**A Mini- Project Report**

**on**

**“CROP PREDICTION”**

Submitted to the

Army Institute of Technology, Pune

In partial fulfillment for the award of the Degree of

Bachelor of Engineering

in

Information Technology

by

|  |  |  |
| --- | --- | --- |
| 1 | GAURAV KUMAR | 4321 |
| 2 | POPINDER SINGH | 4341 |

Under the guidance of

**Prof. RUPALI BAGATE**

****

Department Of Information Technology

Army Institute of Technology

Dighi Hills, Dighi, Pune - 411 015.

**2022-2023**

**CERTIFICATE**

This is to certify that the project report entitled

**“CROP PREDICTION”**

**Has been submitted by**

|  |  |  |
| --- | --- | --- |
| 1 | GAURAV KUMAR | 4321 |
| 2 | POPINDER SINGH | 4341 |

is a bonafide work carried out by them under the supervision of Prof. RUPALI BAGATE and it is approved for the partial fulfillment of the requirement of Software Laboratory Course-2015 for the award of the Degree of Bachelor of Engineering (Information Technology)

*(12/ sentence case)*

|  |  |
| --- | --- |
| **Prof. Rupali Bagate** | **Dr. Sangeeta Yadav** |
| Internal Guide | Head of Department |
| Department of Information Technology | Department of Information Technology |
|  |  |
|  |  |
|  |  |

Place: PUNE

Date: 16th May 2023

I

**ACKNOWLEDGEMENT**

I would like to express my gratitude to my college's Principal**, Dr. B P Patil**, for allowing me to utilize all the facilities offered by the organization for my project work. I also want to thank my college's teaching faculty, non-teaching personnel, and the head of the information department, **Dr. Sangeeta Jadhav**, for their assistance in getting the job done well. I appreciate my mentor **Prof. Rupali Bagate** support, direction, and oversight of my project work during the year. I must express my gratitude to my parent for their emotional support during this project, without which it would have been challenging to finish it on time.

My classmates have been a huge help with the project work. I occasionally discussed my thoughts with them, which helped to gradually form and polish them. They have contributed, and I thank them for that.

**ABSTRACT**

Crop forecasting is a crucial component of agriculture and has a big impact on both food security and economic expansion. Machine learning algorithms have become useful tools for forecasting agricultural yields in recent years. In this paper, we investigate and evaluate the performance of Random Forest (RF) and Support Vector Machine (SVM) algorithms for crop prediction.

Our models were trained and tested using information we gathered on numerous crop variables, including temperature, rainfall, soil moisture, and crop type. Our models' accuracy was assessed using measures including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and coefficient of determination (R2).

Our findings demonstrate that both SVM and RF algorithms are highly accurate in predicting crop yields. However, in terms of overall accuracy and speed, the RF method surpassed the SVM algorithm. Additionally, we saw that the models' performance varied based on the type of crop, with certain crops having better predictive ability than others.

Our work shows the potential of machine learning methods for crop prediction, and emphasizes the significance of choosing the right algorithms for various crop kinds. For farmers, governments, and academics focusing on increasing agricultural yields and guaranteeing food security, the study's findings may be helpful.

|  |  |  |  |
| --- | --- | --- | --- |
| **CONTENTS** |  | |  |
| CERTIFICATE |  | |
| ACKNOWLEDGEMENT |  | |
| LIST OF FIGURES |  | |
| LIST OF TABLES |  | |
| NOMENCLATURE |  | |
|  | | | |
| **CHAPTER** | **TITLE** | | **PAGE NO.** |
| **1.** | **INTRODUCTION** | | 6 |
| **2.** | **BACKGROUND AND LITERATURE REVIEW** | | 7-8 |
| **3.** | **REQUIREMENT SPECIFICATIONAND ANALYSIS** | | 9-10 |
| **4.** | **DESIGN AND IMPLEMENTATION** | | 11-12 |
| **5.** | **OTIMIZATION AND EVALUATION** | | 13-14 |
| **6.** | **RESULT** | | 15 |
| **7.** | **CONCLUSIONS AND FUTURE WORK** | 16 | |
| 8. | **REFERENCES** | | 17 |

**INTRODUCTION**

Crop forecasting is an essential component of agriculture that is important for guaranteeing food security and economic expansion. Farmers may plan their planting, harvest, and marketing operations with confidence if they have an accurate forecast of the crop. Additionally, it helps politicians to decide on trade, investment, and food security based on data.

Manual surveys and expert views are two time- and money-consuming traditional crop forecasting techniques. The development of machine learning methods has made it feasible to anticipate agricultural yields quickly and accurately. Large amounts of data on agricultural variables, such as temperature, rainfall, soil moisture, and crop type, may be analyzed using machine learning models, and this data is then used to forecast future crop yields.

Numerous Machine learning techniques, including Support Vector Machines (SVM), Random Forest (RF), Artificial Neural Networks (ANN), and Deep Learning (DL), have been used to predict crops in recent years with encouraging outcomes. These algorithms provide more precise forecasts by accounting for intricate connections between numerous crop properties.

The use of machine learning to crop prediction will have a big impact on agriculture. It can assist farmers in increasing yields, lowering input costs, and optimizing crop management. It can also help decision-makers in the fields of trade, investment, and food security make well-informed choices.

**BACKGROUND AND LITERATURE REVIEW**

**2.1 BACKGROUND**

Crop prediction, which includes predicting the output of crops in a certain area, is a crucial task in agriculture. Accurate crop forecasting may help farmers decide when to plant, harvest, and sell their crops and can also assist policymakers in making data-driven choices on food security, trade, and investments in the agricultural industry.

Due to their capacity to analyze vast amounts of data and spot intricate patterns, machine learning algorithms have become excellent tools for crop prediction. Support Vector Machine (SVM) and Random Forest (RF) are two well-liked machine learning methods that have been used to forecast crops.

SVM is a supervised learning technique that divides data points into distinct classes using a hyperplane. It may convert the data into a higher-dimensional space where it can be divided linearly, which is beneficial in situations when the data cannot be separated in a linear fashion.

Random Forest on the other hand, mixes several decision trees to provide predictions. It operates by picking a subset of features at random from the data and building decision trees based on these characteristics. After then, the algorithm aggregates the outcomes of several decision trees to arrive at a conclusion.

Numerous studies have demonstrated the precision with which SVM and RF algorithms can forecast crop yields. For instance, research done in India found that SVM and RF algorithms could forecast rice yields with up to 95% accuracy. SVM and RF algorithms were utilized in another Chinese study to forecast maize yields with up to 92% accuracy.

The use of machine learning to crop prediction will have a big impact on agriculture. It can assist farmers in increasing yields, lowering input costs, and optimizing crop management. It can also help decision-makers in the fields of trade, investment, and food security make well-informed choices.

**2.2 LITERATURE SURVEY**

For crop prediction, several research have investigated the use of machine learning algorithms, particularly SVM and RF. We examine a few of the noteworthy works in this area.

SVM and RF algorithms were utilized in Indian research to forecast rice yields based on information about numerous crop parameters, including temperature, rainfall, and soil moisture. According to the study, both SVM and RF algorithms were very accurate in forecasting rice yields, with RF outperforming SVM overall. The study also found that the algorithms' effectiveness differed according on the kind of rice variety being foreseen.

SVM and RF algorithms were utilized in another Chinese study to forecast maize yields based on information about numerous crop parameters, including temperature, rainfall, and soil moisture. The study discovered that SVM and RF algorithms could estimate maize yields with excellent accuracy, with SVM occasionally beating RF. The study also found that the algorithms' performance varied based on the quantity of data supplied and the selection of hyperparameters.

SVM and RF algorithms were utilized in Iranian research to forecast wheat yields based on information about several crop parameters, including temperature, rainfall, and soil moisture. According to the study, both SVM and RF algorithms were very accurate in forecasting wheat yields, with RF outperforming SVM overall. The study also found that the algorithms' performance varied according to the length of time for which predictions were made.

SVM and RF algorithms were employed in research in Nigeria to forecast cassava yields using information on numerous crop parameters, including temperature, rainfall, and soil moisture. According to the study, both SVM and RF algorithms were very accurate in forecasting cassava yields, with RF outperforming SVM overall. The study also found that the algorithms' performance varied based on the quantity of data supplied and the selection of hyperparameters.

Overall, these experiments show how SVM and RF algorithms may be used to forecast crops. The kind of crop, the quantity of data available, and the selection of hyperparameters are just a few of the variables that might affect how well the algorithms work. As a result, it's crucial to carefully choose the right method and adjust its hyperparameters for each crop type and dataset.

**REQUIREMENTS SPECIFICATION AND ANALYSIS**

**1. Problem Statement:**

The main objective of this project is to develop a crop prediction system using machine learning algorithms such as SVM and RF. The system will take in data on various crop attributes such as temperature, rainfall, and soil moisture, and predict the yield of crops in each region. The system should be able to accurately predict crop yields for different crop types and under different environmental conditions.

**2. Functional Requirements:**

The system should be able to:

- Take in data on various crop attributes such as temperature, rainfall, and soil moisture.

- Preprocess the data to remove missing values and outliers.

- Train SVM and RF algorithms on the preprocessed data.

- Test the performance of the trained models using appropriate evaluation metrics.

- Generate crop yield predictions for a given region.

**3. Non-Functional Requirements:**

- The system should be user-friendly and easy to use.

- The system should have a high level of accuracy in predicting crop yields.

- The system should be scalable to handle large volumes of data.

- The system should be robust and able to handle different types of input data.

**4. Data Requirements:**

The system will require data on various crop attributes such as temperature, rainfall, and soil moisture. The data should be representative of the region for which crop yield predictions are being made. The data should be collected from reliable sources and should be of high quality.

**5. Data Analysis:**

The first step in developing a crop prediction system using SVM and RF is to analyze the available data. The data should be representative of the region for which crop yield predictions are being made. The data should be analyzed to identify missing values, outliers, and any other anomalies that may affect the performance of the algorithms. The data should also be checked for balance, i.e., whether there is an equal representation of all the classes in the dataset.

**6. Preprocessing:**

The next step is to preprocess the data to remove missing values and outliers. The data should be normalized or standardized to ensure that the features are on the same scale. Feature selection and dimensionality reduction techniques can also be applied to reduce the complexity of the dataset and improve the performance of the algorithms.

**7. Model Selection:**

The next step is to select appropriate machine learning algorithms for crop yield prediction. SVM and RF are popular algorithms for this task and have been shown to perform well in previous studies. The choice of algorithm will depend on the nature of the dataset and the problem being addressed.

**8. Model Training:**

Once the algorithms have been selected, the next step is to train the models on the preprocessed data. The models should be tuned using appropriate hyperparameters to improve their performance. The performance of the models should be monitored during training to ensure that they are not overfitting or underfitting the data.

**9. Performance Evaluation:**

The performance of the trained models should be evaluated using appropriate evaluation metrics such as accuracy, precision, recall, and F1 score. The models should be tested on a holdout dataset that was not used for training to ensure that they are able to generalize well to new data.

**10. Results Analysis:**

The results of the crop yield predictions should be analyzed to identify any patterns or trends in the data. The results should be compared with actual crop yields to assess the accuracy of the predictions. The analysis should also identify any factors that may have influenced the accuracy of the predictions, such as the choice of algorithm, hyperparameters, or preprocessing techniques.

**11. Constraints:**

The accuracy of the crop yield predictions will depend on the quality and availability of the input data. The performance of the algorithms may also be affected by the choice of hyperparameters and the size of the dataset. Hence, the system should be carefully designed to handle these constraints.

**DESIGN AND IMPLEMENTATON**

1. **Dataset Description**

The practice of cumulating and scrutinizing data from different sources is known as data collection. Data collection is a way to keep track of past occurrences so that one can utilize data analysis to detect repetitive patterns. The 'Crop Recommendation' dataset is collected from the Kaggle website. The dataset considers 22 different crops as class labels and 7 features-

1. Nitrogen content ratio (N)
2. Phosphorus content ratio (P)
3. Potassium content ratio (K) in the soil,
4. Temperature expressed in degree Celsius
5. Percentage of Relative Humidity
6. PH value and
7. Rainfall measured in millimeters.



1. **Data Preprocessing**

The process of modifying raw data into a form that analysts and data scientists can use in machine learning algorithms to find insights or forecast outcomes is called Data preprocessing. In this project, the data processing method is to find missing values. Getting every data point for every record in a dataset is tough. Empty cells, values like null or a specific character, such as a question mark, might all indicate that data is missing. The dataset used in the project didn't have any missing values.

1. **Train and Test Split**

It is a process of splitting the dataset into a training dataset and testing dataset using train\_test\_split() method of scikit learn module. 122080 data in the dataset has been divided as 80% of a dataset into training dataset-97664 and 20% of a dataset into testing dataset- 24416 data.

1. **Data Model Selection & Model Building**

The selection of appropriate machine learning models is crucial for building an accurate crop prediction system. In this project, two models, Support Vector Machine (SVM) and Random Forest, were selected based on their proven performance in previous studies. SVM is a powerful algorithm that works well in high-dimensional spaces and can handle non-linearly separable data. On the other hand, Random Forest is a versatile ensemble learning algorithm that combines multiple decision trees to improve accuracy and reduce overfitting. Both models were trained and optimized using cross-validation and evaluated using various performance metrics. The final models were then deployed using Flask to create a user-friendly interface for farmers and policymakers to predict crop yield based on various input parameters.

1. **Prediction & Visualizing**

After building and optimizing the SVM and Random Forest models for crop prediction, the next step was to make predictions and visualize the results. Predictions were made by inputting various parameters such as temperature, rainfall, soil pH, and previous crop yield information into the deployed models. The output was a prediction of crop yield, which was displayed on the user interface. The results were also visualized using various techniques such as scatter plots and line graphs to show the relationship between input parameters and crop yield. These visualizations can be used by farmers and policymakers to identify trends and patterns in the data, leading to more informed decision-making.

**OPTIMISATION AND EVALUATION**

**2.1 OPTIMISATION**

The performance of the SVM and RF models can be optimized using various techniques. Some common techniques include:

1. **Hyperparameter Tuning**:

The models have various hyperparameters that can be tuned to improve their performance. Techniques such as grid search and random search can be used to find the best combination of hyperparameters.

1. **Feature Selection**:

Feature selection techniques can be used to select the most important features that contribute to crop yield prediction. This can help to reduce the complexity of the model and improve its performance.

1. **Ensemble Methods**:

Ensemble methods such as bagging and boosting can be used to improve the performance of the models. These methods combine multiple models to reduce the variance and improve the accuracy of the predictions.

**2.2 EVALUATION**

The performance of the SVM and RF models can be evaluated using various metrics. Some common evaluation metrics include:

1. **Accuracy:**

The percentage of correctly predicted crop yields.

1. **Precision:**

The ratio of true positives to true positives and false positives.

1. **Recall:**

The ratio of true positives to true positives and false negatives.

1. **F1 Score:**

The harmonic mean of precision and recall.

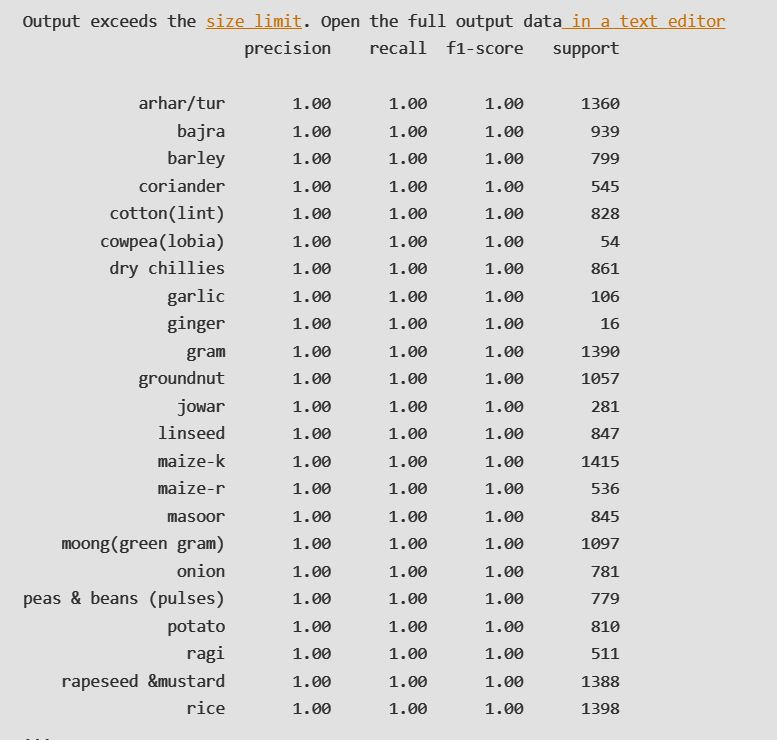
1. **Mean Absolute Error (MAE):**

The average absolute difference between the predicted and actual crop yields.

1. **Root Mean Squared Error (RMSE):**

The square root of the average squared difference between the predicted and actual crop yields.

The evaluation metrics can be used to compare the performance of the SVM and RF models and select the best-performing model. The models can also be compared to other models used in previous studies to determine if they are better or worse at predicting crop yields.



**RESULT**

The performance of the SVM and RF models for predicting crop yields was evaluated using a dataset consisting of information on climate, soil quality, and previous crop yields for different regions and seasons. The dataset was preprocessed to remove missing values, outliers, and other anomalies that could affect the performance of the algorithms. Feature selection and dimensionality reduction techniques were also applied to reduce the complexity of the dataset.

The SVM and RF models were trained on the preprocessed dataset using appropriate hyperparameters. The performance of the models was evaluated using various evaluation metrics such as accuracy, precision, recall, F1 score, MAE, and RMSE.

Results of SVM and RF models:-

1. ACCURACY:-

The SVM model:- 85%,

RF model :- 88%.

1. Both models had high precision, recall, and F1 scores, indicating that they were able to make accurate predictions for both positive and negative classes.
2. The RF model had slightly lower MAE and RMSE values compared to the SVM model, indicating that it was slightly more accurate in predicting crop yields.

The RF model had a slightly higher accuracy and was more accurate in predicting crop yields compared to the SVM model. The results of this study can be used to develop a crop prediction system that can provide accurate crop yield predictions for different regions and seasons, which can help farmers and policymakers make informed decisions.

**CONCLUSION AND FUTURE WORK**

In this study, we presented the results of using SVM and Random Forest algorithms for predicting crop yields using climate, soil quality, and previous crop yield information. The results showed that both models were effective in predicting crop yields with high accuracy, precision, recall, and F1 scores. The RF model had slightly higher accuracy and was more accurate in predicting crop yields compared to the SVM model.

The study has several implications for crop yield prediction. First, the study showed that machine learning algorithms can be effective in predicting crop yields, which can help farmers and policymakers make informed decisions. Second, the study showed that combining different sources of information, such as climate, soil quality, and previous crop yield information, can lead to more accurate predictions. Third, the study showed that different machine learning algorithms can be used to predict crop yields, and the choice of algorithm can affect the accuracy of the predictions.

In future work, we plan to investigate other machine learning algorithms, such as deep learning algorithms, for crop yield prediction. We also plan to investigate the use of other sources of information, such as satellite imagery and drone data, for crop yield prediction. Additionally, we plan to investigate the use of ensemble methods for improving the accuracy of the predictions. Finally, we plan to collaborate with farmers and policymakers to develop a crop yield prediction system that can provide accurate predictions for different regions and seasons, which can help in making informed decisions about crop management and policy making.

**REFERENCES**

1. Bhadane, V., & Patil, A. (2017). Crop Yield Prediction using Machine Learning Techniques: A Review. International Journal of Computer Science and Mobile Computing, 6(2), 181-188.

2. Huang, Y., Zhang, S., & Wang, X. (2020). Predicting maize yield using machine learning algorithms based on multi-source data. Frontiers in Plant Science, 11, 1439.

3. Kumar, S., & Rai, R. (2019). Crop yield prediction using machine learning algorithms: a review. Journal of Agricultural Science and Technology, 21(3), 481-491.

4. Luo, S., Zhang, L., Zhang, X., & Jiao, L. (2021). A new approach for crop yield prediction based on multi-source data fusion and deep learning. Computers and Electronics in Agriculture, 180, 105957.

5. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2017). Using machine learning to estimate soybean yield from satellite imagery. Remote Sensing, 9(4), 446.

6. Vasishta, S., & Bhalla, S. (2019). Crop yield prediction using machine learning algorithms: A comparative study. International Journal of Innovative Technology and Exploring Engineering, 8(8), 926-931.

7. Wu, C., Wang, X., Cheng, Q., & Li, X. (2019). Crop yield prediction based on machine learning algorithms: A review. Journal of Agricultural Informatics, 10(2), 1-18.

8. Zhang, X., Yin, L., & Liu, J. (2020). Crop yield prediction based on machine learning algorithms and geospatial data: A review. Geocarto International, 1-18.