraging data-driven insights.

**2. Problem Statement**

To recommend the most appropriate crop to grow on a given land using soil characteristics (Nitrogen, Phosphorus, Potassium), environmental factors (temperature, humidity), pH value, and rainfall data. The classification model will guide farmers in selecting crops that are most likely to succeed in given conditions.

**3. Objectives**

* Collect and preprocess agricultural data for machine learning.
* Train multiple classification algorithms to recommend suitable crops.
* Evaluate and compare the performance of different models.
* Visualize evaluation metrics for better interpretability.

**4. Methodology**

* **Data Collection**: Dataset is obtained from Kaggle: Link: https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset. It contains the following features:
* N, P, K (soil nutrients)
* Temperature
* Humidity
* pH
* Rainfall
* Label: Crop type
* **Data Preprocessing**:
* No missing values were found in the dataset.
* Features are standardized using **StandardScaler**.
* Target labels (crop names) are encoded using **LabelEncoder**.

* **Model Building**:  
   Dataset is split into 80% training and 20% testing sets.

Models used:

* K-Nearest Neighbors (KNN)
* Decision Tree Classifier
* Random Forest Classifier
* Naive Bayes Classifier
* **Model Evaluation**:

Metrics used:

* Accuracy
* Precision
* Recall
* F1 Score

**5. Data Preprocessing**

The dataset is cleaned and prepared as follows:

* **Standardization:** All numerical features are scaled to have zero mean and unit variance.
* **Encoding:** The crop names are converted to numeric labels.
* **Splitting:** Dataset is split into training and test sets (80:20).

**6. Model Implementation**

Logistic Regression is used due to its simplicity and effectiveness in binary classification problems. The model is trained on the processed dataset and used to predict the loan default status on the test set.

**7. Evaluation Metrics**

The following metrics are used to evaluate the model:

* **Accuracy**: Measures overall correctness.
* **Precision**: Indicates the proportion of predicted defaults that are actual defaults.
* **Recall**: Shows the proportion of actual defaults that were correctly identified.
* **F1 Score**: Harmonic mean of precision and recall.
* **Confusion Matrix**: Visualized using Seaborn heatmap to understand prediction errors.

**8. Results and Analysis**

* The model provided reasonable performance on the test set.
* Confusion matrix heatmap helped identify the balance between true positives and false negatives.
* Precision and recall indicated how well the model detected loan defaults versus false alarms.

**9. Conclusion**

The logistic regression model successfully classified loan defaults with satisfactory performance metrics. The project demonstrates the potential of using machine learning for automating loan approval processes and improving risk assessment. However, improvements can be made by exploring more advanced models and handling imbalanced data.

**10. References**

* scikit-learn documentation
* pandas documentation
* Seaborn visualization library
* Research articles on credit risk prediction

















