International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 06 | June - 2021 ISSN: 2582-3930

CROP YIELD PREDICTION BASED ON ENSEMBLE MODEL

D.Vinodha[1], Aravinda bose S[2], Gokulnath S[3], Magesh S [4]

[1]Assistant Professor, Dept. of Computer Science and Engineering, S. A. Engineering College, Chennai, India [2,3,4]Student, Dept. of Computer Science and Engineering, S. A. Engineering College, Chennai, India

---

Abstract - India is a country where agriculture and agriculture related industries are the dominant source of living for the people. Agriculture is in poor condition since before comparing previous years. The vital reason for this is without a well organized pattern about farming and proper instructions to the farmers. Agriculture is an extensive source of economy of the country. Apart from this India also ail from natural disasters like flood and drought which forfeit the crops. One's strategy should be spot on while harvesting the crops, factors like season, soil moisture and weather condition should be well planned and also when to harvest the crop to get the maximum yield. In this project study, we provide an effective recommendation system for the farmers using machine learning techniques. In the proposed system we use ensemble based hybrid model to make the classifier strong. The input dataset is been obtained from the public repository. The input historical dataset in provided to machine learning platforms for data processing. The clustering of data is performed using k-means clustering and for classification of data we use knn and regression techniques to define the conditions and solutions. This would provide effective recommendation for the farmers in defining the best crop which can be farmed to obtain huge profitable yield analyzing the environmental factors.

.

1.INTRODUCTION

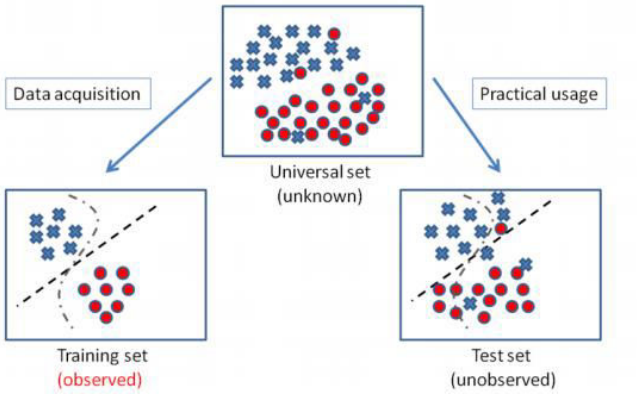
Agriculture is the one amongst the substantial area of interest to society since a large portion of food is produced by them. Currently, many countries still experience hunger because of the shortfall or absence of food with a growing population. Expanding food production is a compelling process to annihilate famine. Developing food security and declining hunger by 2030 are bene_cial critical objectives for the United Nations. Hence crop protection; land assessment and crop yield prediction are of more considerable signi_cance to global food production [1]. A country's policymaker depends on precise forecast, to make appropriate export and import assessments to reinforce national food security. Cultivators

2. EXTRACTION

A. MACHINE LEARNING

"Optimizing a performance criterion using example data and past experience", said by E. Alpaydin, gives an easy but faithful description about machine learning. In machine learning, data plays an indispensable role, and the learning algorithm is used to discover and learn knowledge or properties from the data. The quality or quantity of the dataset will affect the learning and prediction performance. The textbook (have not been published yet) written by Professor Hsuan-Tien Lin, the machine learning course instructor in National Taiwan University (NTU), is also titled as "Learning from Data", which emphasizes the importance of data in machine learning. Fig. 1 shows an example of two-class dataset.

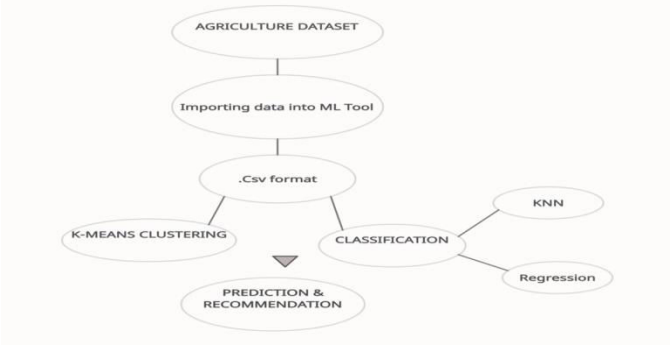
Training Set and Test Set In machine learning, an unknown universal dataset is assumed to exist, which contains all the possible data pairs as well as their probability distribution of appearance in the real world. While in real applications, what we observed is only a subset of the universal dataset due to the lack of memory or some other unavoidable reasons. This acquired dataset is called the training set (training data) and used to learn the properties and knowledge of the universal dataset. In general, vectors in the training set are assumed independently and identically sampled (i.i.d) from the universal dataset. In machine learning, what we desire is that these learned properties can not only explain the training set, but also be used to predict unseen samples or future events. In order to examine the performance of learning, another dataset may be reserved for testing, called the test set or test data. For example, before final exams, the teacher may give students several questions for practice (training set), and the way he judges the performances of students is to examine them with another problem set (test set).



3. DOMAIN INTRODUCTION

Data Mining is widely applied to agricultural issues. Data Mining is used to analyze large data sets and establish useful classifications and patters in the data sets. The overall goal of the Data Mining process is to extract the information from a data set and transform it into understandable structure for further use. This paper analyzes the crop yield production based on available data. The Data mining technique was used to predict the crop yield for maximizing the crop productivity.

Architecture of Crop yield prediction process



To start with any data mining problem, it is first necessary to bring all the data together. The data used for this proposed work are obtained for the years from 2000 to 2012 for district of Tamil Nadu in India. The preliminary data collection is carried out for districts of Tamil Nadu in India. Each area in this collection is identified by the respective longitude and latitude of the region. The data are taken in nine input variables. The variables are Year, District, Crop, Area, Tanks, Bore Wells, Open Wells, Production and Yield. Table 1 shows the description of input variables

SPECIFICATION OF MODULES

Dataset Acquisition-The climate data obtained from indianwaterportal.org and the crop production data obtained from faostat3.fao.org is taken into account for the study. The climate data contains various variables which are responsible for the rainfall for a specific region and the quantum of crop production for that region is taken into account for this work. The historical climate data of the CSV type with a record for every line of text alike to the data for a given month of an year. The crop production data of the CSV type with a record for each line of text belongs to the data for a specific month 1

of an year. In the preprocessing stage, though there are many measured parameters available in the raw climate dataset, the less relevant features responsible for the study are ignored and the important features are only taken into account.

Clustering is one of the most common exploratory data analysis technique used to get an intuition about the structure of the data.
Kmeans algorithm is an iterative algorithm that tries to partition the dataset into *K* pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to **only one group**. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum. The less variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster.

4.EXISTING SYSTEM

In existing system, the farming land properties keep on changing as per the cultivation and environmental changes. Hence machine learning based historical data analysis of that farm and crop production is essential to recommend the farmer which crop has to be cultivated in the specific farm based on the previous data. Lack of this system only lead to financial and time loss for the farmers costing their life's.

4. PROPOSED SYSTEM

In the proposed system, we provide effective farmer recommendation system analyzing the previous historical data. For processing the huge data we need machine learning tools for analyzing and predicting the solutions. In this project study, we provide an effective recommendation system for the farmers using machine learning techniques. In the proposed system we use ensemble based hybrid model to make the classifier strong. The input dataset is been obtained from the public repository. The input historical dataset in provided to machine learning platforms for data processing. The clustering of data is performed using k-means clustering and for classification of data we use knn and regression techniques to define the conditions and solutions, So that machine itself can match the condition for each clustered data and provide the respective solution for it. This would provide effective recommendation for the farmers in defining the best crop which can be farmed to obtain huge profitable yield analyzing the environmental factors.

5. RESULTS & DISCUSSION

The proposed system has provided the information about the crop yield prediction by considering the historical dataset for suitable crops in various landforms using machine learning model (Ensemble model).

6. CONCLUSIONS

The Results shows that we can attain an accurate crop yield prediction using the ensemble model comparing the output efficiency of two machine learning algorithms. It is suitable for massive crop yield prediction in agricultural planning. This makes the farmers to take the right decision for right crop such that the agricultural sector will be developed by innovative ideas. For future work, the crop prediction process will also consider the biological factors.

REFERENCES

[1] Toshichika Iizumia , Yonghee Shinb , Wonsik Kima , Moosup Kimb , Jaewon Choib, "Global crop yield forecasting using seasonal climate information from a multi-model ensemble", International journal on climate services, 2018.

[2] David B. Lobell and Gregory P. Asner, "Comparison of Earth Observing- 1 ALI and Landsat ETM+ for Crop identification and Yield prediction in Mexico", international journal on geo science an remote sensing, June 2003

[3] P. S. Vijayabaskar , Sreemathi.R , Keertanaa.E "Crop prediction using predictive analysis", international conference on computation of power energy information and communication, 2017

[4] Yogesh gandge, Sandhya, "A Study on Various Data Mining Techniques for Crop Yield Prediction", international conference on Electrical, Electronics, Communication, Computer and Optimization Technique, 2017

[5] Jin Chaun, Qin Qiming, Zhu Lin, Nan Peng, Abduwasit Ghulam, "TVDI based Crop Yield Prediction model for Stressed Surface", international conference on Autonous Region, 2007.

[6] Joon-Goo Lee, Haedong Lee, Aekyung moon, "Segmentation method of COI for Monitoring and prediction of the crop growth", international conference on embedded system research section, 2014.

[7] Monali paul, santosh k. vishwakarma, ashok verma, "Analysis of soil behavior and prediction of crop yield using data mining approach", International conference on computational intelligence and Communication Networks 2015.

[8] K Aditya shastr , Sanjay H A and Kavya H "A novel data mining approach for soil and crop classification" The 9th International conference on computer Science & Education, August 22-24 2014.

[9] Liying yang, "classifier selection for ensemble learning based on accuracy and diversity" published by Elseiver ltd. Selection and peer- review under responsibility of 2015.

[10] Ayeman E Khedr, Mona Kadry, Ghada Walid, "proposed framework for implementing data mining techniques to enhance decisions in agriculture sector applied case on food security", Information center ministry of agriculture, Egypt, International conference 2014.

[11] Sofianita Mutalib, S-N-Fadhlun Jamian , Shuzlina Abdul-Rehman and Azlinah Mohamed"soil and crop classification of application of self organizing map and k-means" 2010 10th international conference on intelligent systems design and applications.