ISSN No:-2456-2165

Crop Recommendation System for Madhya Pradesh Districts using Machine Learning

Dipti Dubey[1], Nitesh Gupta[2], Subhodhni Gupta[3], Shashi Gour[4]
[123] Department of Computer Science, SAM Global University, Bhopal, MP, India
[4] Directorate of Extension Services, JNKVV, Jabalpur, MP, India

Abstract:- Recommendation systems have become increasingly common in recent years. The use of recommender systems has become common in an assortment of commercial applications. There are software tools and techniques that provide suggestions for items of use to the user called recommender systems. Recommender systems help users to make various decisions such as what item to buy, what music to listen to, which book to read, etc.

But in the field of agriculture, the selection of crop and cropping techniques has a significant impact on the productivity and financial success of farmers. So many parameters including rainfall, soil properties, crop rotation, land preparation, and uncontrollable elements like weather, have an impact on crop recommendations are involved with uncertainty a good recommender system is required. It is unfortunate that there is no universal system to assist farmers in agriculture. Most Indian farmers cultivate at their own discretion and follow the pattern and norms of ancestral farming. Due to a lack of adequate technical knowledge, they do not get enough production and profit.

By using recommendations that are appropriate for them, the proposed agricultural recommender system will assist the farmers in minimising their losses and maximizing their profits. Based on the necessary parameters, this system helps the farmers in selecting a suitable crop for farming.

Keywords:- *Crop Recommendation, Supervised Learning, Classification Models.*

I. INTRODUCTION

Many economies around the world are based on agriculture. However, the traditional farming practices have become outdated and inefficient. With the increasing global population, there is a need to produce more food in a sustainable manner. The agriculture recommendation system is a technology that is revolutionizing farming practices by providing farmers with personalized recommendations that improve crop yield and reduce costs. An agriculture recommendation system is a software application that uses data analytics, machine learning, and artificial intelligence to provide farmers with personalized recommendations.

In previous research papers for crop recommendation, various parameters like soil nutrients (N, P & K), rainfall, temperature, humidity etc. have been used by most of the researchers. But all farmers do not have information related to these parameters so we have used simple factors like district, season, area (in ha) and soil type to recommend the crop name.

This recommendation system is made using various classification models such as logistic regression, k-nearest neighbor, decision tree, random forest and gradient boosting algorithms, which analyze the production of different crops in the previous years. The research focuses on the major crops for Madhya Pradesh stored in database system. The dataset contains details of District, Season, Area, Production, Soil Type and Average Rainfall.

Analysing these data, the algorithms give the result which recommend the preferred profitable output. For the accuracy of the prediction, data of past Fifteen years is being used by the algorithms for the purpose of learning and result analysis. In this paper, discusses the literature review in section II, discusses the methodology in section III explains. In Section IV, experimental results and analysis is explained. Lastly is given the conclusive idea of this research.

II. REVIEW OF LITERATURE

The review work carried out after the literature survey from few International journals like SCI, IEEE, some R&D Reports and books. The best selected papers for our research work are as follows.

Bondre, D. A., et al. (2019, October 1) present a novel system aimed at predicting crop yield by utilizing machine learning algorithms. Specifically, they apply algorithms such as Support Vector Machine and Random Forest on agricultural data obtained from previous harvests. Moreover, in addition to predicting crop yield, the system also provides recommendations for fertilizer that is best suited for each individual crop.[1]

Reddy, D. A , et al. (2019, February 15) introduce a method that considers three crucial parameters: soil characteristics, soil types, and crop yield data collection. By utilizing these parameters, this method recommends a suitable crop for farmers to cultivate. The proposed recommendation system employs an ensemble model with majority voting techniques, such as random tree, CHAID, K-

Nearest Neighbour, and Naive Bayes as learners. This model yields high specific accuracy and efficiency. The resulting classified image contains ground truth statistical data, including weather, crop yield, and state and district wise crops. It predicts crop yield under specific weather conditions for a given crop.[2]

Kumar, R et al.(2015, May 6) present the Crop Selection Method (CSM) as a potential solution for the crop selection problem. Their aim is to maximize the net yield rates of crops throughout the season and attain the highest level of economic growth for the country. The authors contend that the implementation of the proposed method could result in a significant increase in the net yield rates of crops.[3]

Palanivel, K., et al. (2019, January 1) discusses the use of machine learning techniques for predicting crop yield in agriculture. Various algorithms such as artificial neural networks, support vector machines, linear and logistic regression, decision trees, and Naive Bayes are used for implementing prediction. The paper proposes an approach for prediction of crop yield using machine learning techniques in big data computing paradigm.[4]

Nigam, A., et al. (2019, November 1)present a research paper with a focus on the prediction of crop yield. Various machine learning techniques are applied and subsequently compared based on their mean absolute error. This prediction will aid farmers in determining the optimal crop to grow by taking into account factors such as temperature, rainfall, area, and more.[5]

Doshi, Z., et al. (2018, August 1) present an intelligent system, named AgroConsultant, to aid Indian farmers in arriving at an informed decision concerning the choice of crops to cultivate. This decision is based on several factors such as the sowing season, the geographic location of the farm, soil properties, and environmental factors, including temperature and rainfall.[6]

K., & Mathur, P.. (2020, October 30) have presented a paper that aims to forecast crop yield through the utilization of factors such as area, yield, production, and area under irrigation. To achieve this objective, four distinct machine learning techniques, namely Decision Tree, Logistic Regression, Lasso Regression, and Ridge Regression, were employed to estimate crop yield. Cross-validation techniques were applied to validate the results, utilizing metrics such as mean absolute error, mean squared error, and root mean squared error. The Decision Tree algorithm was found to outperform the other machine learning techniques.[7]

Nishant, P. S., et al. (2020, June 5) present a paper that proposes a predictive model for estimating crop yield across various crop types in India. The novelty of this research lies in the utilization of simple parameters such as State, district, season, and area to enable the user to predict crop yield in any given year. Kernel Ridge, Lasso, and ENet algorithms are used in the paper to predict crop yield in order to achieve this. Additionally, the paper uses the concept of Stacking Regression to enhance the accuracy of these algorithms and provide a more precise prediction.[8]

Jadhav, A., et al. (2022, April) discusses the use of machine learning and deep learning technologies to provide crop recommendations to farmers in India. The aim is to overcome the challenges faced by farmers and provide effective solutions to avoid undesirable results. The paper reports the results of implementing four machine learning algorithms, namely Decision Tree, Random Forest, Logistic Regression, and XGBoost, on the dataset. The features selected for the dataset include N (Nitrogen), P (Phosphorous), K (Potassium) values of the soil, temperature, humidity, rainfall, and pH value. The Random Forest algorithm provided the highest accuracy of 98.9%, followed by XGBoost with 98.2% accuracy. Logistic Regression gave 95.6%, while Decision Tree gave the least accuracy at 95.3%. [9]

III. PROPOSED METHODOLOGY

The various steps involved in this research work are data collection, pre-processing, building recommender model, training and testing the recommender model. Construction of the recommender is the specific task and required the domain knowledge with some degree of expertise in agri-informatics, and other commercial multivariate parameters. In order to meet the specified task, the proposed recommender comprises of two approaches which deal with different kinds of issues.

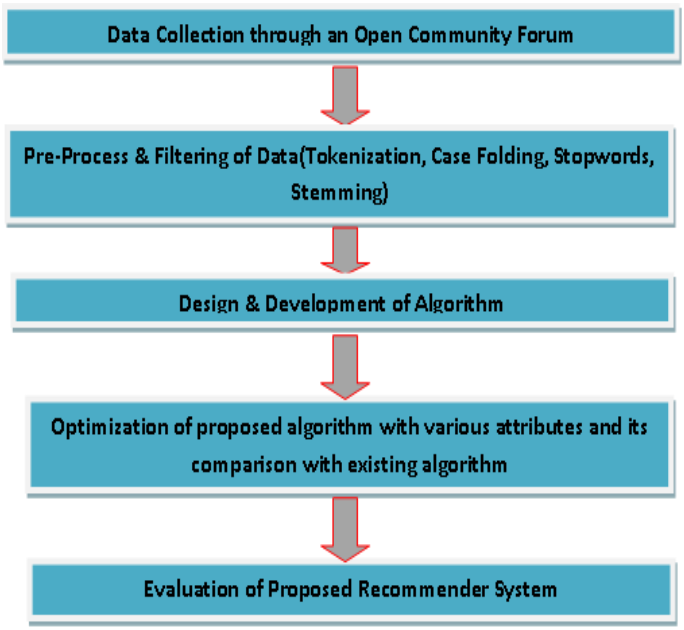

Fig 1 Proposed Methodology Flowchart

A. Data Collection

Data is very important in machine learning systems. As the location and climate in the agriculture field changes, the crop production also changes accordingly. Therefore, to implement this recommendation system, we have selected Madhya Pradesh state which is located in India. Since the geographical location of different districts is different in Madhya Pradesh also, we have collected district wise data.

ISSN No:-2456-2165

This data is gathered from various government websites. The data about the crops of each district of Madhya Pradesh of last fifteen year was gathered from www.dac.gov.in. This dataset has Name of District, Crop Name, Crop Year, Season, Area (in Hectare), Production (in Tonnes) and Yield information from 2004 to 2019. Because type of soil is also a very important parameter hence data of soil type was gathered from dolr.gov.in.

B. Data Preprocessing

Data preprocessing is the first and most important step to converting the raw data into a clean data set for a machine learning model. Usually, real-world data is not complete, consistent, accurate and often lacks specific attribute values/trends. While doing any analysis with data, it is essential to clean it and put it in a formatted way, in that situation, data preprocessing helps to clean, format, and organize the raw data, into clean data for Machine Learning models.

Here we have used different machine learning algorithms like gradient boosting classification, KNN algorithm, logistic regression, random forest classification and decision tree classification, they require data scaling to give good results and we have used it to implement the recommendation system.

- *Data Integration*

Because the data has been collected from multiple sources. To build the recommender model, a dataset of various features from multiple sources is collected and combined.

- *Data Scaling*

Data Scaling is a data preprocessing step for numerical features The dataset used has different scales, so here we use MinMaxScaler class to re-scale our data. MinMaxScaler also known as min-max scaling or min-max normalization, it is the simplest method and consists of rescaling the range of features to scale the range in [0, 1] i.e. the minimum of feature is made equal to zero and the maximum of feature equal to one The original distribution's shape is preserved while the values are scaled to a specific value range.

- *Identifying and handling missing values*

In data preprocessing, it is important to correctly identify and handle missing values, failing to do so is very likely to lead to incorrect and erroneous conclusions from the data. This will hinder our ML project. There are two ways to handle missing data: (a) Deleting a particular row (b) Calculating the mean.

Here we have used calculate mean method to handle missing values, as this method is useful for features with numerical data. Here, we can calculate the mean, median or mode of a particular feature or column or row that has missing values and replace the result with the missing values. This method can add variation to the dataset, and it can efficiently negate any loss of data.

- *Data Partitioning*

For training and testing purpose we split the entire dataset into two parts: A total of 80% of the dataset was used to train the model, while 20% of the data were saved for model testing.

- *Machine Learning Algorithms used to recommend crop*

We have used supervised machine learning algorithms to get the desired results. Supervised machine learning is an algorithm that learns from labeled training data to help predict outcomes for unforeseen data. The system may provide targets for any new input once it has received The system may provide targets for any new input once it has received sufficient training. In this research study, we focused on various machine learning classification techniques, such as Gradient Boosting, KNN Algorithm, Logistic Regression, Random Forest, and Decision Tree, to find the best recommendation system that provides accurate results.

IV. RESULTS

After performing the data cleaning and visualizations, we implemented our machine learning algorithms on the features of the dataset. The five algorithms we have used are Gradient Boosting Classifier, KNN Classifier, Logistic Regression, Random Forest Classifier and Decision Tree Classifier. The features selected of the dataset are District, Crop, Year, Season, Soil Type, Area and Production.

After implementing the five algorithms on our dataset, we can see that Random Forest Classifier gives us the highest accuracy out of all the five algorithms (78%), followed by KNN (73%), Decision Tree Classifier (68%), Gradient Boosting Classifier gave an accuracy of (64%) while Logistic Regression gave the least accuracy at (50%).

Table 1 Machine Classification Learning Algorithms

	Algorithms	**Accuracy**
0	Random Forest	0.776352
1	KNN	0.729760
2	Decision Tree	0.678156
3	Gradient Boosting	0.644088
4	Logistic Regression	0.497971

ISSN No:-2456-2165

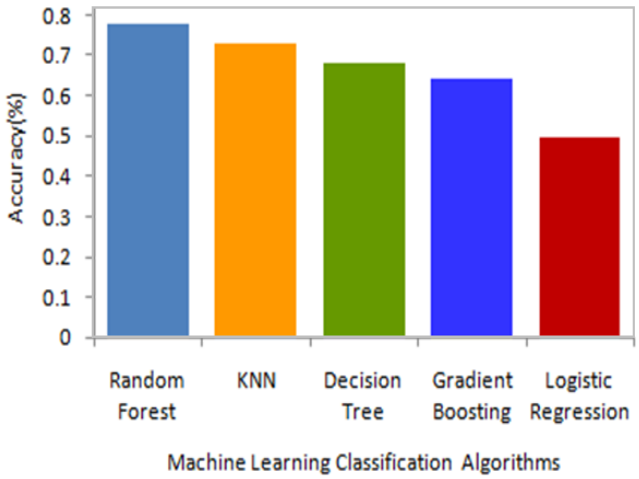

Fig 2 Machine Classification Learning Algorithms

V. CONCLUSION AND FUTURE SCOPE

In this paper, various machine learning algorithms are recommended to predict crop name based on soil type, rainfall, season and area. Experiments have been carried out on a dataset from the Government of India and it has been established that the Random Forest Classifier gives the highest accuracy. Accurate forecast of different specified crops in different districts will help the farmers of Madhya Pradesh. This will not only help the farmers to choose the right crop to grow in the next season but will also promote the use of technology in the agriculture sector. Further development is to integrate the crop recommendation system with another subsystem, yield predictor that would also provide the farmer an estimate of production if he plants the recommended crop.

REFERENCES

[1]. Bondre, D. A., & Mahagaonkar, S.. (2019, October 1). Prediction of crop yield and fertilizer recommendation using machine learning algorithms.
[2]. Reddy, D. A., Dadore, B., & Watekar, A.. (2019, February 15). Crop Recommendation System to Maximize Crop Yield in Ramtek region using Machine Learning. International Journal of Scientific Research in Science and Technology.
[3]. Kumar, R., Singh, M. P., Kumar, P., & Singh, J. P.. (2015, May 6). Crop Selection Method to maximize crop yield rate using machine learning technique
[4]. Palanivel, K., & Surianarayanan, C.. (2019, January 1). An Approach for Prediction of Crop Yield Using Machine Learning and Big Data Techniques. Social Science Research Network.
[5]. Nigam, A., Garg, S., Agrawal, A., & Agrawal, P.. (2019, November 1). Crop Yield Prediction Using Machine Learning Algorithms. Institutions.
[6]. Doshi, Z., Nadkarni, S., Agrawal, R., & Shah, N.. (2018, August 1). AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms.
[7]. K., & Mathur, P.. (2020, October 30). Crop Yield Estimation in India Using Machine Learning.
[8]. Nishant, P. S., Venkat, P. S., Avinash, B. L., & Jabber, B.. (2020, June 5). Crop Yield Prediction based on Indian Agriculture using Machine Learning.
[9]. Jadhav, A., Riswadkar N., Jadhav P., Gogawale Y.., (2022, April). Crop Recommendation System Using Machine Learning Algorithms. International Research Journal of Engineering and Technology (IRJET)