How to Cite:
Poornima, R., Sathya, G., & Shalini, B. (2022). Crop yield prediction in agriculture based on climatic conditions and ground water level prediction for southern regions. *International Journal of Health Sciences*, *6*(S8), 478–487.
https://doi.org/10.53730/ijhs.v6nS8.9725

Crop yield prediction in agriculture based on climatic conditions and ground water level prediction for southern regions

Mrs. R. Poornima
Assistant.Professor of Computer Science and Engineering, K.S.Rangasamy College of Technology, Tiruchengode – 637 215, Namakkal District, TamilNadu, India

G. Sathya
Students of Computer Science and Engineering, K.S.Rangasamy College of Technology, Tiruchengode – 637 215, Namakkal District, TamilNadu, India

B. Shalini
Students of Computer Science and Engineering, K.S.Rangasamy College of Technology, Tiruchengode – 637 215, Namakkal District, TamilNadu, India

***Abstract*---**The main aim of this project is to predict the rainfall ,ground water level and temperature for next ten years to predict the crop yield. The prediction of crop yield based on previous years downfall, temperature and ground water level results in taking necessary steps to improve crop yield in future years.Understanding crop yield can help ensure food security and reduce impacts of climate change. Crops are sensitive to various weather phenomena such as temperature, weather and rainfall. Therefore, it becomes crucial to incorporate these features when predicting the yield of a crop. Forecasting is a complicated process. In this work, ARMA (Auto Regressive Moving Average) method is employed to forecast crop yield. Past ten years of knowledge set is taken for temperature, rainfall and ground water level for our country. Moreover, classification of downfall, temperature and ground water level data set records is done using ANN approach to predict the model for future test record data sets. It will be helpful in analyzing crop yield in the past and so as to predict the future levels.

***Keywords*---**crop yield prediction, agriculture, climatic conditions, ground water level prediction.

Introduction

Agriculture is one of the most important profitable sectors in Southern regions. It plays an important part in pastoral development and sustainability. The position of farming may drop factors like unpredicted fall, global climate change, use of inordinate fungicides etc. The main idea of this study is to supply a strategy for crop yield production supported the historical climatic and production data. Crop yield forecasting supported the previous years of temperature and downfall can help growers take necessary ways to enhance crop yield within the coming season. Understanding crop yield can help ensure food security and reduce impacts of global climate change.

The project tried to develop a method such that the crop yields can be predicted beforehand using only temperature and rainfall of previous years. Accurate rainfall prediction is a difficult task because rainfall depends on various features such as cloud cover,and many other climatic factors but we wanted to provide useful information using only two features about crop yield i.e. temperature and downfall. In existing system, the data set is taken from Indian meteorological sites saved in excel files as 'csv' files. It contains temperature, rainfall and spring water level data. These data are taken for text pre-processing first and then converted into input data format. Then ARMA model is used to forecast the future years' levels and matplot library is used for showing the forecast graph.

In proposed system, all the prevailing methodology is administered. Like existing system, the data set is taken from Indian meteorological sites saved in excel files as 'csv' files. It contains temperature, rainfall and ground water level data. These data are taken for text pre-processing first and then converted into neurons for input. Then ANN model is used to predict the future years' levels and then is applied to classify the crop yield in future years.

Related Works

S.NO	SUMMARY OF THE PAPER	TECHNIQUE USED	METHODS/ ALGORITHM
[1]	In any of Data Mining actions the training data is to be collected from past data and the gathered data is used in terms of training which has to be exploited to study how to categorize future yield predictions.	Data Mining	K Nearest Neighbour algorithm

[2]	In agriculture, several data mining strategies are used and analysed to predict crop cultivation in the future.This research has focused on comparing the feature selection methods of different prediction models, and suggested the best method to forecast crop cultivation in the future	Boruta, SFFS, and RFE	Support vector machine,Radial Basis function
[3]	A deep learning model is used that takes advantage of state-of-the-art modeling and solution techniques to predict crop yield based on environmental data and management practices..	Batch Normalization,Dropout, Stochastic Gradient Descent	Convolution Neural Network,Reccurent Neural Network
[4]	It focuses on implementing crop yield prediction system by using Data Mining techniques by doing analysis on agriculture dataset. The result is based on comparison among the classifiers.	J48, LWL, LAD Tree and IBK classifiers	KNN, Bayesian Network, and Decision Tree.
[5]	The present study focuses on the effect of various parameters like rainfall, mean temperature, area under irrigation on production of major crops. It also aims to predict crops production by applying various regression algorithms	Regression Technique	Multiple Linear Regression, Random Forest Regression and Multivariate Adaptive Regression Splines (Earth). Apriori Algorithm

Proposed Methodology

This study includes the prevailing method of crop yield prediction and suggests new innovative method. In existing study, the info set is taken from Indian meteorological sites saved in excel files as 'csv' files. It contains rainfall data. These rainfall data are taken for pre-processing of text first and converted into statistic data set format. Then artificial neural network is employed to classify the new test records with the given about data set. It'll be helpful to predict the future years' levels. Drawbacks are:

- Only classification is formed among past records Future rainfall level is not predicted yet.
- Aims in classification only and helps to assess future crop yield production.
- It couldn't be preferred when the outlier data is more.

In proposed system, the prevailing methodology is also carried out. Like existing system, here also, Indian meteorological sites dataset is taken and saved in excel files as 'csv' files which contain rainfall / spring water level data. These data are then converted into statistic data set format and ARMA model is used for predicting the future years' levels and then a fuzzy logic is applied for classifying crop yield in future years. This study is administered by spliting work into the following modules.

ANN based classification of rainfall data

Artificial Neural Network: It includes single input layer, 1 hidden layer (can be made to 2 if required) and 1 output layer, whose input is instance's feature vector and output is that the instances' category. It also includes one propagation process, then forward propagation then back propagation. The method of propagation is in Algorithm 1.

Algorithm 1

INPUT: DESCRIPTION FOR VECTOR ENCODED (Name of State, Year and values of Annual rainfall are encoded to binary strings. deem example, if four states are taken then first Delaware value is 0001, second state value is 0010 then on. Likewise if 15 years are taken, then first year value is 000000000000001, second year value is 000000000000002 then on continuing. Likewise, the whole columns are taken for input layer neurons are concatenated and then bit strings are prepared for each neuron.

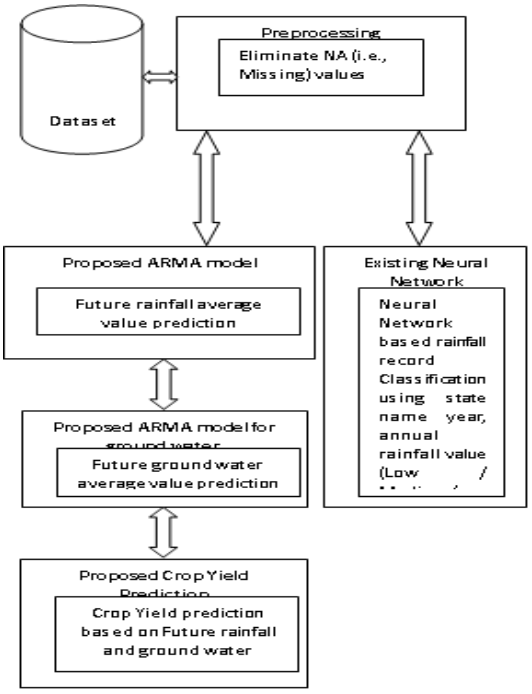

Fig 3.1 Architecture digram

OUTPUT: 1 Output neuron with value zero to atleast one. The category count is about to three so for any given input neurons with thousand neuron values, if value between zero and 0.333 is ready as the output in the output neuron, then category belongs to the primary (out of thousand categories), 0.334 to 0.666 belongs to next category then on. Fetch all records, convert into bit strings, 32 hidden layer, 3 output layer neurons are set and network is formed to run. The output is generated and displayed.For all records, above process is formed and output is displayed

Arma model based prediction for rainfall

Here, rainfall water data set are taken for Indian rainfall data for past years. Then the info is converted into data frames and preprocessed so that zero values in the entire columns records are eliminated. The info is then converted into time-series format such that 12 records (for each month) for all years present in data set. Then using "arima' function in R, given data set is employed to predict the model and predicted after for upcoming years. Using ts.plot() upcoming years values also are plotted.

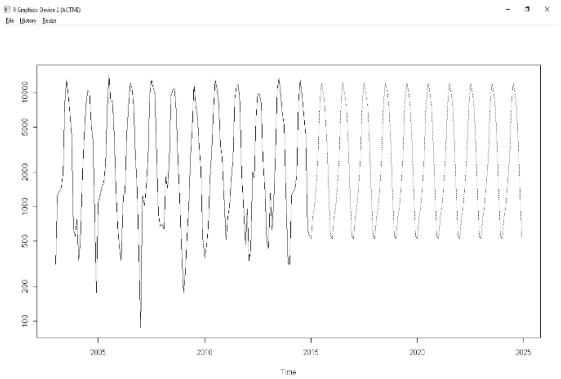

FIG 3.2 Rainfall predicted data from 2016 to 2025

It is found that no major decrease in rainfall for future 10 years (2016 to 2025).

Arma model based prediction for ground water

In this third module, water table data set is read out for Indian rainfall data for past ten years. The info is converted to data frame and zero values in all columns records are removed/eliminated. The data is converted to time-series format such that 12 records (monthwise) for all years present in data set. Using 'arima' function, given data set is employed to predict the model and then prediction is carried out for next upcoming years. Using ts.plot() next upcoming years values are plotted.

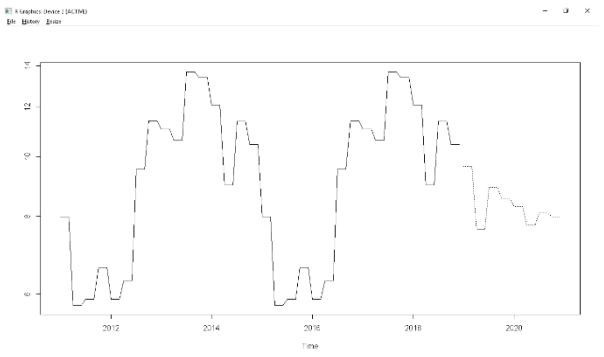

Fig 3.3 Ground water level predicted data for 2021 and 2022

It is found that light decrease in 2021 and 2022 when comparing with 2020

Crop yield prediction based on rainfall and temperature

In this fourth module, using previous module rainfall / temperature outcomes, crop yield prediction process supported fuzzy logic is carried out. The classification table for crop yield prediction based on downfall and Temparature is as follows

CLASSIFIC -ATION	AVERAGE YEARLY RAIN	AVERAGE YEARLY TEMPARATURE
Very Good	62-87 cm	20-25 C
Good	50-62 cm or 87-100 cm	15-20 C
Average	25-50 cm or 100-150 cm	12-15 C or 25-33 C
Bad	10-25 cm or 150-200 cm	8-12 C or 33-40 C
Very Bad	All other Values	All other Values

Table for crop yield prediction based on Rainfall and Temperature

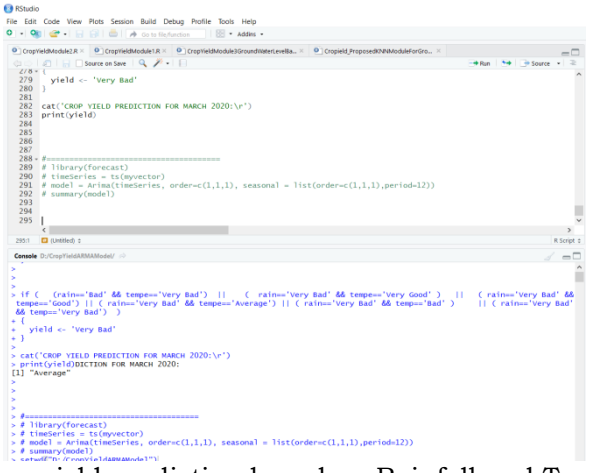

Fig 3.5 crop yield prediction based on Rainfall and Temparature

Ground water level classification based on KNN model

In this fifth module, India ground water-level data set is taken for ten years old data. Then it is converted into R data frames and zero values in all records are removed. The info from i) 'MONSOON', ii) 'POMKH', iii) 'POMRB' and iv) 'PREMON' column are taken, then average value calculated, and middle value is prepaped and values lesser than middle value are updated as zero and greater than together. These 2 values are applied for classfactor column. KNN classification is allied than for prediction using class factor column. For test runs, K value is given as 5 then 6 and model is predicted. Then accuracy value is calculated and considered for the KNN model.

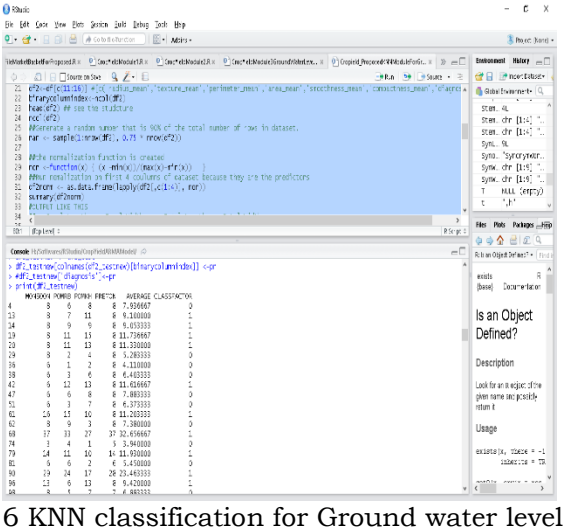

Fig 3.6 KNN classification for Ground water level data

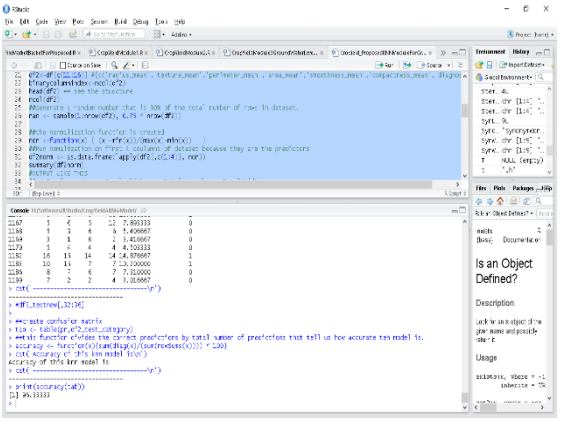

Fig 3.7 Accuracy of KNN prediction model

Implementation

Input Design

Input design is that the process of converting user-originated inputs to a computer understandable format. Input design is one among the most expensive phases of the operation of computerized system and is often the major problem of a system. An outsized number of problems with a system can usually be tracked backs to fault input design and method. Every moment of input design should be analyzed and designed with utmost care.

The system takes input from the users, processes it and produces an output. Input design is link that ties the knowledge system into the world of its users. The system should be user-friendly to realize appropriate information to the user. The choices made during the input design are

- To provide cost effective method of input.
- To achieve the very best possible level of accuracy.
- To make sure that the input is understand by the user.

System analysis decide the subsequent input design details like, what data to input, what medium to use, how the info should be arranged or coded, data items and transactions needing validations to detect errors and eventually the dialogue to guide user in providing input. Input data of a system may not be necessarily is raw data captured in the system from scratch. These also can be the output of another system or subsystem.

The design of input covers all the phases of input from the creation of initial data to actual entering of the data to the system for processing. The planning of inputs involves identifying the data needed, specifying the characteristics of every data item, capturing and preparing data from computer processing and ensuring correctness of knowledge. Ambiguity in input results in a total fault in output. The goal of designing the input file is to make data entry as easy and error free as possible.

Output Design

Output design generally refers to the results and knowledge that are generated by the system for many end-users.output is the main reason for developing the system and the basis on which they evaluate the usefulness of the application. The output is meant in such a way that it is attractive, convenient and informative.Because the outputs are the most important sources of information to the users, better design should improve the system's relationships with user and also will help in decision-making. Form design elaborates the way output is presented and therefore the layout available for capturing information.

Results and Disussion

The project tried to develop a way such that the crop yield can be predicted beforehand using only temperature and rainfall of previous years. Accurate rainfall prediction may be a difficult task because rainfall depends on various features such as cloud cover, evapotranspiration, and lots of other climatic factors but we wanted to extract useful information about crop yield using only two features i.e. temperature and rainfall. The proposed method uses Auto Regressive Moving Average and Seasonal ARIMA models to predict temperature. Since we would like to predict temperature from past values, these models suit our needs. A time-series created from the dataset is fed into the model to predict temperature. Similarly, ARMA and ARMA with exogenous variables (ARMAX) models are used to predict rainfall. The ARMAX model is employed in case of rainfall so that other factors such as cloud cover, temperature and evapotranspiration also can be taken into account. We used a symbolic logic system to predict yield. The fuzzy model takes within the predicted values from the model with least errors and gives the yield for that season.

Conclusion

According to the results, temperature is foremost forecasted by the ARIMA model and thus the a fineness of prediction made for rainfall by ARMA model is additionally good. Rainfall, which is a pivotal factor for forecasting of crop yield is delicate to estimate precisely. Climate factors may change other remaining variables which can impact the forecasting of downfall.

Scope for future development

The proposed work makes use of symbolic sense to estimate crop yield which works on a group range rather of separate values, thus, the error in forecasted rainfall data doesn‘t beget problems as long because the difference between factual and estimated values isn't drastic. The model can successfully predict crop yield for a given time when the downfall an temperature values for the former time value is understood . The model can successfully prognasticate ground water position for a given time when the former time value is understood . In addition, this design classifies the lowermost water position data set records using KNN to predict the model for awaited test record data sets. It will be helpful in inspect the lowermost water situations within the history also on predict the longer term

situations. In future, logistic retrogression are frequently applied to further classify the data.

References

1. Fathima, Sowmya K, Sunita Barker, Dr. Sanjeev Kulkarni " Analysis of crop yield prediction using data mining technique" International Research Journal of Engineering and Technology (IRJET) Volume: 07 Issue: 05 | May 2020
2. G.Mariammal, S.Raja &, A. Suruliandi " Crop prediction based on soil and environmental characteristics using feature selection techniques" International Conference for Emerging Technology (INCET) | 2021
3. Saeed khakhi, V.Archontoulis , Lili wang, Sotirios" A CNN-RNN Framework for Crop Yield Prediction" International Research Journal of Engineering and Technology (IRJET) Volume: 06 Issue: 11 | Dec 2018.
4. Shruti Mishra Priyanka Paygude Snehal Chaudhary Sonali Idate "Use of Data Mining in Crop Yield Prediction" International Conference on Inventive Systems and Control |2018
5. Suvidha Jambekar & Shikha Nema "Prediction of Crop Production in India Using Data Mining Techniques" International Conference for Emerging Technology (INCET) 2019
6. Rinartha, K., & Suryasa, W. (2017). Comparative study for better result on query suggestion of article searching with MySQL pattern matching and Jaccard similarity. In *2017 5th International Conference on Cyber and IT Service Management (CITSM)* (pp. 1-4). IEEE.
7. Rinartha, K., Suryasa, W., & Kartika, L. G. S. (2018). Comparative Analysis of String Similarity on Dynamic Query Suggestions. In *2018 Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS)* (pp. 399-404). IEEE.
8. Novoa, R. B. (2021). State of the art and future applications of digital health in Chile. *International Journal of Health & Medical Sciences*, *4*(3), 355-361. https://doi.org/10.31295/ijhms.v4n3.1772