International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 08 | Aug 2020 www.irjet.net p-ISSN: 2395-0072

CRAPCSS-Crop Recommendation and Pest Control Suggestion System

Annapoorna.S[1], Apoorva Herle[2], C M Sushma[3], Neha.Y.Jain[4], Mr.Vijay Kumar.S[5]

[1]*Annapoorna S, Student*
[2]*Apoorva Herle, Student*
[3]*C M Sushma, Student*
[4]*Neha Y Jain, Student*
[5]*Mr.Vijay Kumar S, Asst. Professor, Dept. of ISE, BNMIT, Bangalore, India*

Abstract - In the field of agriculture, huge proportions of crops are lost due to the imprecise selection of the crop to be cultivated in a particular portion of land. The farmers are in most cases, not appraised of the requirements and necessities of the crops i.e. the minerals, soil moisture, gritty texture, ph levels and other soil prerequisites. This can cause anguish and affliction to farmers, both mentally and financially. Another problem that a farmer, in most occasions stumbles upon is the pest and diseases that can affect the crops that are being grown, which they are generally blinded to at a very early stage of the process. This obstacle of farmers is attempted to be addressed in our paper and we have taken up this task to try and help solve this problem with the help of a Recommendation and Prediction System. With our model acting as a catalyst to this hindrance, we attempt to identify and predict the pre-eminent and much suited crop to the farmers and diagnose the pest that may be the cause of the affect as well as advocate the most appropriate and effective pest control techniques. In this paper, we have applied SVM algorithm, Logistic Regression algorithm and we have found that SVM classifier model gives the better accuracy as compared to other algorithms.

Key Words: Crop Identification, Pest Control, Recommendation System, SVM Classification Model, Prediction Model

1. INTRODUCTION

Agriculture is an integral sector of Indian Economy. The contribution of Indian agriculture sector to India's Gross Domestic Product (GDP) is about 18% and which in average provides employment to 50% of the country's workforce, which clearly shows that agriculture plays a vital role in the gross economy. Agriculture is a major contributor to the Indian economy.

The main stream Indian population depends either explicitly or implicitly on agriculture for their livelihood. It is very important to monitor the soil nutrients for a good yield. Agriculture sector faces many problems such as irregular rainfall, floods, draught, climate change etc. To overcome these problems technological solution is needed which can help the farmers. The productivity of farming is not only depending on natural resources but it also depends on input provided to the system.

2. LITERATURE SURVEY

The literature survey on the recommendation and prediction of crops using normalized dataset employs a few common methodologies and techniques, but this paper focuses on a varied approach to classifying and predicting using the SVM algorithm.

The following papers mentioned subsequently highlights upon the methodologies unique to them.

Recommendation System for Crop Identification and Pest Control Technique in Agriculture aims to resolve the problems of farmers with the help of recommendation system. By the help of this model, best suitable crop to the farmer is predicted and the pest that may affect is detected. In this paper, SVM classification algorithm, SVM and Logistic Regression algorithm is applied and it is found that SVM classification model gives the better accuracy as compared to other algorithms. The crop prediction using classification is accomplished using SVM classification method [1].

Ensemble is a data mining model also known as the Committee Methods or Model Combiners that combine the power of multiple models to acquire greater prediction, efficiency than any of its models could achieve alone. In this system, one of the most familiar ensemble technique called Majority Voting technique is used. In the voting technique any number of base learners can be used. There has to be at least two base learners. The learners are chosen in a way that they are competent to each other yet being complimentary also. Higher the competition higher is the chance of better prediction. Drawbacks include only fewer crops are used and thus cannot be considered universally, it is very site specific and not wide range, use of CHAID tree makes it unrealistically short and uninteresting due to this the multiple splits are hard to relate to real business conditions [2].

Technique used is sliding window nonlinear regression to predict based on different factors affecting agriculture productions such as rainfall, temperature and market price. The price of the crops predicted is not stable and this system overcomes this problem by using demand grade for each crop. The process of predicting the crop yield uses the data mining approach through which the analyzed soil datasets are predicted. The crop yield estimation using classification technique estimates the crop yield and selects the most

suitable crop for cultivation, thereby improving the value and gain of farming area. Drawbacks include that the system does not consider the demand existing in the agricultural economy and uneducated farmers cannot come to know about the system, even when farmers know about it but they do not know how to use it. And the methods to meet the demand and conveying the suggestions are not feasible [3].

This paper presents a technique named CSM to select sequence of crops to be planted over season. CSM method may improve net yield rate of crops to be planted over season. The proposed method resolves selection of crop (s) based on prediction yield rate influenced by parameters (e.g. weather, soil type, water density, crop type). CSM algorithm works on prediction of crop yield rate based on favorable condition in advance and gives a sequence of crops with highest net yield rate. It takes crop, their sowing time, plantation days and predicted yield rate for the season as input and finds a sequence of crops whose production per day are maximum over season. CSM method retrieves all possible crops that are to be sown at a given time stamp. Yield rate of these crops are evaluated, if yield rate per day of these crops are fair (within tolerance) then those crops are selected for crop sequences. Drawbacks include performance and accuracy of CSM method depends on predicted value of influenced parameters, if the price factor is not included in the crop selection method, then system may lead to select a wrong crop to grow and with crop rotation, some years you just have to plant it in another crop instead of the high yielding one. This may lead to lower financial returns at times [4].

This works on previously cultivated crop yields and depends on season, location, temperature and rainfall primarily. Rainfall predictor has a major role to play which is a separate data model. Neural networks are the algorithm used for the predictions for easier development of recommendations system. This assists the farmers in making informed decisions. So a random vector is produced which is autonomous of the past arbitrary vector with same discrimination and a tree is created by utilizing the set. Drawbacks is that it does multi-dimensional prediction that is the crop yield in a particular defined area is dependent on the rainfall predictor, and in case of faulty rainfall predictor dataset using it crop yield predictions may result in faulty predictions and crop demand and supply as well as other economic aspects like farming harvest prices and retail prices are not considered [5].

Two important and well known classification algorithms K-Nearest Neighbor (KNN) and Naive Bayes (NB) are applied to the soil dataset which is taken from the soil testing laboratory Jabalpur, M.P. Their accuracy is obtained by evaluating the datasets. Each algorithm has been run over the training dataset and their performance in terms of accuracy is evaluated along with the prediction done in the testing dataset. The prediction is done on the category attribute for each instance of soil. Low (L), medium (M) and high (H) can be defined as the categories in which the soils are classified. Confidence values for low (L), medium (M) and high (H) categories has been computed, Drawbacks are that KNN predictions are based on intuitive assumptions that objects close in distance are potentially similar which can be a good sense of discrimination between the KNN when making predictions and the complexities may occur due to the usage of fewer datasets and other efficient models such as Decision tree or principal component analysis can be used for a much better accuracy [6].

3. METHODOLOGY

In the proposed method, identification and classification of crops suitable for a particular land holdings using SVM method is implemented. Figure 1 represents the architectural design of the proposed system. The system has 3 primary modules namely preprocessing, feature extraction and classification.

With advance in technology, many updated technology has been applied in agriculture sector to improve the health of crops named Precision agriculture. Indian farmers tend to choose unsuitable crop for their soil and this problem can be solved by precision agriculture where the soil characteristics like soil type, texture, pH value etc. are used for detecting that which crop is suitable for cultivation in that soil. This minimizes the risk of cultivating inexact crop which collectively results in better crop yield from a particular land holding. Once the seeds are sown it's important for farmers to prevent it from insects and weed. With our research work we can help farmers to be more technically sound and they can spread awareness among their fellow farmers about the correct crop to choose and also by preventing crop loss every year, they can be financially stable. Data collection for crop prediction with attributes such as pH, rainfall, temperature and other soil parameters.

On applying a non-linear classification technique called the kernel-trick in SVM helps in indirectly mapping the input into high-dimensional feature spaces. The dataset is trained using classification models like linear SVM algorithm and decision tree to predict the crop. The model also makes the user aware about the pests that generally affect the crop predicted and its removal measurement.

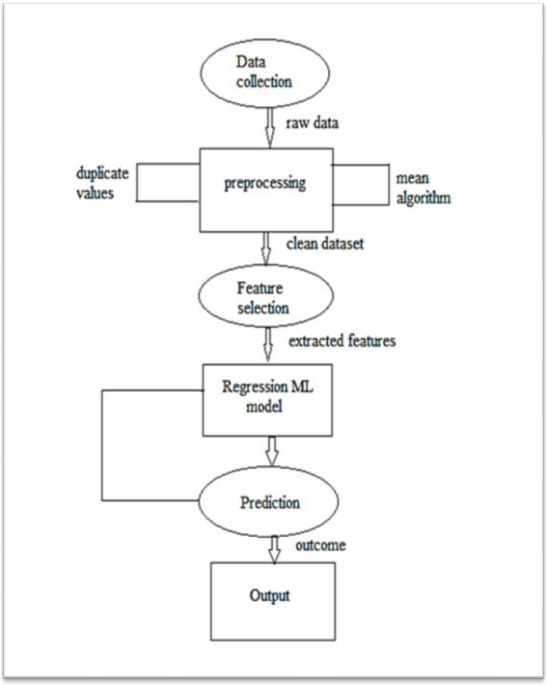

Fig 3.1: Architectural design.

3.1 Pre-processing

The function of preprocessing module is to upgrade and enhance the data present in the CSV file and discarding unwanted disturbances and outliers. Figure 3.1.1 represents the initial data flow levels of the raw collected information. Figure 3.1.2 is the follow up data flow diagram that defines the clear approach of structuring and normalizing of raw dataset into usable clean dataset. Mean imputation: The data we get is rarely homogenous.

Sometimes data can be missing and it needs to be handled so that it does not reduce the performance of our machine learning model. In our project, we replace the missing data by the mean or median of the entire column. For this, we used the sklearn. preprocessing library which contains a class called imputer which helped us in taking care of our missing data. Our object name is imputer and the imputer class can take parameters like missing values, which is the placeholder for the missing values, strategy which is the imputation strategy wherein you replace missing values using the mean along the axis and axis which can be assigned 0 or 1; 0 to impute along columns and 1 to impute along rows.

Fig 3.1.1: Dataflow diagram level 0.

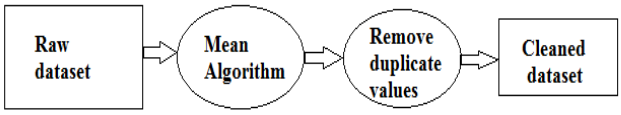

Fig 3.1.2: Dataflow diagram level 1.

3.2 Feature selection

Feature Selection is the process where you automatically or manually select those features which contribute most to your prediction variable or output in which you are interested in. It Reduces Overfitting: Less redundant data means less opportunity to make decisions based on noise. It Improves Accuracy: Less misleading data means modeling accuracy improves. It Reduces Training Time: fewer data points reduce algorithm complexity and algorithms train faster.

You can get the feature importance of each feature of the dataset by using the feature importance property of the model. Feature importance gives a score for each feature of your data, the higher the score more important or relevant is the feature towards the output variable. Feature importance is an inbuilt class that comes with Tree Based Classifiers, we will be using Extra Tree Classifier for extracting the features for the dataset in the project. Figure 3.2.1 notates each of the main processes that together form the complete system. The cleaned dataset obtained as the output in level 1 is considered the input in level 2 which is subjected to the feature selection algorithm to obtain the featured data as an output.

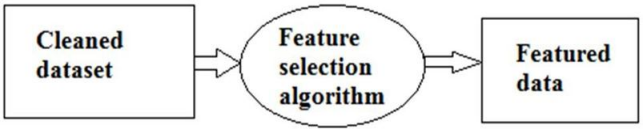

Fig 3.2.1: Dataflow diagram level 2.

Figure 3.2.2 on performing feature selection algorithm on the cleaned data, certain attributes showcase more priority and thus gives the appropriate sense of selection for the model.

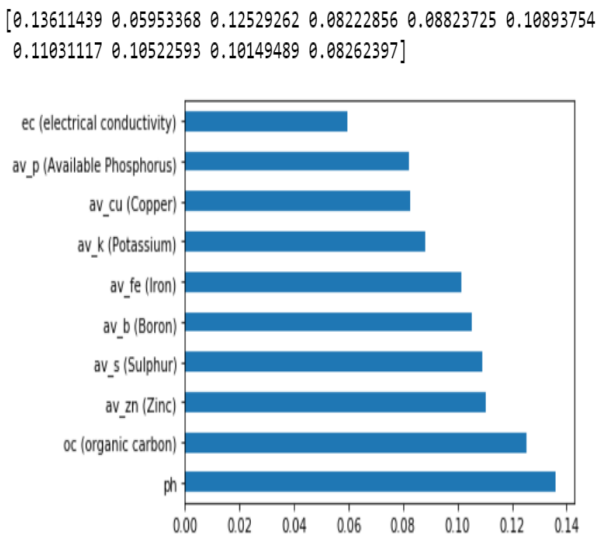

Fig 3.2.2: Feature selection based on certain attributes.

3.3 Classification

The support vector machine (SVM) is a predictive analysis data-classification algorithm that assigns new data elements to one of labeled categories. SVM is, in most cases, a binary classifier; it assumes that the data in question contains two possible target values.

In our project, we have used the sklearn.svm module within which we have used the submodule called svc within the SVM module for classification purpose with the parameters such as kernel (kernel = 'sigmoid'), gamma (gamma = 'auto'), C which behaves as a regularization parameter in SVM (C= 5.2).

Figure 3.3.1 depicts that the data that is integrated is subjected for the selection of the attribute from the previous dataflow diagram. The raw data is taken as the input for processing. In this project, the classification is distributed among train set and test set wherein the test size is 20% and train size is 80% of the total structured dataset.

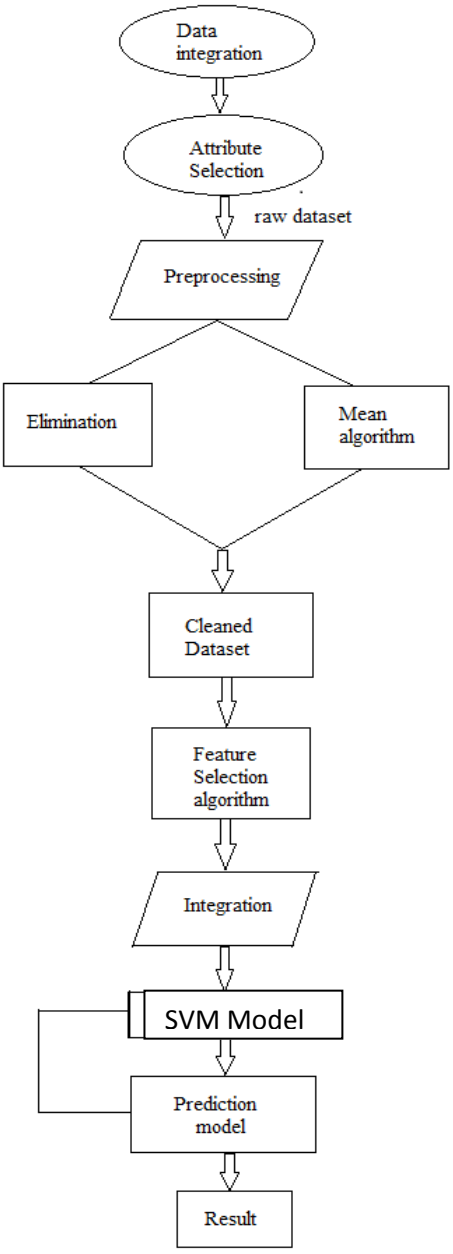

Fig 3.3.1: Dataflow diagram level 3.

Figure 3.3.2 illustrates the sequence of actions that the recommendation system undergoes. The user acts as a catalyst for the feature selector to collect the features and transfer the collected data to the data management module which integrates the data and transfers he same to the Support Vector Machine algorithm where the training and analyzing of data takes place and finally the resulted data is transferred to the database as an output.

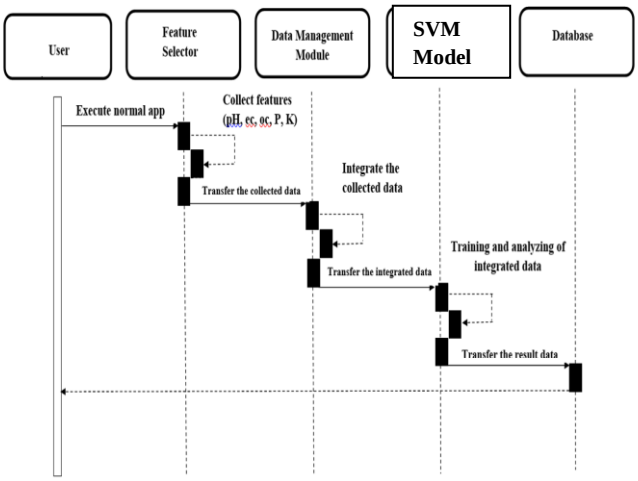

Fig 3.3.2: Sequence diagram for recommendation system.

4. RESULTS AND DISCUSSION

This chapter discusses the results obtained by implementing different modules of the proposed system. Our model is designed by using Spyder IDE (ANACONDA 3) and the accuracy of our project accounts to 97%.

```
[528]  M↓
    from sklearn.metrics import classification_report
    print(classification_report(y_test, y_predict))
```

	precision	recall	f1-score	support
arecanut	1.00	0.69	0.82	26
banana	0.95	1.00	0.97	18
cashewnut	1.00	1.00	1.00	19
coconut	1.00	1.00	1.00	20
groundnut	1.00	1.00	1.00	16
jowar	1.00	1.00	1.00	21
maize	1.00	1.00	1.00	22
mango	1.00	1.00	1.00	17
millets	1.00	1.00	1.00	21
nutmeg	1.00	0.94	0.97	16
orange	1.00	1.00	1.00	19
paddy	1.00	0.95	0.98	22
pepper	0.71	1.00	0.83	22
potato	1.00	1.00	1.00	23
red chilli	1.00	1.00	1.00	27
rice	1.00	1.00	1.00	18
rubber	1.00	1.00	1.00	22
wheat	1.00	1.00	1.00	15
accuracy			0.97	364
macro avg	0.98	0.98	0.98	364
weighted avg	0.98	0.97	0.97	364

Fig 4.1: Accuracy snapshot of the model from project.

After training our model with ample amount of data values we have successfully predicted the crop and the pest control techniques to quiet a satisfactory level and the rules are induced from algorithms like SVM classifier. The rules induced from these models helps in building RECOMMENDATION SYSTEM. The input is taken in as a form format and the training set formed is precisely classified. This model can predict for any situation and for any crop and if it fails to predict, it will display a message saying the same.

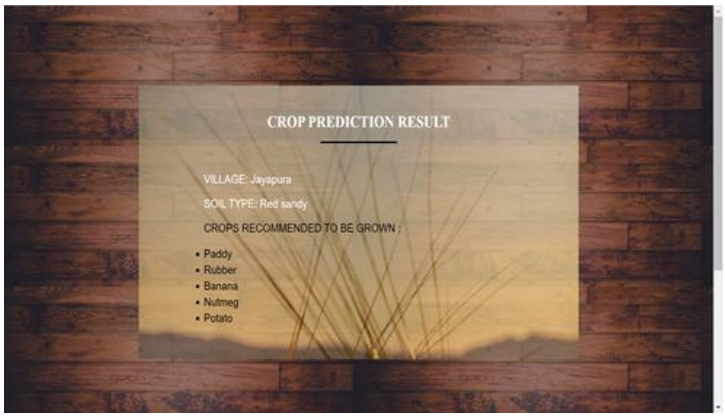

Fig 4.2: Snapshot of prediction of suitable crop.

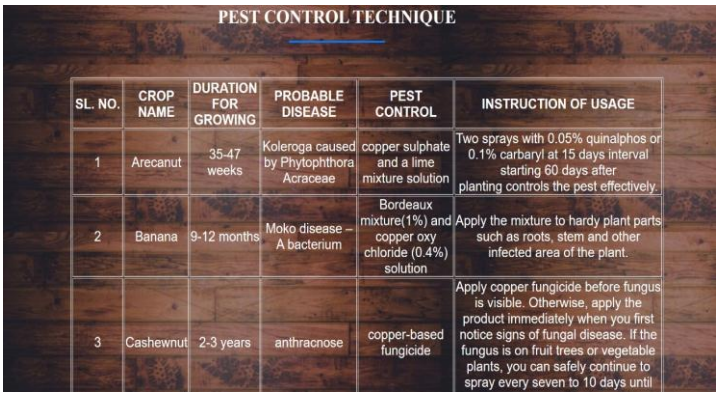

Fig 4.3: Snapshot of apt pest control technique for suitable crop predicted.

On prediction and recommendation of the appropriate crop, the suitable pest control technique is suggested in order to reduce wastage and financial losses.

5. CONCLUSION AND FUTURE WORK

Agriculture which is an important part of our economy, is essential to ensure that even the smallest investment which is done in the agriculture sector should be taken care of. So it is essential to check if the right crop has been chosen for a land holding which matches its requirements to benefit the nation and farmer in particular. After training our model with an ample amount of data values we have successfully predicted the crop to quiet a satisfactory level. This model can predict for any given situation and for any crop, and if it fails to predict it, it will show a message stating the same.

The future enhancements for this project are that we can develop a recommendation system for all the states in India and also devise an optimum methodology to implement this recommendation system globally. This project can also use live data using the IOT components. Another enhancement that can be made is that it can be deployed on a platform where it is accessible to everyone in the form of an application which will be user friendly. This will make our

project accessible to local audience throughout and our goal to reach millions will be achieved.

ACKNOWLEDGEMENT

We consider it a privilege to express through the pages of this paper, a few words of gratitude to all those distinguished personalities who guided and inspired me in the completion of the project. The authors wish to thank Prof. Vijay Kumar S, Prof. Manjunath G S and Dr. Shashikala from the department of Information Science and Engineering, BNMIT, Bengaluru for their relentless support, guidance and encouragement. We would also like to thank the other faculty members at BNMIT for their continuous support towards the successful completion of our project.

REFERENCES

[1] Kumar, A., Sarkar, S., & Pradhan, C. "Recommendation System for Crop Identification and Pest Control Technique in Agriculture.", 2019 International Conference on Communication and Signal Processing (ICCSP). doi:10.1109/iccsp.2019.8698099

[2] Paul, M., Vishwakarma, S. K., & Verma, A. . "Analysis of Soil Behaviour and Prediction of Crop Yield Using Data Mining Approach.", 2015 International Conference on Computational Intelligence and Communication Networks (CICN). doi:10.1109/cicn.2015.156K. Elissa, "An Overview of Decision Theory," unpublished. (Unplublished manuscript)

[3] Pudumalar, S., Ramanujam, E., Rajashree, R. H., Kavya, C., Kiruthika, T., & Nisha, J. (2017). "Crop recommendation system for precision agriculture.", Eighth International Conference on Advanced Computing (ICoAC). doi:10.1109/icoac.2017.7951740

[4] Doshi, Z., Nadkarni, S., Agrawal, R., & Shah, N. . "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms.", 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). doi:10.1109/iccubea.2018.8697349

[5] Raja, S. K. S., Rishi, R., Sundaresan, E., & Srijit, V "Demand based crop recommender system for farmers.", 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR). doi:10.1109/tiar.2017.8273714

[6] Kumar, R., Singh, M. P., Kumar, P., & Singh, J. P.. "Crop Selection Method to maximize crop yield rate using machine learning technique.", 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM).