IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

CROP RECOMMENDATION AND WEATHER FORECASTING USING MACHINE LEARNING

Sayyam Oswal, Kiran Patil, Sughosh S MutalikDesai, Raghavendra, Prof.Rakesh.J.Kadakol

ISE, KLS Gogte Institute of Technology, Belagum

***Abstract*—** Agriculture faces significant challenges in ensuring sustainable and resilient crop production amidst a changing climate and unpredictable weather patterns. Accurate crop suitability assessment and reliable weather forecasting play crucial roles in optimizing agricultural practices and enhancing productivity. Machine learning (ML) has emerged as a powerful tool for addressing these challenges by leveraging vast amounts of historical and real-time data to identify patterns and make predictions.To address this challenge, this paper presents a novel approach in the form of a user-friendly yield prediction system for farmers. The proposed system leverages mobile applications or websites to connect with farmers, utilizing GPS technology for location identification and allowing users to input area and soil type data. Employing various machine learning algorithms including Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest (RF), Multivariate Linear Regression (MLR), and K-Nearest Neighbor (KNN), the system predicts crop yields or recommends the most profitable crops based on user inputs. Experimental results reveal that the Random Forest algorithm achieves the highest accuracy of 95%. With agriculture's historical significance in India and its current ranking as the second-largest producer globally, this research emphasizes the critical need for improving crop yield and agricultural productivity. By providing farmers with advanced decision-making tools, such as the proposed Crop Recommendation System, we aim to mitigate uncertainties arising from fluctuating market prices and empower farmers towards more sustainable and profitable agricultural practices, ultimately contributing to India's socio-economic development and well-being.

Keywords— *GPS Technology, Accuracy, Agricultural Productivity, Socio-Economic Development.*

I. INTRODUCTION

Agriculture has long been the backbone of India's economy, providing sustenance and livelihoods to a significant portion of its population. With its allied sectors, agriculture contributes substantially to the country's Gross Domestic Product (GDP) and serves as a vital source of income for rural communities.

Agriculture has an extensive history in India. Recently, India is ranked second in the farm output worldwide. Agriculture-related industries such as forestry and fisheries contributed for 16.6% of 2009 GDP and around 50% of the total workforce. Agriculture's monetary contribution to India's GDP is decreasing. The crop yield is the significant factor contributing in agricultural monetary. The crop yield depends on multiple factors such as climatic, geographic, organic, and financial elements. It is difficult for farmers to decide when and which crops to plant because of fluctuating market prices. Citing to Wikipedia figures India's suicide rate ranges from 1.4-1.8% per 100,000 populations, over the last 10 years. Farmers are unaware of which crop to grow, and what is the right time and place to start due to uncertainty in climatic conditions. The usage of various fertilizers is also uncertain due to changes in seasonal climatic conditions and basic assets such as soil, water, and air. In this scenario, the crop yield rate is steadily declining. The solution to the problem is to provide a smart user-friendly recommender system to the farmers

In this paper, we present the design, implementation, and evaluation of the Crop Recommendation System, highlighting its potential to revolutionize agricultural practices in India. By empowering farmers with actionable insights and recommendations, we envision a future where sustainable and profitable agriculture becomes the norm, ultimately leading to the socio-economic upliftment of rural communities and the overall advancement of the agricultural sector in India. The proposed methodology incorporates various ML algorithms, including support vector machines (SVMs), random forests (RFs), and neural networks (NNs), to analyze historical meteorological data, soil parameters, and crop yield records. These models are trained to identify the optimal environmental conditions for different crops and predict future weather patterns.

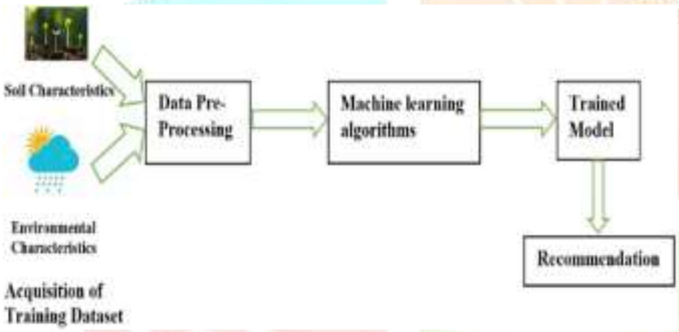

FIG. 1 : Generic Block Diagram

II. LITRATURE SURVEY

Umamaheswari S, Sreeram S, [1]: Blockchain's most basic promise for the agriculture industry is that it removes the need for third parties otherwise required to ensure trust within buyer-seller relationships, or for that matter any source-destination relationship. In an environment enabled by blockchain technology, transactions become peer-to-peer with no use for intermediaries.

Manjula E. [2] : Data Mining is emerging research field in crop yield analysis. Yield prediction is a very important issue in agricultural. Any farmer is interested in knowing how much yield he is about to expect. In the past, yield prediction was performed by considering farmer's experience on particular field and crop. The yield prediction is a major issue that remains to be solved based on available data. Data mining techniques are the better choice for this purpose.

Verdouw C, Bogaardt MJ.[3]: Smart Farming is a development that emphasizes the use of information and communication technology in the cyber-physical farm management cycle. New technologies such as the Internet of Things and Cloud Computing are expected to leverage this development and introduce more robots and artificial intelligence in farming. This is encompassed by the phenomenon of Big Data, massive volumes of data with a wide variety that can be captured, analysed and used for decision-making.

Wolfert S, Ge L [4]: Big data are being used to provide predictive insights in farming operations, drive real-time operational decisions, and redesign business processes for game-changing business models. Several authors therefore suggest that Big Data will cause major shifts in roles and power relations among different players in current food supply chain networks. The landscape of stakeholders exhibits an interesting game between powerful tech companies, venture capitalists and often small start-ups and new entrants. At the same time there are several public institutions that publish open data, under the condition that the privacy of persons must be guaranteed

Kritika N, Prasanth DJ [5]: Blockchain can be of great help in this sector. A transparent and trusted system can be built by putting all the information about agricultural events on a blockchain. Farmers can also get instant data related to the seed quality, climate environment related data, payments, soil moisture, demand and sale price, etc. all on a single platform.The intent of this project is to store the sensor data in a blockchain and build a smart contract deployed in the Ethereum blockchain to facilitate buying and selling of crops and land.

Djodiltachoumy S [6]: The yield prediction is a major issue that remains to be solved based on available data. Data mining techniques are the better choice for this purpose. Different Data Mining techniques are used and evaluated in agriculture for estimating the future year's crop production. This research proposes and implements a system to predict crop yield from previous data. This is achieved by applying association rule mining on agriculture data. This research focuses on creation of a prediction model which may be used to future prediction of crop yield. This paper presents a brief analysis of crop yield prediction using data mining technique based on association rules for the selected region i.e. district of Tamil Nadu in India. The experimental results shows that the proposed work efficiently predict the crop yield.

III. METHODOLOGY

1) ***Problem Statement:*** *Crop Suitability and Weather Forecasting Recommendation System using – A Machine learning Approach.*

Recent advancements leverage accelerometers and gyroscope sensors with machine learning to distinguish between normal conditions and accidents effectively. the design, implementation, and evaluation of the Crop Recommendation System, highlighting its potential to revolutionize agricultural practices in India. By empowering farmers with actionable insights and recommendations, we envision a future where sustainable and profitable agriculture becomes the norm, ultimately leading to the socio-economic upliftment of rural communities and the overall advancement of the agricultural sector in India.

2) ***Data Collection:***

In addition to the methods mentioned, data can also be collected through surveys, questionnaires, APIs, and data feeds. Data collection involves gathering relevant information from various sources, such as social media platforms, databases, public records, or sensor devices.

3) ***Data Storing:***

Data can be stored in various formats such as CSV, JSON, Excel, databases (e.g., MySQL, PostgreSQL), or cloud-based storage systems like Amazon S3 or Google Cloud Storage. Proper data storage practices involve organizing and structuring the data in a way that allows for efficient retrieval and analysis.

4) ***Feature Engineering:***

Feature engineering is the process of transforming raw data into meaningful features that can improve the performance of machine learning models.It includes tasks such as handling missing values (e.g., imputation techniques), dealing with skewed data (e.g., logarithmic transformation), performing exploratory data analysis (EDA), applying encoding techniques (e.g., one-hot encoding, label encoding), and addressing issues like scaling, normalization, standardization, handling duplicate values, handling class imbalance in datasets, and identifying outliers.

5) ***Feature Selection:***

Feature selection aims to identify the most relevant and informative features for the machine learning model. Techniques like correlation analysis and covariance can help identify the relationships between features. Libraries like scikit-learn provide methods such as Recursive Feature Elimination (RFE), SelectKBest, and Principal Component Analysis (PCA) for feature selection. Feature selection helps reduce dimensionality, improve model interpretability, and enhance model performance.

6) ***Model Creation:***

Model creation involves selecting an appropriate machine learning algorithm based on the problem statement, dataset characteristics, and desired outcomes. Hyperparameter tuning is crucial to optimize the model's performance. Techniques like grid search, random search, or Bayesian optimization can be employed. Evaluation metrics like accuracy, precision, recall, F1 score, and area under the curve (AUC) are used to assess the model's performance.

7) ***Model Deployment / GUI Creation:***

After training and validating the model, it can be deployed for real-world use. Model deployment can involve building APIs or web services to provide predictions on new data or embedding the model into an application. Creating a graphical user interface (GUI) allows users to interact with the model easily, providing inputs and receiving outputs in a user-friendly manner.

8) ***Algorithm:***

Artificial Neural Network (ANN):-

1. Gather the dataset containing input features and corresponding target outputs.
2. Preprocess the data by normalizing or standardizing the input features.
3. Split the dataset into a training set and a testing/validation set.
4. Design the architecture of the Artificial Neural Network (ANN), including the number of layers, neurons, and activation functions.
5. Initialize the weights and biases of the network.
6. Train the ANN by feeding the training data through forward propagation, calculating the loss/error, and updating the weights using backpropagation with an optimization algorithm like gradient descent.
7. Iterate the training process by repeating forward propagation, error calculation, and weight updates until convergence or a specified number of epochs.

8. Evaluate the performance of the trained ANN using appropriate metrics on the testing/validation set.
9. Use the trained ANN to make predictions on new data by feeding the input features through the network and obtaining the corresponding output.

Support Vector Machine (SVM):

1. Import the libraries such as numpy and pandas.
2. Fitting the SVM Classifier to the training set.
3. Predict the test set result.
4. Creating a confusion matrix.
5. Visualizing the training set result Step-6: Visualizing the test set result.

Random Forest Algorithm:

1. Import numpy, pandas, matplotlib, and sklearn's RandomForestClassifier.
2. Split the data into training and test sets.
3. Initialize the Random Forest model by setting hyperparameters such as the number of trees in the forest, maximum depth of the trees, and number of features to consider at each split.
4. Fit the Random Forest model to the training set. Predict the target variable for the test set using the trained model.
5. Evaluate the performance of the model using appropriate metrics such as accuracy, precision, recall, or mean squared error, depending on the task.

K-Nearest Neighbors (KNN):-

1. Collect a labelled dataset.Pre-process the data and split it into training and testing sets.
2. Choose the value of k.
3. Calculate the distance between the new data point and all points in the training set.
4. Find the k nearest neighbours based on the smallest distances.
5. For classification, assign the majority class label among the neighbours as the predicted class; for regression, take the average of the target values.
6. Evaluate the model's performance on the testing set using appropriate metrics.
7. Refine the model if needed by adjusting the value of k or pre-processing techniques.

Multivariate Linear Regression (MLR):-

1. Collect the dataset with the dependent variable and multiple independent variables.
2. Preprocess the data by handling missing values, outliers, and normalizing if needed.
3. Select relevant independent variables if there are too many using techniques like correlation analysis or feature selection methods.
4. Split the dataset into a training set and a testing/validation set.
5. Train the MLR model on the training set to estimate the regression coefficients using methods like Ordinary Least Squares.
6. Evaluate the model's performance using metrics like MSE, RMSE, or R^2 on the testing/validation set.
7. Interpret the coefficients to understand the impact of the independent variables on the dependent variable.
8. Use the trained model to make predictions on new data by plugging in the values of the independent variables.
9. Refine the model if necessary by iterating through steps 3 to 8, considering different feature combinations or exploring other modeling techniques.

9) *Future Scope and Further Enhancements:*

Explore the integration of additional data sources, such as geospatial information or real time weather data, to improve the accuracy and timeliness of crop suitability and weather forecasting. Investigate advanced techniques like spatial analysis or geospatial data to account for location-specific factors that influence crop suitability.Continuously monitor and update the recommendation system with new data to adapt to changing environmental conditions and improve the accuracy of predictions.

10) *Real-Time Monitoring and Updates:*

Implement a real-time monitoring system that continuously collects and updates weather data, crop performance metrics, and other relevant variables. Utilize this real-time data to provide up-to-date recommendations and forecasts, allowing farmers to adapt their practices based on current conditions

11) *Result Analysis*

We performed modelling all the five algorithms mentioned above and saw the results as shown in the Table 5.1. The first one is ANN which produces the accuracy of 86%. Likewise, all the other algorithms work with some default as well as minimal parameters. The Random Forest Algorithm produces the highest accuracy of 95% and MLR is produced least accuracy i.e 60%.

Algorithm	Accuracy (%)
Artificial Neural Network (ANN)	86
Support Vector Machine (SVM)	75
Multivariate Linear Regression (MLR)	60
Random Forest (RF)	95
K Nearest Neighbor (KNN)	90

Table 5.1 Accuracy Score Table for various Algorithms

Fig.2 : This figure is a tabular representation of Accuracy Score for various Algorithm . Machine learning algorithms are employed to analyze the combined dataset, identifying correlations between environmental factors and crop performance.

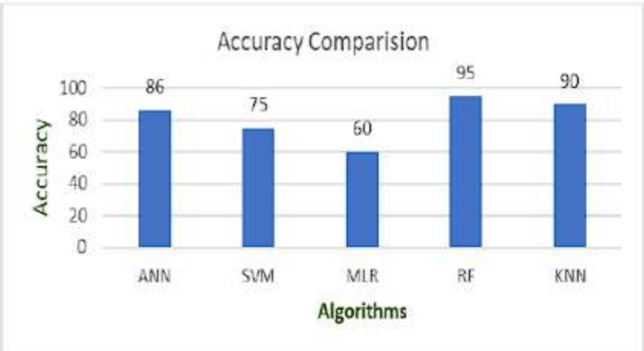

Fig.3 : This figure is a graph that represents the Accuracy scores for different models. After conducting a comparative analysis of various machine learning models for Arrhythmia detection, we evaluated their predictive performance using accuracy, precision, recall, and F1 score metrics. The objective was to determine the effectiveness of these models in correctly classifying different Crop and soil categories. And factors like computational complexity, interpretability, and scalability should be considered when selecting the most suitable model for real-world deployment.

CONCLUSION

In this machine learning project focused on crop suitability and weather forecasting recommendation system, we explored and compared multiple algorithms, including Support Vector Machines (SVM), Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN) and Random Forest. Our goal was to determine the best algorithm that would provide accurate predictions for crop suitability and weather forecasting. After thorough experimentation and evaluation, we found that Random Forest exhibited the highest accuracy among the tested algorithms. Random Forest, with its ensemble-based approach, demonstrated superior performance in predicting crop suitability and weather conditions. The algorithm's ability to handle complex relationships between features and reduce overfitting through ensemble averaging proved to be advantageous in our project. The high accuracy achieved by Random Forest indicates its potential to effectively analyze and classify the suitability of different crops based on various environmental factors and weather conditions.

Furthermore, the Random Forest algorithm's ability to handle large datasets and capture feature importance was instrumental in generating reliable recommendations for crop suitability and weather forecasting. By considering multiple decision trees and aggregating their outputs, the algorithm provided robust predictions, enhancing the reliability of our recommendation system. While SVM, ANN, and KNN also yielded promising results, Random Forest surpassed them in terms of accuracy. However, it is essential to note that the choice of the best algorithm may vary depending on the specific dataset, problem domain, and evaluation metrics employed.

REFERENCES

[1] Umamaheswari S, Sreeram S, Kritika N, Prasanth DJ, "BIoT: Blockchain-based IoT for Agriculture", 11th International Conference on Advanced Computing (ICoAC), 2019 Dec 18 (pp. 324-327). IEEE.

[2] Jain A. "Analysis of growth and instability in the area, production, yield, and price of rice in India", Journal of Social Change and Development, 2018;2:46-66

[3] M. Ajay Kumar,Sree Kamya,K. Kiranmai, Pragathi (2023),"Dynamic Accident Detection And Alert System Using Arduino",Journal of Emerging Technologies and Innovative Research (JETIR).

[4] Jones JW, Antle JM, Basso B, Boote KJ, Conant RT, Foster I, Godfray HC, Herrero M, Howitt RE, Janssen S, Keating BA, "Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. Agricultural systems", 2017 Jul 1;155:269-88

[5] Kumar R, Singh MP, Kumar P, Singh JP, "Crop Selection Method to maximize crop yield rate using a machine learning technique", International conference on smart technologies and management for computing, communication, controls, energy, and materials (ICSTM), 2015 May 6 (pp. 138-145). IEEE.

[6] Veenadhari S, Misra B, Singh CD, "Machine learning approach for forecasting crop yield based on climatic parameters", In 2014 International Conference on Computer Communication and Informatics, 2014 Jan 3 (pp. 1-5). IEEE.

[7] S. Pavani, Augusta Sophy Beulet P., "Heuristic Prediction of Crop Yield Using Machine Learning Technique", International Journal of Engineering and Advanced Technology (IJEAT), December 2019, pp (135-138).