Crop Yield Prediction using KNN Model

Dr. V. Latha Jothi [(1)], Neelambigai A [(2)], Nithish sabari S [(3)], Santhosh K [(4)]
(1) Professor, [(2)(3)(4)] Final Year Students
Department of CSE
Velalar College of Engineering and Technology
Thindal, Erode

***Abstract*:-Agriculture plays a dominant role in the growth of the country's economy. Climate and different environmental modifications have become a major threat in the agriculture field. This makes the problem of predicting the yielding of crops an exciting challenge. Data Mining techniques are the better selections for this purpose. Different Data Mining techniques are used in agriculture for estimating the upcoming year's crop production. Crop Yield Prediction includes predicting yield of the crop from previous historical data like rainfall, temperature and groundwater level. KNN model is using to classifies the groundwater level dataset to predict the future test data record dataset. It could be useful in analysing the ground water levels in the past and which predict the future levels.**

Keywords: Crop yield; Data mining; Rainfall prediction; Temperature prediction; Groundwater level prediction; KNN model.

I. INTRODUCTION

Agriculture is the most important supply of Indian Economy. For the better crop yield, the farmers always require a timely guidance to predict the future of crop productivity and also an analysis is to be made which will help the farmers to utilize full capacity in the crop production for their crops. Yield prediction is a vital agricultural problem. The volume of data is vast in Indian agriculture. For agriculture problems data mining is applied widely. As every farmer is interested by knowing that how much yield is expected to make. In the past, with the farmer's previous experience for a particular crop, one can make the predictions for crop.

Data Mining is the method of extraction, transforming, loading and predicting the meaningful facts from big information to extract some patterns and also transform it into understandable structure for further use. In this paper the main goal is to create a user-friendly interface for farmers, which gives the analysis of crop yield prediction which is based on available datasets. To maximize and predict the crop yield productivity, one can make use of different data mining techniques. Applying the data mining techniques on historical climate and crop production data several predictions can be made on the basis of knowledge gathered which in turn can help in increasing crop productivity.

II.BACKGROUND

Data mining is a field of the intersection of computer science and statistics used to discover patterns in the information bank. The main goal of the data mining process is to extract useful information from the dossier of data and mild it into an understandable structure for future use. There are specific processes and techniques used to hold out data mining successfully.

This section discusses about various related works already done in data mining techniques using previous years dataset. Most of the researchers focused on the problem for yield prediction.

A Modified linear regression method can be used to predict rainfall using average temperature and cloud cover [3] in various districts in southern states of India. The Linear Regression method is modified in order to obtain the most optimum error percentage by iterating and adding some percentage of error to the input values. This method provides an estimate of rainfall using different atmospheric parameters like average temperature and cloud cover to predict the rainfall.

A Seasonal Auto Regressive Integrated Moving Average (SARIMA) model which includes iterative estimation, analysis and forecasting levels predicts the monthly rainfall [5]. Mean Absolute Percentage Error (MAPE) is used to calculate the accuracy.

Data mining techniques can be used to predict crop yield, where the information gained for each attribute is calculated to acquire [6] a ranking of the attributes such as rainfall, potential evapotranspiration, maximum and minimum temperature, cloud cover and wet day frequency to select the attributes.

Utilizing abundant surface groundwater available at the end of the wet season while benefiting from timely access to shallow groundwater [2] from the process of capillary rises so that the farmers can have a better crop yield with or even without the expensive irrigations.

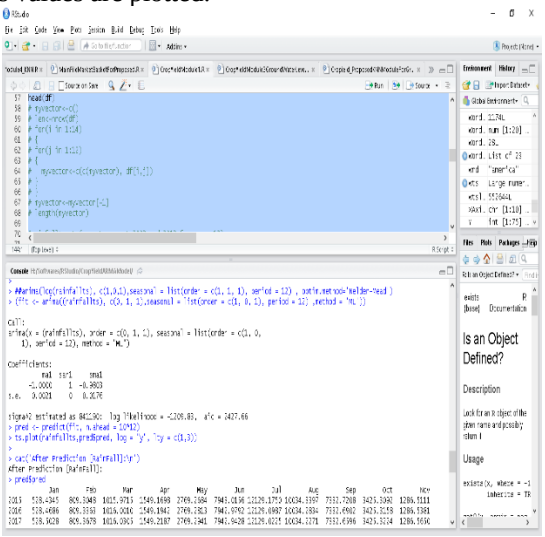
ARIMA and Multiple Linear Regression can be used to predict rainfall for all states of India. In MLR equation, parameters are taken from the dataset and variables are extracted from the dataset by means of correlation. ARIMA is used for modelling time series and rainfall prediction [4].

III.EXPERIMENTAL STUDY

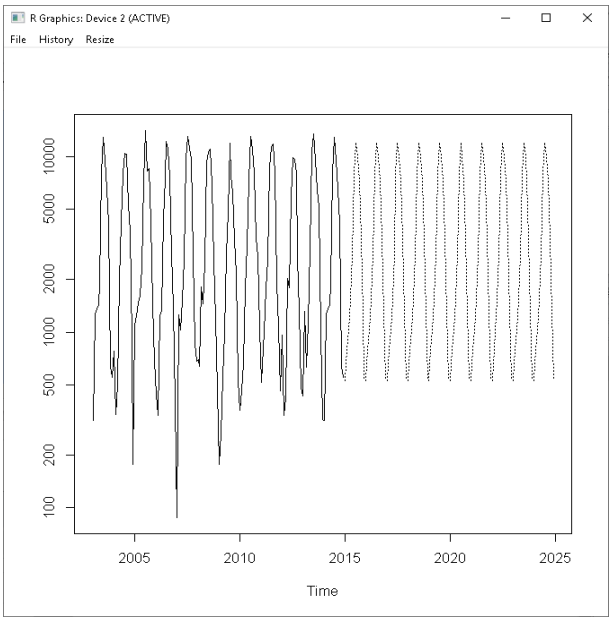
The Auto Regression Moving Average (ARMA) and K Nearest Neighbors (KNN) models are applied to the predict crop yield in upcoming years. Rainfall, Temperature and groundwater level dataset are taken from Indian metrological department. It contains record amount of the rainfall, temperature and groundwater level in previous years. Each row is represented a year the average rainfall, temperature of every month distributed in 12 columns. By using KNN model the dataset is divided into 75% of training data and 25% of testing data to predict the crop yield production.

3.1 ARMA MODEL BASED PREDIC/TION FOR RAINFALL

In this module, rainfall water data set is taken for Indian data for past ten years. The data is converted into data frame and pre-processed such that zero values in all column's records are eliminated. The data is converted into time series format such that twelve records (for each month) for all years present in the data set. Then using 'arima' function, the model is prepared for the given data set and predicted for upcoming years. Using ts. plot () the upcoming years values are plotted.



AFTER PREDICTION: RAINFALL



GRAPH FOR RAINFALL PREDICTION

3.2 ARMA MODEL BASED PREDICTION FOR TEMPERATURE

In this module, temperature data set is taken for Indian data for past ten years. The data is converted into data frame and pre-processed such that zero values in all column's records are eliminated. The data is converted into time series format such that twelve records (for each month) for all years present in the data set. Then using 'arima' function, the model is prepared for the given data set and predicted for upcoming years. Using ts.plot() the upcoming years values are plotted. Using the previous rainfall and temperature outcomes, fuzzy logic-based crop

yield prediction is carried out the following algorithm.

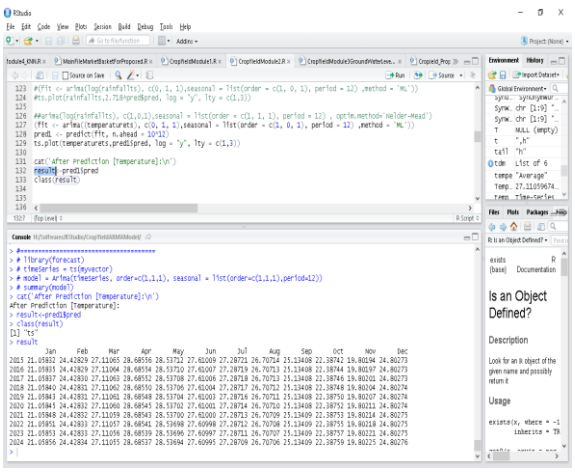
IF (rain=='very good' and temp=='very good') or (rain=='very good' and temp=='good'): THEN yield = 'very good'

IF (rain=='very good' and temp=='average') or (rain=='good' and temp=='very good') or (rain=='good' and temp=='good') or (rain=='good' and temp=='average'): THEN yield = 'good'

IF (rain=='very good' and temp=='bad') or (rain=='very good' and temp=='very bad') or (rain=='good' and temp=='bad') or (rain=='good' and temp=='very bad') or (rain=='average' and temp=='very good') or (rain=='average' and temp=='good') or (rain=='average' and temp=='average'): THEN yield = 'average'

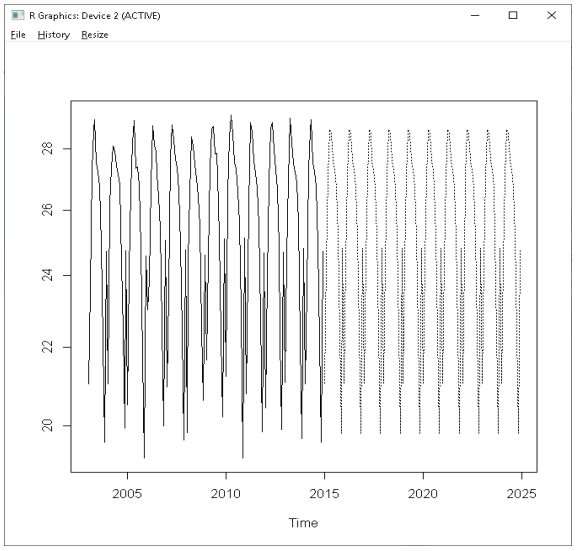
IF (rain=='average' and temp=='bad') or (rain=='average' and temp=='very bad') or (rain=='bad' and temp=='very good') or (rain=='bad' and temp=='good') or (rain=='bad' and temp=='average') or (rain=='bad' and temp=='bad'): THEN yield = 'bad'

IF (rain=='bad' and temp=='very bad') or (rain=='very bad' and temp=='very good') or (rain=='very bad' and temp=='good') or (rain=='very bad' and temp=='bad') or (rain=='very bad' and temp=='average') or (rain=='very bad' and temp=='bad') or (rain=='very bad' and temp=='very bad'): THEN yield = 'very bad'



AFTER PREDICTION: TEMPERATURE

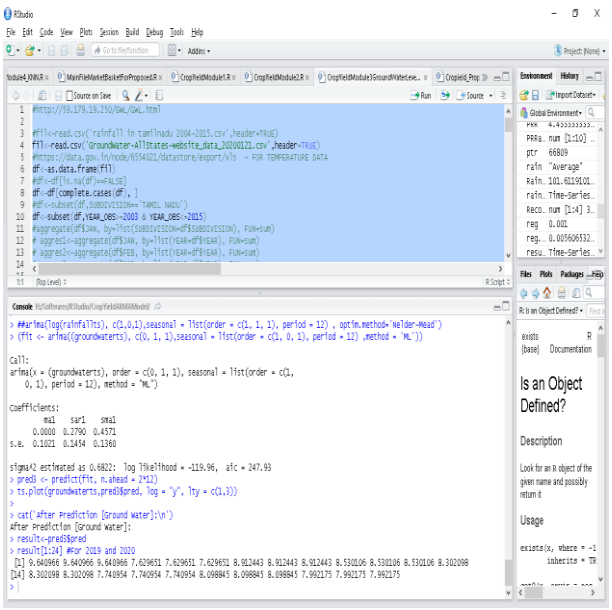
CLASSIFICATION	AVERAGE YEARLY RAIN	AVERAGE SEASON TEMPERATURE
Very Good	62-87 cm	20-25 C
Good	87-100 cm	15-20 C
Average	100-150 cm	12-15 C or 25-33 C
Bad	150-200cm	8 -12 C or 33-40 C
Very Bad	All other values	All other values



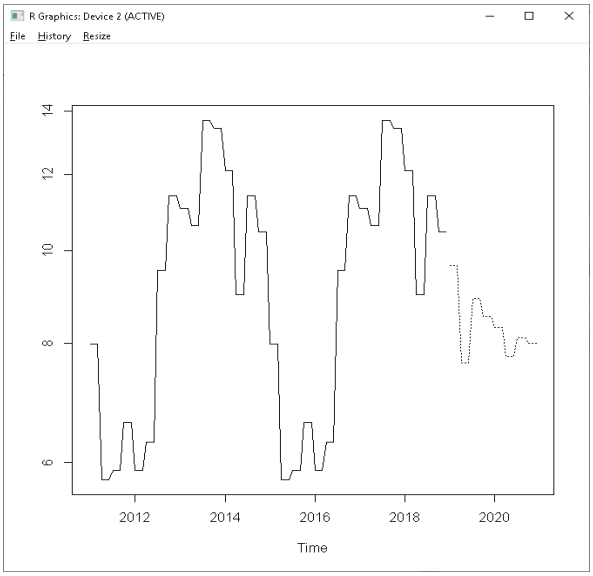
GRAPH FOR TEMPERATURE PREDICTION

3.3 ARMA MODEL BASED PREDICTION FOR GROUND WATER

In this module, groundwater level data set is taken for Indian data for past ten years. The data is converted into data frame and pre-processed such that zero values in all column's records are eliminated. The data is converted into time series format such that twelve records (for each month) for all years present in the data set. Then using 'arima' function, the model is prepared for the given data set and predicted for upcoming years. Using . plot() the upcoming years values are plotted.



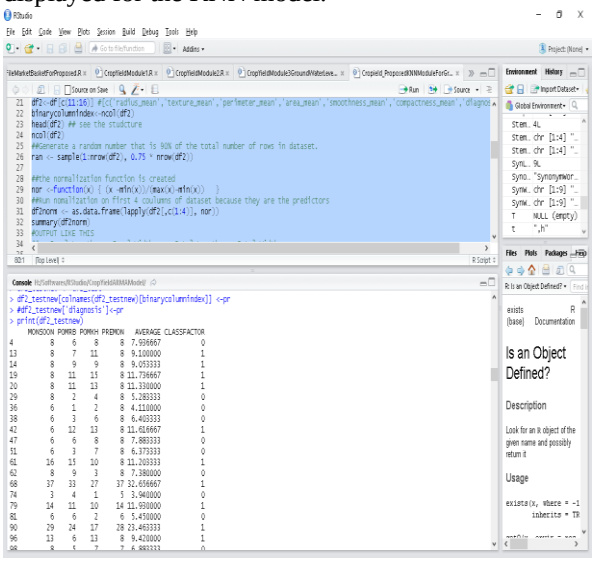
AFTER PREDICTION: GROUND WATER LEVEL



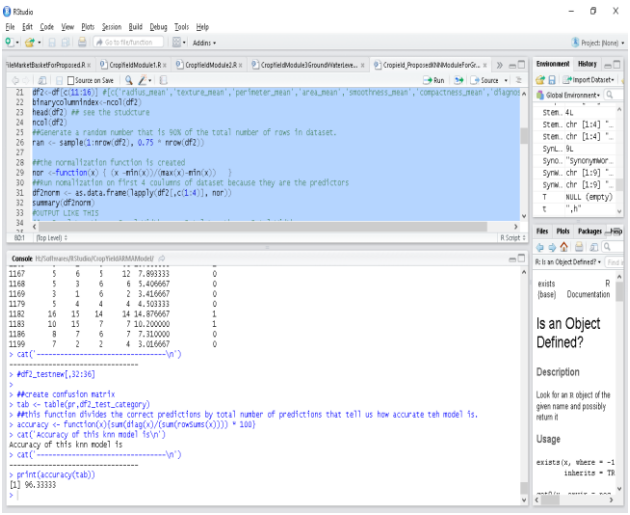
GRAPH FOR GROUNDWATER LEVEL DATA

3.4 GROUND WATER LEVEL CLASSIFICATION BASED ON KNN MODEL

In this module, groundwater level data set is taken for Indian data for past ten years. The data is converted into data frame and pre-processed such that zero values in all column's records are eliminated. The data from 'MONSOON', 'POMKH', 'POMRB' and 'PREMON' column are taken, average value calculated, middle value is found out and values below middle value are considered as zero and above as one. These two values are applied in class factor column. Then KNN classification is predicted using the class factor column. For test run, K value is given as 6 and the model is predicted. The accuracy is calculated and displayed for the KNN model.

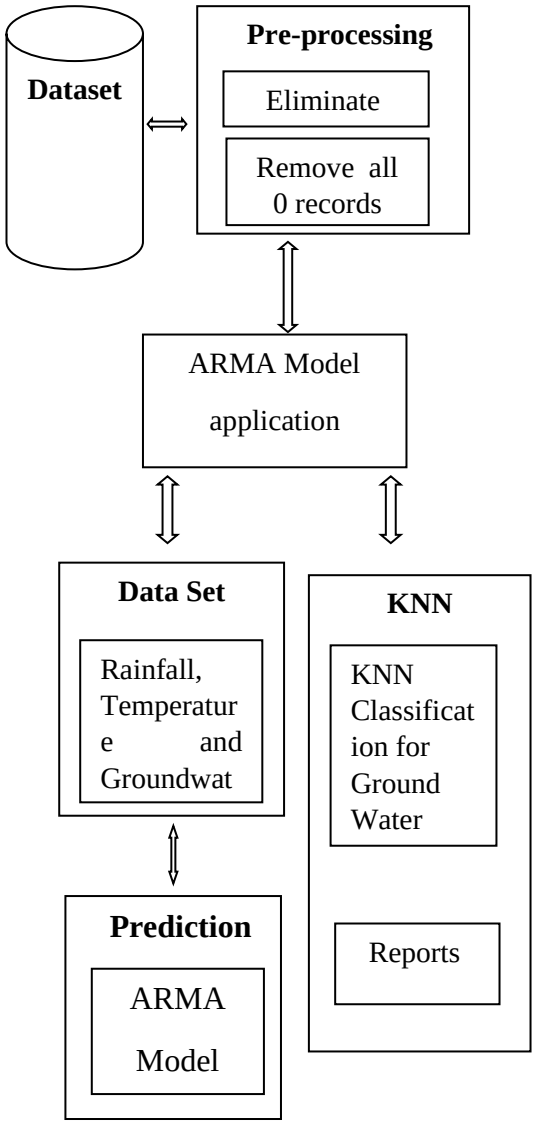


KNN CLASSIFICATION FOR GROUND WATER LEVEL DATA



ACCURACY OF THE KNN MODELI

IV.WORK FLOW



V.CONCLUSION AND FUTURE ENCHANCEMENTS

According to the results, temperature is best predicted by the ARIMA model and the accuracy of predictions made for rainfall by ARMA model is also good. Rainfall, which is an important factor for the prediction of crop yield is difficult to estimate precisely. Climate factors may change due to other remaining variables which may influence the prediction of rainfall.

Also, the proposed work makes use of fuzzy logic to estimate crop yield which works on a set range rather than discrete values, therefore, the error in predicted rainfall data does not cause problems as long as the difference between actual and estimated values is not drastic. The model can successfully predict crop yield for a given year when the rainfall and temperature values for the previous years is known. The model can successfully predict ground water level for a given year when the previous years' value is known. In addition, this project classifies the ground water level data set records using KNN to predict the model for future test record data sets. It will be helpful in analysing the ground water levels in the past and so as to predict the future levels. In future, logistic regression can be applied to further classify the data.

VI.REFERENCE

[1] Balachandran.S, Ashokan R. and Sridharan S (2006). "Global surface temperature in relation to northeast monsoon rainfall over Tramlined", Journal of Earth System Science, vol. 115, no.3, pp. 349-362.

[2] Mohammad Rafiuzzaman, Ibrahim Cil." A fuzzy logic based agricultural decision support system for assessment of crop yield potential using shallow ground water table" International Journal of Computer Applications Volume 149 – No.9, 2016.

[3] Mohapatra S.K, Upadhyay.A, and Gola C. "Rainfall prediction based on 100 years of meteorological data." 2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN), pp. 162-166. IEEE, 2017.

[4] Renuka, Sujata Terdal." Evaluation of Machine Learning Algorithms for Crop Yield Prediction". International Journal of Engineering and Advanced Technology (IJEAT) Volume-8 Issue-6, August 2019

[5] Shivam Bang, Rajat Bishoni, Ankit singh Chauhan, Akshay kumar, Indu chawal." Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models" 2019 Twelfth International Conference on Contemporary Computing (IC3).IEEE, 2019.

[6] Sundaram, Meenakshi S, and Lakshmi M. "Rainfall prediction using seasonal auto regressive integrated moving average model." Computer science 3, no. 4 (2014), pp. 58-60, 2014.

[7] Veenadhari S, Misra B, and Singh C.D. "Machine learning approach for forecasting crop yield based on climatic parameters." 2014 International Conference on Computer Communication and Informatics, pp. 1-5. IEEE, 2014.