Second International Conference on Nexgen Technologies
Sengunthar Engineering College, Tiruchengode, Namakkal Dist. Tamilnadu (India)
8th - 9th March 2019 www.conferenceworld.in ISBN : 978-93-87793-75-0

CROP PREDICTION USING MACHINE LEARINING ALGORITHM

Mr.K.Bharanidharan[1], Mr.V.Kannan [2], Mr.V. Karthick[3], Dr.B. Mahalakshmi[4]

Final year student of Computer Science and Engineering, K.S.Rangasamy College of Technology, Tiruchengode, (India)[1,2,3]

Associate professor of Computer Science and EngineerinG, , K.S.Rangasamy College of Technology, Tiruchengode, (India)[4]

ABSTRACT

Prediction of crop yield helps in managing storage of crop as well as it directs the transportation decision, and risk management issue related to crop. The Proposed method helps us to predict the crop yield and suggest the best crop in different type of condition based on different parameter related to climates. Machine learning techniques can be used to improve prediction of crop yield under different climatic scenarios, parameters such as area, production rate, mean temperature, mean rainfall details. Hence, multiple parameters are considered. Support Vector Machine and K-Nearest Neighbor are used in analysis of agricultural dataset and results are compared to predict the accuracy.

I INTRODUCTION

In the world of increasing population and climatic change, so the necessity food resources are more and more. Farmers are facing a difficult situation to cultivate a crop which gives more production which sustainable with changing climates and market economic pressure. The agricultural yield is primarily depends on weather conditions. Yield prediction is an important agricultural problem. Every farmer is interested in knowing, how much yield he is about expect. In the past, yield prediction was performed by considering farmer's previous experience on a particular crop. Data mining is the computer-assisted process of digging through and analyzing enormous sets of data and then extracting the data. Data mining tools predict behaviors and future trends, allowing businesses to make proactive, knowledge-driven decisions. Understanding the relative Importance of these Climate factors to crop yield variation could provide valuable information about crop planting and management under climate change condition for policymakers and farmers.

The dataset is collected from publicly available Indian Government records for the crop yield prediction. The parameter includes in the dataset are precipitation, minimum temperature, average temperature, maximum temperature and rainfall, area, production and yield for the season from January to December for the years 2000 to 2015. The Support Vector Machine and K-Nearest Neighbors classifier using R tool has been applied on the current dataset.

This paper examines comparative analysis of supervised machine learning techniques to predict the crop in different weather conditions. This prediction will help the farmer to choose whether the particular crop is suitable for that area and in the particular season. This prediction is carried out by the Support Vector Machine and K-Nearest Neighbors algorithms where high accuracy can be achieved. The analysis of results improves crop yield prediction.

II. RELATED WORK

Support Vector Machines (SVMs) is one of the supervised machine learning techniques. There are a number of examples of where it has been used in the agricultural domain. Yunous vagh et al.,(2012) reported as on Crop production has been linked to a number of factors such as seasonal temperature, temperature variations, radiation, evaporation, soil moisture and crop management practices, this work investigated the effect of temperature variation on crop production [2]. Spatial scales are important in that the scale related results are specific to the associated group or agency. For example, the national scale may be used by governments to determine their economic strategy from food reserves, while results from smaller scale relationships, being used to detect food shortages and associated mitigation possibilities, and for seasonal forecasting by farmers at the farm level Root Mean Square Errors (RMSEs) were established as initial criteria for selection of the final algorithm upon which the test set would run. The third criterion was the RMSE of the cross-validation, the algorithm with the best performance in these three criteria turned out to be GP with a correlation. The results showed a high positive correlation between stochastic monthly temperature and the wheat yield, but the situation was rather complex.

Tripathi et al., (2006) reported on how Support vector machine was applied for reduction of Precipitation for climate change scenarios To reduce the generalization error bound and to achieve generalized performance, Support vector machine was used to forecast the demand and supply of pulp wood [3]. Support vector machine was also applied to provide insights into crop response patterns related to climate conditions by providing the features contribution analysis for agricultural yield prediction.

II PROPOSED SYSTEM

The proposed system implemented with various parametric factors like temperature, rainfall, production, soil type and based on land type. Support vector machine and k-Nearest Neighbors are used to analyze the datasets and provide the classifications. Barota package is applied on the preprocessed dataset. This package gives confirmed attributes, tentative attributes and removed attributes. In this system confirmed attributes and tentative attributes are used and predict more accuracy. Those attributes are combined into the separate dataset for the further implementation of algorithm. Supervised machine learning algorithms are applied on the predicted attribute dataset.

IV. IMPLEMENTATION

The implementation process includes dataset collection, data preprocessing, feature selection, proposed methodology, getting results.

A. Dataset Collection

Now a day, India becomes digitalized. All agriculture datasets are available in governmental web sites (https://data.gov.in). Whereas in Tamil Nadu government had provided an individual data sets in their web sites. Individual data sets like temperature, rainfall, production, soil type and based on land type are available individually in their web sites. In order to preprocess the data, the individual data set are combined as single huge data set with common attributes.

B. Data Preprocessing

Data preprocessing is the major process of prediction of crop yield. Preprocessing is the process of extracting knowledge form large amount of data. Combining these datasets on the basis of common attributes crop name, district, season and year.

C. Feature Selection

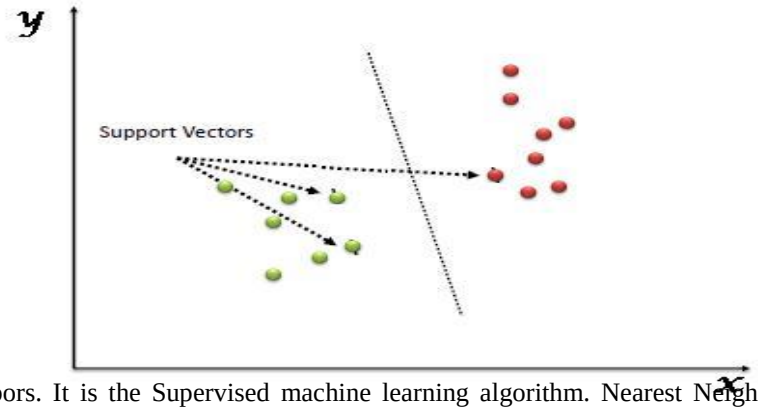
Attribute selection is the major part to get accurate results. In preprocessed data set which contains many attributes. To find the attributes which are more accuracy for the algorithm Baruta package is used.

D. Baruta Package

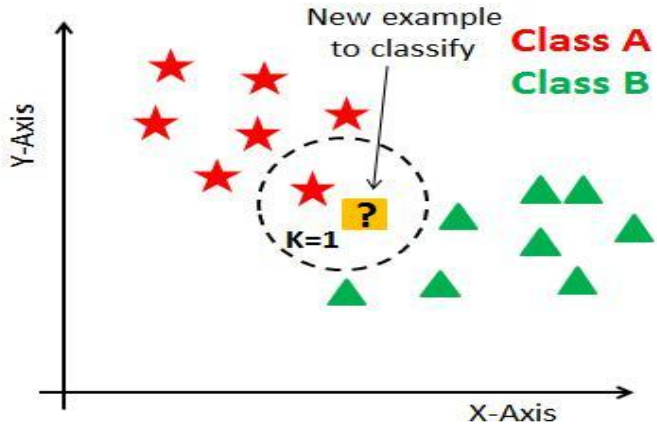
It will create the duplicate dataset, and shuffles the values, which created dataset is shadow of original dataset. This package works on the basis of Random Forest classifier. This package will provides attributes and to create a dataset for methodology implementation. The new data set will be created from the result of Baruta package. The predicted data set will be the final data set used for the algorithm implementation.

E. Methodology

a) Support Vector Machine: Support Vector Machines (SVM) are supervised learning models for classification and regression problems. They can solve linear and non-linear problems and work well for many practical problems. Kernel function (nonlinear) to map the data into a different space where a **hyperplane** (linear) cannot be used to do the separation. Hyperplane can classifics the nearest data points.



b)K-Nearest Neighbors. It is the Supervised machine learning algorithm. Nearest Neighbors classification is used to find the nearest data points in the datasets.K is the integer value that must be greater then 0. Algorithm is applied on the trained datasets.Using metrics package to get accuracy of results.



V.RESULTS AND DISCUSSION

This section discusses the results obtained after applying the Supervised Machine learning techniques on crop yield dataset of Tamilnadu state, India. R tool was used to construct the algorithm. Random classifier takes 100 decisions for feature selection. From the Comparison of algorithms results, the prediction of accuracy is found. In more or less both algorithms gives certain range of accuracy.

Comparison of Support Vector Machine and K-Nearest Neighbors, K-NN gives more accuracy then SVM. The SVM algorithm achieved the accuracy of 87.36% and K-NN achieved 87.68% of accuracy.

VI. CONCLUSION

The crop data set is analyzed and the optimal parameters for the crop production are determined. Hence the proposed system is considered with the soil and other factors for the crop and to increase the crop production under the different climatic conditions.

The proposed method will be faster and more efficient than manual prediction especially over large datasets.

The Performance evaluation of the proposed method is measured in teams of accuracy.

ACKNOWLEDGEMENTS

We Acknowledge DST-File No.368. DST – FIST (SR/FIST/College-235/2014 dated 21-11-2014) for financial support and DBT–STAR-College-Scheme-ref.no:BT/HRD/11/09/2018 for providing infrastructure support.

REFERENCES

[1] Ami Mistry, Vinita Shah, "Brief survey of data mining techniques applied to applications of agriculture ," International Journal of Advance Research in computer and communication engineering, vol. 5, pp. 301–304, Feb 2016.

[2] Yunous Vagh, Jitian Xiao, "Mining Temperature Profile data for Shire-level crop yield prediction", IEEE, pp.77-83, Jul 2012.

[3] S. Tripathi, V.V. Srinivas and R.S. Nanjundiah, "Downscaling of precipitation for climate change scenarios: a support vector machine approach", Journal of Hydrology, vol. 330, no. 3, pp.621-640, 2006.

[4] Niketa Gandhi and J. Armstrong, "A review of the applications of data mining techniques for decision making in agriulture," IEEE, 2016.

[5] Yogesh Gandge and Sandhya,"A study on various data mining techniques for crop yield prediction " , IEEE,pp. 220-223,2016.

[6] Ramesh, Medar Gandge and Sandhya,"A survey on various data mining techniques for crop yield prediction " , International journal ofAdvance Research in Computer Science and Management,vol.2,2016.

[7] Jain A., Murty M. N., Flynn P.J. (2013), "Data clustering: a review".ACM ComputSurv, doi:10.1.1.18.2720, Vol.31, No.3, pp. 264–323.

[8] MotiurRahman M, Haq N, Rahman RM. Application of data mining tools for rice yield prediction on clustered regions of Bangladesh. IEEE. 2014;2014:8–13.

[9] MotiurRahman M, Haq N, Rahman RM. Application of data mining tools for rice yield prediction on clustered regions of Bangladesh. IEEE. 2014;2014:8–13.

[10] ShakilAhamed, A. T. M., Mahmood, N. T., Hossain, N., Kabir, M. T.,Das, K., Rahman, F., and Rahman, R. M. (2015). "Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh". IEEE/ACIS 16th International Conference on Software Engineering,ArtificialIntelligence,Networkingand Parallel/Distributed Computing (SNPD). doi:10.1109/snpd.2015.7176185,Vol.12,No.2,pp.12-36.

[11] Ng, R. T., and Jiawei Han. (2002). "CLARANS: a method for clustering objects for spatial data mining". IEEE, Transactions on Knowledge and Data Engineering. doi:10.1109/tkde.2002.1033770. Vol.14,No 5, pp.1003-1016.