**A DATA ANALYTICS PROJECT REPORT**

**on**

**“CROP RECOMMENDATION”**

**Submitted to**

**KIIT Deemed to be University**

**BACHELOR’S DEGREE IN**

**COMPUTER SCIENCE AND ENGINEERING**

**BY**

**RAVIKANT DIWAKAR** (21051499)

**UNDER THE GUIDANCE OF**

**MS. SRICHETA PARUI**

****

**SCHOOL OF COMPUTER ENGINEERING**

**KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY**

**BHUBANESWAR, ODISHA - 751024**

Contents

1. ABSTRACT 1
2. INTRODUCTION 2
   1. Motivation ……………………………………………………….….. 2
   2. Literature Review ……………………………………………….…... 3
   3. Project Goals and Objectives ……………………………………….. 4
   4. Model Development ………………………………………………… 4
3. METHODOLOGY 6
   1. System Architecture ………………………………………………… 6
   2. Datasets ……………………………………………………………… 7
   3. Crop Recommendation ……………………………………………… 8
   4. Analysis Tools and Technologies …………………………………… 8
4. RESULT 10
   1. Model Performance ……………………………………………...…. 10
   2. Website Deployment ……………………………………………….. 11
5. FUTURE SCOPE 13
6. CONCLUSION 14

List of Figures

3.1 System Architecture ………………………………………………… 6

3.2 Random data from the dataset for crop recommendation …………… 7

4.1 Accuracy Comparison ………………………………………..……. 10

4.2 Landing page of the website ………………………………………. 11

4.3 Result of crop recommendation ………………………………….... 12

*CROP RECOMMENDATION*

1. **Abstract**

Agriculture plays a crucial role in the economies of many developing countries, farmers face many obstacles in meeting the expectations of a changing global community. Crop productivity is severely affected by elements like soil erosion, fertilizer shortages, and climate change. Furthermore, as farmers depend more on instinct and little trial and error than on data-driven decision-making, traditional agricultural methods frequently provide less-than-ideal results. we use machine learning algorithms to the crop forecast problem. Our mission is to provide farmers with the knowledge and skills necessary to choose crops that are appropriate for their local climate and soil nutrient levels. We examine several algorithms to find the best crop prediction models, such as Decision Tree, Guassian Naive Bayes, Support Vector Machine (SVM) and Random Forest Through our experiments we compare the performance of these algorithms by on various assumptions. Out of all the models that were looked at, Random Forest comes out to be the most accurate, according to the data. Our goal is to transform crop planning techniques by utilizing machine learning, giving farmers the ability to maximize yields while reducing resource waste. This study is a fundamental step toward a more productive and sustainable agriculture where decisions are made based on data-driven insights.

**Keywords :** Machine Learning, Crop Recommender System, Random Forest, Decision Tree, Naive Bayes, SVM

*School of Computer Engineering, KIIT, BBSR* 1

*CROP RECOMMENDATION*

1. **Introduction**

Machine learning has become a powerful tool for solving challenging problems in many different fields, including agriculture. Crop recommendation is one such problem, where choosing the right crops is essential to increasing agricultural profitability and production. Our goal in this project is to create a website for crop recommendations in the agricultural industry. Using machine learning techniques to help farmers make informed crop selection selections is the main goal of this research. Our goal is to deliver farmers customized suggestions based on their unique soil properties, environmental circumstances, and past production data by utilizing machine learning algorithms. We collect and analyze data from a variety of sources, such as soil databases, climate records, and crop performance data, in order achieve this goal. These datasets form the basis for developing and assessing machine learning models that can forecast which crops would perform best under particular circumstances. For selecting the best model for crop selection, our approach compares and assesses a number of machine learning techniques, including Decision Tree, Guassian Naive Bayes, Support Vector Machine (SVM) and Random Forest. The website's foundation incorporates models that show reasonable accuracy and reliability, so ensuring farmers' ease of access and usability. In addition, the website is designed to be deployed on cloud platforms, improving the scalability and accessibility of the recommendation system. This makes it easier for farmers to access easy guidance services anywhere and provide them with the insights needed to make better crop choices This service proposes to modify crop recommendations by integrating advanced machine learning algorithms with web -Weaker grassroots communication -And improved livelihoods Through continuous improvement and refinement of our recommendation systems, we hope to equip farmers with practical tools to help them meet the challenges faced by modern agriculture to cultivate good crops.

* 1. **Motivation**

The impact of climate change on agriculture has become more apparent in recent years. Globally, farmers face major challenges due to extreme weather, unpredictable rainfall and climate change. Furthermore, inappropriate crop selection is often due to limited availability of improved agricultural techniques.The impact of climate change on agriculture has become more apparent in recent years.

*School of Computer Engineering, KIIT, BBSR* 2

*CROP RECOMMENDATION*

Globally, farmers face major challenges due to extreme weather, unpredictable rainfall and climate change. Furthermore, inappropriate crop selection is often due to limited availability of improved agricultural techniques.Farmers often depend on their own experiences and common information, both of which can be limited and capable of inaccuracy. They might therefore find it difficult to choose crops wisely, which could result in lower yields and economic losses. It is important to use technology wisely in order to tackle these issues. We can give farmers access to cutting-edge crop selection by using the power of machine learning and web-based platforms. By making agricultural knowledge available to everyone with a smartphone and an internet connection, these technologies have the potential to make accessible the field. It help farmers to optimize agricultural operations and make data-driven decisions regarding crop selection by combining machine learning algorithms with user-friendly online interfaces. We can increase agricultural productivity, reduce the effects of climate change, and eventually improve the standard of living for farmers worldwide by widely deploying these technologies.

* 1. **Literature Review**

The use of machine learning (ML) in agriculture has revolutionized how farmers predict crop yields and make decisions. Researchers have invested considerable effort in studying various ML algorithms for their effectiveness in this field. Since Random Forest continuously predicts crop yields with high accuracy, it gets the most accuracy among these algorithms [1][4]. Its ability to handle large datasets with numerous input variables makes it a reliable choice for modeling complex data [2]. In addition to predicting outputs, machine learning (ML) is essential for developing recommendation systems for farmers. These systems try to maximize yields by utilizing historical data as well as present conditions to provide farmers with useful advice on which crops to plant [3]. The integration of ML into applications has further facilitated the accessibility of such recommendations, enabling farmers to make informed decisions promptly [3]. The literature also highlights the significance of extensive datasets that include a range of crop growth-influencing variables, including soil type, weather, and farming techniques. The predictive performance of ML models is heavily dependent on the quality and scope of the data they are trained on [1]. As such, ongoing efforts to collect and analyze agricultural data are crucial for the continuous improvement of these models [2]. The application of ML in crop recommendation systems represents a transformative shift towards precision agriculture. By utilizing the predictive capabilities of algorithms such as Random Forest, these systems serve to promote sustainable farming practices in addition to augmenting crop yields [4].

*School of Computer Engineering, KIIT, BBSR* 3

*CROP RECOMMENDATION*

**2.3 Project Goals and Objectives**

The main aim of our project is to use machine learning to transform how crops are recommended and deployed in agriculture. By providing farmers with exact guidance that is customized to their particular climate and soil conditions, we want to assist them improve crop yield. We want to enable farmers to make better choices through data-driven decision-making, which should increase overall agricultural productivity. To achieve this, we explore various machine learning algorithms such as Decision Tree, Gaussian Naive Bayes, Support Vector Machine (SVM), and Random Forest. We will evaluate these algorithms based on their accuracy, precision and recall. Once we’ve identified a model that works well, we focus on making it even better and creating easy-to-use, easy-to-use web applications. Scalability and accessibility are important for our system to effectively handle a large number of users and handle many requests. One of our main objectives is sustainability. Our goal is to encourage methods that will enable farmers to maximize crop yields while conserving resources, so fostering a more sustainable agricultural system. Additionally, in order encourage interaction and have a lasting effect on agriculture, we think it is important to share our knowledge with relevant parties. Through this effort, we hope to ultimately enhance environmental sustainability, economic development, and food security.

**2.4 Model Development**

In the Model Development section of our project, Several classification algorithms were utilized to address the assigned work. Among these algorithms are Random Forest, Support Vector Machine (SVM), Gaussian Naive Bayes, and Decision Tree. These algorithms are all appropriate for our goals because they each have special qualities and strengths.

**Decision Tree :**

Decision trees are useful for providing insights into the decision-making process since they are simple to understand and intuitive. In the end, they create a tree-like structure where each node indicates a choice based on a feature by dividing the data into smaller groups based on feature values.

*School of Computer Engineering, KIIT, BBSR* 4

*CROP RECOMMENDATION*

**Gaussian Naive Bayes :**

This algorithm is based on Bayes' theorem and assumes that features are independent of each other, given the class label. Gaussian Naive Bayes, despite its simplistic premise, frequently delivers good results in reality, particularly when dealing with high-dimensional data.

**Support Vector Machine (SVM) :**

SVM is a powerful algorithm for both classification and regression tasks. A practical approach is to find the hyperplane in the feature space that best partitions the classes. Because SVM can use kernel operations to translate data into a higher dimensional space where separation is possible, it is particularly useful when the data cannot be separated linearly.

**Random Forest :**

Random forest is a cluster learning method that consists of multiple decision trees and combines their predictions into a comprehensive classification. It is known for its robustness and trouble-free handling of high-dimensional data. Additionally, random forests provide insights into contextual factors, which can help to understand change strategies affecting crop production.

*School of Computer Engineering, KIIT, BBSR* 5

*CROP RECOMMENDATION*

1. **Methodology**
   1. **System Architecture**

The Web server and the ML container are the two major parts of the system architecture. To store and retrieve data, the Web server communicates with a database, usually a SQL database. This architecture is designed to effectively manage a large number of requests. Its scalability and deployability enable it to smoothly integrate into the current infrastructure while also accommodating growing demands.

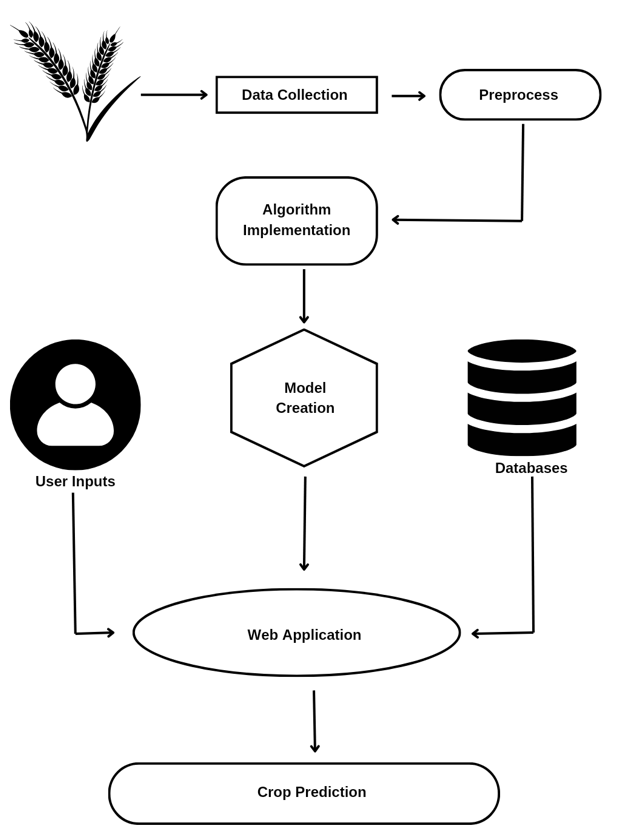


Figure 3.1 : System Architecture

*School of Computer Engineering, KIIT, BBSR* 6

*CROP RECOMMENDATION*

* 1. **Datasets**

The dataset consists of 2200 rows, with each row containing 8 columns representing various environmental and soil parameters. The first three columns show the values of nitrogen, phosphorus and potassium (NPK) in the soil, followed by the columns showing temperature, humidity, pH and precipitation, which are the environmental parameters an average value

The Nitrogen, Phosphorous, and Potassium levels in the soil are crucial indicators of soil fertility, while Temperature, Humidity, and Rainfall are essential environmental factors affecting crop growth. The pH value indicates the acidity or alkalinity of the soil, influencing nutrient availability to plants.

The final column, labeled as "Label," provides information on the type of crop best suited to grow under the given conditions. This column serves as the target variable or the value we aim to predict using machine learning algorithms. By analyzing the relationships between these input features (NPK values, environmental conditions) and the corresponding crop types (Label), we can develop a predictive model to recommend the most suitable crops for specific soil and environmental conditions.

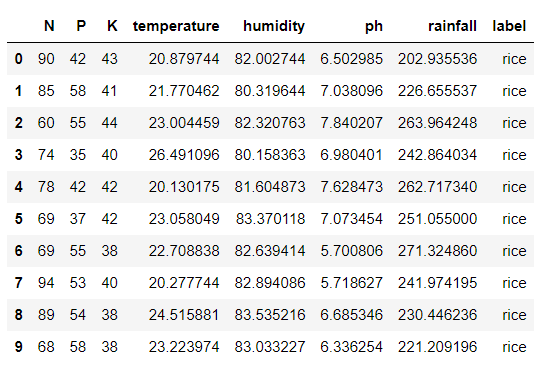


Figure 3.2 : Random data from the dataset for crop recommendation

*School of Computer Engineering, KIIT, BBSR* 7

*CROP RECOMMENDATION*

* 1. **Crop Recommendation**

Crop recommendation is a classification task using machine learning algorithms to classify plant species based on specific features. In our case, the classification criteria include soil NPK (nitrogen, phosphorus, potassium) value, ambient temperature, soil pH, soil pH, and spatial rainfall.

To accomplish this task, we trained our dataset using four standard machine learning algorithms:

* Random Forest
* Gaussian Naive Bayes
* Support Vector Machine (SVM)
* Decision Tree

These algorithms were selected because they are known to perform well on labeled datasets for classification problems, particularly with the features available in our dataset. Additionally, they offer relatively fast runtime, making them suitable choices for our classification task.

In the observations section, we provide the accuracy of each algorithm, indicating how well they performed in categorizing the plants based on the given features and labeled dataset. This information helps evaluate the effectiveness of each algorithm in the crop recommendation process.

* 1. **Analysis Tools and Technologies**

In the Analysis Tools and Technologies section, we utilized a range of tools and technologies to address both aspects of the project: the web and the machine learning components.

For the machine learning aspect, we employed:

* Python: A versatile programming language widely used in machine learning for its rich ecosystem of libraries and ease of use.
* NumPy: A fundamental package for scientific computing in Python, providing support for large, multi-dimensional arrays and matrices.
* Pandas: A powerful library for data manipulation and analysis, offering data structures and operations for structured data.

*School of Computer Engineering, KIIT, BBSR* 8

*CROP RECOMMENDATION*

For the web component, we utilized:

* Streamlit: A popular framework for building interactive web applications with Python, allowing us to create intuitive user interfaces for presenting and interacting with machine learning models.
* HTML, CSS, JS (JavaScript): Standard web technologies used for designing and styling the user interface of the web application, enhancing its appearance and interactivity.

*School of Computer Engineering, KIIT, BBSR* 9

*CROP RECOMMENDATION*

1. **Result**

**4.1 Model Performance**

In the Model Results section, we delved into an exploration of various algorithms to determine the optimal model for crop prediction. The algorithms under scrutiny included Decision Tree, Gaussian Naive Bayes, Support Vector Machine (SVM), and Random Forest. Through a series of experiments, we rigorously evaluated the performance of each algorithm using diverse criteria. After analysis, Random Forest emerged as the standout performer among all the models examined. Its accuracy, as revealed by the data, surpassed that of the other algorithms tested. This outcome underscores the effectiveness of Random Forest in accurately predicting crop outcomes based on the given parameters. Further details regarding the comparative analysis and evaluation metrics are provided in subsequent sections.

In Figure 3, we present a graphical representation of the accuracy comparison among the various algorithms evaluated for crop prediction. The graph provides a visual insight into how each algorithm performed relative to one another in terms of accuracy.

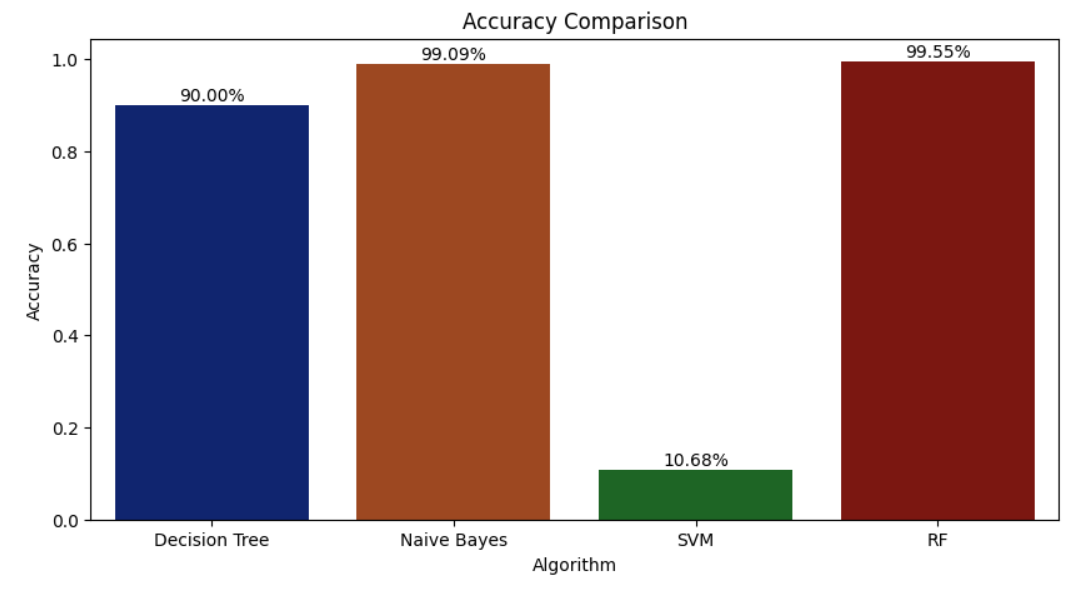


Figure 4.1 : Accuracy Comparison

*School of Computer Engineering, KIIT, BBSR 10*

*CROP RECOMMENDATION*

**4.2 Website Deployment**

We provide an overview of the outputs generated by the deployed application. The web application, developed using Streamlit, incorporates a machine learning backend to facilitate crop recommendation based on input parameters. Our intention is to deploy this application on a cloud service platform, ensuring widespread accessibility and scalability.

Figure 4 displays the landing page of the website, serving as the initial interface for users. It provides an introduction to the application's features and guides users on how to proceed.

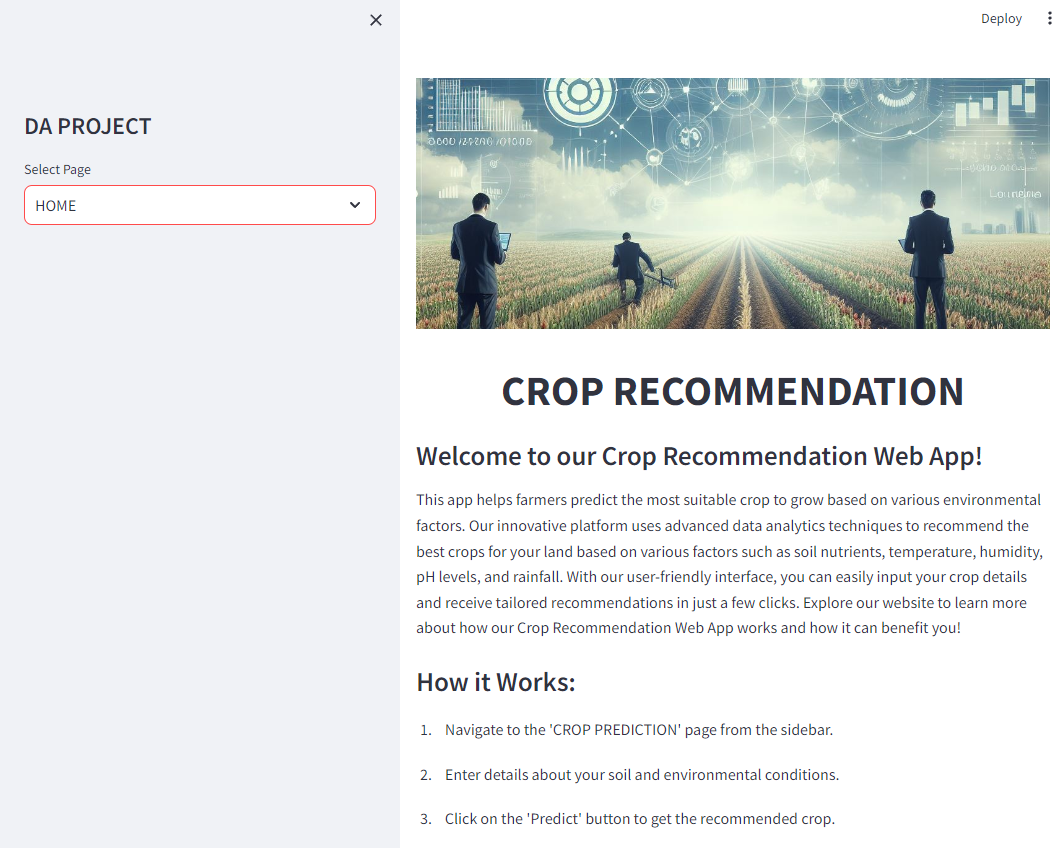


Figure 4.2 : Landing page of the website

*School of Computer Engineering, KIIT, BBSR 11*

*CROP RECOMMENDATION*

Figure 5 displays the crop recommendation outcomes generated by the deployed application. It provides insights into recommended crops based on user-input parameters. With its intuitive interface, farmers and stakeholders can access tailored crop recommendations, improving agricultural decision-making.

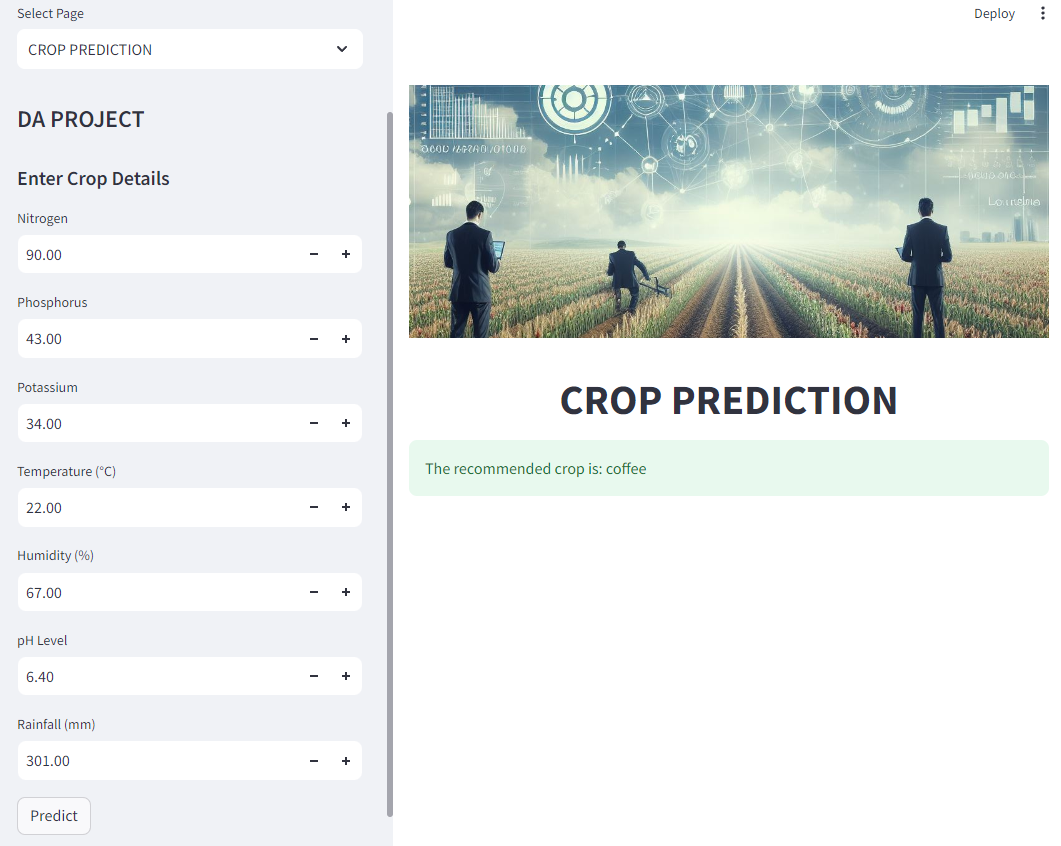


Figure 4.3 : Result of crop recommendation

*School of Computer Engineering, KIIT, BBSR 12*

*CROP RECOMMENDATION*

1. **Future Scope**

As we look ahead, there's a lot we can do with machine learning (ML) in farming. Right now, we've seen how ML can help predict crop yields and suggest what crops to grow. But there's still a whole world of possibilities out there. One thing we can do is use more kinds of information, like pictures from satellites, drones flying over fields, and special sensors in the ground. By combining all this data, we can make our predictions even more accurate and helpful for farmers. We can also make special ML models for different places where farming happens. These models would understand the unique things about each area, like what the weather is like or what kinds of soil are there. That way, the recommendations they give to farmers will be just right for where they are. It's important to make sure that everyone can use these ML tools easily, no matter where they are or what their background is. And we need to think about important stuff like who owns the data and how we keep it safe, so that everyone benefits from ML in farming, not just a few people.

*School of Computer Engineering, KIIT, BBSR 13*

*CROP RECOMMENDATION*

1. **Conclusion**

Our study on crop recommendation using machine learning (ML) gives up a potentially innovative path for the farming industry. Through our research, we have demonstrated the reliability of machine learning algorithm Random Forest in particular can be at forecasting crop yields and providing farmers with insightful advice. We studied 4 different algorithms and concluded that random forest was the best for the selected data set. The random forest achieved an accuracy of 99.55%. These technological innovations are not just technological marvels, they are also important tools that enable the farmers to make data-driven decisions that drive sustainability and production efficiency. Integrating ML into crop recommendation systems has the potential to transform agriculture by providing farmers with personalised, timely and contextual advice. This can lead to more efficient use of available resources, less negative impact on the environment, and ultimately better living conditions for the people who grow our food Improving these systems, making it more accessible and ensuring that farmers’ demands are met in the future will be critical. Our research has reached the foundation for a powerful and accessible future, but ML is not necessarily the same. This project represents a major step towards sustainable and successful agriculture where choices are guided by insights gleaned from data. We want to give farmers the skills and information they need to meet the demands of modern agriculture and harness the power of machine learning to grow better crops for a better future.

*School of Computer Engineering, KIIT, BBSR 14*

*CROP RECOMMENDATION*

***References***

[1 ] Kampalli R., & Priyadarsini K. (2021). A Review on Crop Yield Prediction Using Machine Learning Methods.

1. Jhajhariaa K., Mathura P., Jain S., & Nijhawana S. (2023). Crop Yield Prediction using Machine Learning and Deep Learning Techniques. In Proceedings of the International Conference on Machine Learning and Data Engineering.
2. Rao M. S., Singh A., Reddy N. V. S., & Acharya D. U. (2022). Crop Prediction Using Machine Learning.
3. Pande S. M., Ramesh, P. K., Anmol, A., Aishwarya, B. R., Rohilla, K., & Shaurya, K. (2021). Crop Recommender System Using Machine Learning Approach.

*School of Computer Engineering, KIIT, BBSR 15*