MACHINE LEARNING PROJECT

PROJECT TITLE: **The Crop Recommendation Model**

Under the guidance of: **Professor Liangwen Wu**

TEAM MEMBERS:

Tanvi Patel - 1316221

Pratiksha Kande - 1309129

Pankti Bhatt - 1287645

TABLE OF CONTENTS

INTRODUCTION

Agriculture plays a crucial role in the economy of any country. However, farmers often face challenges in determining which crops to grow based on various factors such as soil type, climate changes, and weather patterns. To address this issue, we have developed a crop recommendation system that can provide farmers with personalized recommendations based on their location, soil type, and weather patterns. This system can help farmers make informed decisions about which crops to grow and when, leading to higher crop yields and increased profits.

METHODOLOGY

Dataset Description

- We used a Kaggle Dataset named **Crop Recommendation Dataset** by **Atharva Ingle**.
- This dataset was built by augmenting datasets of rainfall, climate and fertilizer data available for India.
- **Data Fields:**
    - N - Ratio of Nitrogen content in soil
    - P - Ratio of Phosphorous content in soil
    - K - Ratio of Potassium content in soil
    - Temperature - Temperature in degree Celsius
    - Humidity - Relative humidity in %
    - Ph - Ph value of the soil
    - Rainfall - Rainfall in mm

Data Preprocessing

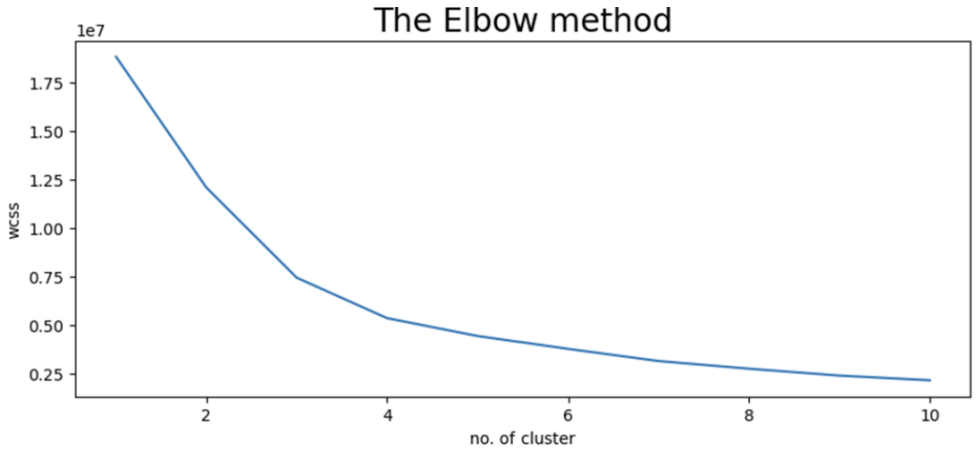
- An interactive Python function to summarize crop characteristics using the @interact decorator.
- The function displays the statistics of the variables and displays the mean, median, minimum, and maximum values of each variable for the selected crop using the agg() method.
- By using this function, we can easily compare the characteristics of different crops and identify patterns in the data.

Baseline Model Used

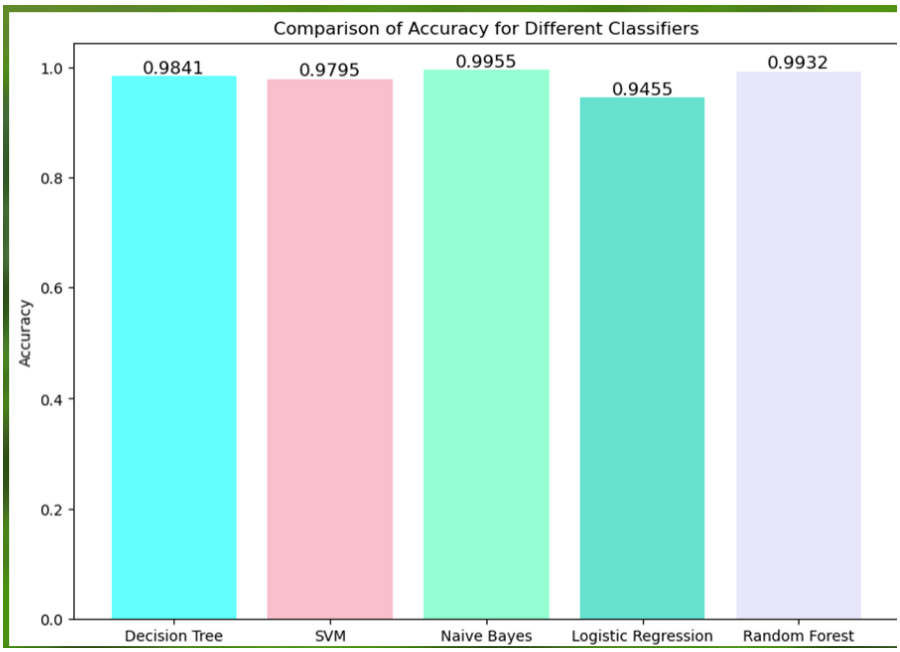
- Decision Tree Classifier
- Random Forest Classifier
- Support Vector Machine
- Naïve Bayes
- Logistic Regression

Implementation

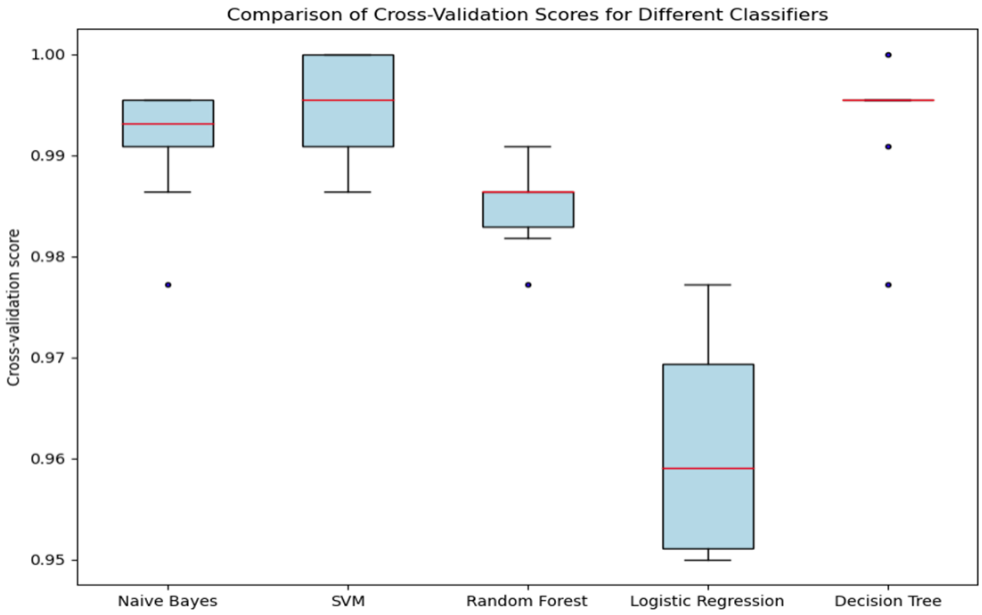
- The Elbow Method for Optimal Number of Clusters
    - The elbow method is used for our crop recommendation dataset to determine the optimal number of clusters to use in our KMeans clustering algorithm.
    - KMeans clustering algorithm for clustering the crops based on their nutrient requirements, temperature, humidity, rainfall, and soil ph.



- Bar Graph for Accuracy Comparison
  - We evaluated the performance of our machine learning models using accuracy metric.
  - We used evaluation metrics such as accuracy, precision, and recall to evaluate the performance of the models.



- Box Plot for Cross-Validation Score Comparison



Results

- Compared the accuracy of different machine learning models using our dataset, and the results are as follows:
    - Decision Tree: 98.41%
    - Naive Bayes: 99.55%
    - SVM: 97.95%
    - Logistic Regression: 94.55%
    - Random Forest: 99.32%
- Based on the evaluation metrics, Naïve Bayes model performed the best with an accuracy of 99.55%. This indicates that Naïve Bayes model is the most suitable model for crop recommendation.

CONCLUSIONS AND FUTURE WORKS

Our crop recommendation system can provide accurate recommendations for different crops based on various environmental factors. This system can potentially benefit farmers by improving crop yields and reducing resource waste. Future work can include improving the dataset with more features and samples to enhance the accuracy of the system. Additionally, implementing deep learning algorithms can also be explored to further improve the system's performance. Another potential direction for future work is to develop a front-end web application for users to easily access the system and obtain crop recommendations.