

Heuristic Prediction of Crop Yield using Machine Learning Technique

S. Pavani, Augusta Sophy Beulet P.



Abstract: - Vast research has been done and several attempts are made for application of Machine learning in agricultural field. Major challenge in agriculture is to increase the production in the farm and deliver it to the end customers with best possible price and good quality. It is found that at least 50 percent of the farm produce never reach the end consumer due to wastage and high-end prices. Machine learning based solutions developed to solve the difficulties faced by the farmers are being discussed in this work. The real time environmental parameters of Telangana District like soil moisture, temperature, rainfall, humidity are collected and crop yield is being predicted using KNN Algorithm.

Keywords: Agriculture, Crop Yield Prediction, KNN, Machine learning.

I. INTRODUCTION

Agriculture plays an essential role within the economy of the state and therefore the higher production of this sector is important for inclusive improvement. Enhanced improvement in Agriculture continues to be the core agenda of the government and possesses center stage in state economy embodying three thrust zones viz., (i) to promote inclusive development, (ii) to upgrade provincial salary and (iii) to sustain food and nourishment security. Approximately 55.49% of the population of the state is reliant for livelihoods in one or the other, there is a profound need to raise the farmer's income and ensure sustainable growth in Telangana to reduce poverty. The share of agriculture sector to the Gross State Value Added (GSVA) in 2015-16 is 14.7 percent and at current prices in 2016-17 is 15.3 percent[1]. The pressure on agricultural land is growing over the years for industrialization, urbanization, housing, infrastructure and other factors. Such kinds of conditions cause agricultural land to be turned for commercial purposes. The scope for growth of the area available for cultivation is reducing a lot that which results in the decrease in sown net field. Prediction of the yield of the crop in advance for particular region depending on the climatic conditions and other factors which contribute to more yield is important, which will make farmers to select the crop or to get better yield of their particular crops in the particular region[2]. Data mining means distinguishing concealed examples from enormous datasets and setting up a relationship between them to take care of the issue through information analysis[3]. The use of data mining in rural areas has brought advantages in the field of research[4]. Classification is important to build up the

essentials in any field of science. It can help to determine the distinction between objects and ideas.

It moreover gives vital data through which research can be made in an efficient way. Soil is one of the primary segments for yielding harvests in agricultural discipline[5]. Soil moisture, humidity, temperature, rainfall plays a vital role in improving the crop yield.

Machine learning is one of the best strategies for predicting unknown values. This work manages the expectation of crop yield using machine learning. This work collects dataset from Telangana State Development Planning Society. A real dataset which consists of rainfall, temperature, soil moisture, humidity for all the districts in Telangana state for one day and one week has been used in this work[6]. This application's evaluations help farmers predict crop yields. Support Vector Machine (SVM), Regression analysis, K Nearest Neighbor (KNN), clustering and various types of techniques are used for prediction. KNN is used in this work. In machine learning, classification is the problem of detecting to which set of the categories, a new object or observation belongs to on the basis of the data set which contains the instances or the observations of the known category[7]. KNN is a language learning algorithm that is non-parametric. The aim is to use a model where information focuses are clustered in a few groups in order to predict the classification of another instance [8]. KNN Algorithm depends on the closeness of the feature: After our training set, how firmly out of test highlights determine how we manage a given point of data. KNN can be utilized for classification-the yield or the output is a class membership (predicts a class-a discrete esteem or value). An element is distinguished by a larger proportion of its neighbors' vote, with the entity being divided among its closest neighbors to the most regular class. It can also be used for regression-output, which is the item's reward (predicts unceasing qualities). This calculation is the average (or middle) of its nearest neighbors' estimates of k. A commonly used metric for distance calculation in KNN algorithm is Euclidean distance.

This technique can be used to predict crop yield with more accuracy for years of data together[9]. The main principle reason of the work is to predict the crop yield in prior according to the details of the factors required for good crop yield.

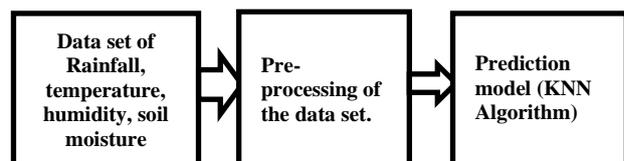


Fig. 1. Flow graph of the methodology

Manuscript published on 30 December 2019.

* Correspondence Author (s)

Pavani Sunkara, School of Electronics Engineering, Vellore institute of Technology, Chennai, India. Email: sunkarapavani1@gmail.com

Augusta Sophy Beulet P.*, School of Electronics Engineering, Vellore Institute of Technology, Chennai, India. Email: augustasophyt.p@vit.ac.in

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

II. METHODOLOGY

A. Datasets

The data sets of different districts of the Telangana state are collected from Telangana State Development Planning Society. The important factors that determine the crop yield are temperature, humidity, soil moisture and rainfall. The samples contain daily recorded data of the above-mentioned factors and soil moisture of about depth of 4 inches. These samples are taken for the month of May 2019. These factors for every district in Telangana state are collected for one week and the crop yield is being predicted by using Machine learning technique. KNN algorithm is used to classify and predict the crop yield[10]. All the available data set is divided into a window of five among which four are the input factors to the prediction model and fifth one is crop yield. Among the data set available maximum of the data is used for training and the remaining data is used for testing. The machine learning technique KNN algorithm is used for prediction of crop yield[11].

B. Learning algorithm

The k-nearest neighbor (k-NN) method is a data mining technique considered to be among the top five techniques for data mining. The method k-NN uses the common definition of “Cicero pares cum paribus facillim e congregantur” (birds of a feather flock together or literally equals with equalseasily associate). It attains the properties of new variable with the help of properties of existing variables. It is applicable in classification as well as regression problems. It endeavors to characterize an obscure example dependent on the known characterization of its neighbors. Give us a chance to assume that a lot of tests with realized grouping is accessible, the alleged preparing set[12]. Naturally, each example ought to be grouped comparably to its encompassing examples. In this way, on the off case that the order of an instance is obscure, by observing the ordering of its nearest neighbor tests it could be expected at that point. Depending on an obscure example and a set of preparations, each of the separations between the obscure example and each of the instances in the collection of preparations can be interpreted. [13].The division with the least esteem is compared to the instance in the preparation that is nearest to the possess sample unidentified. Therefore, the obscure example might be ordered dependent on the characterization of this closest neighbor. So with respect to the application of KNN algorithm towards prediction of crop yield, the nearest neighbors of a particular point(crop yield) like temperature, humidity, rainfall and soil moisture are considered, if these factors have the enough values required for a crop yield then it can be considered as a good crop yield depending on the factors[14]. We can implement KNN by using the below mentioned steps:

- 1) Load the data set.
- 2) Initialize the ‘k’ value.
- 3) For getting the anticipated class, repeat from one to all the numbers of training data set.
- 4) Compute the distance between test data and each line of training information. Here the Euclidean distance is utilized, since it’s the most prominent technique. Different measurements that can be utilized are Chebyshev, cosine, and so forth.

- 5) Classify the computed distances in ascending order based on distance values.
- 6) Get top ‘k’ rows from the classified array.
- 7) Get the most continuous class of these lines.
- 8) Return the anticipated class.

Aside from utilizing as grouping tool like depicted above, KNN has a few further applications [15].

It is prevalent in pursuit applications. For instance, on the off chance that one needs to locate the most comparable records to a specific report, for example for identifying copyright, KNN calculation is perfect. The calculation additionally has numerous utilizations in data mining and AI[16]. One specific utilization of K Nearest Neighbor is in anomaly detection (peculiarity identification). It is the distinguishing proof of things, occasions, or perceptions which don't fit in with the normal examples or different things in a dataset.

C. Prediction of Crop Yield through K-NN

This work represents a review of K-NN technique for the early prediction of cropyield. K-NN analysis is used for predicting the unknown parameter from the known parameters.

In this work we are considering rainfall, temperature, humidity and soil moisture as input parameters which are the main parameters to be considered for a good crop yield, although there are many other factors that can be considered[17]. The unknown value of crop yield can be predicted from the nearest known values of the nearest neighbors by calculation of Euclidean distance between them. Then we would be able to predict crop yield for given rainfall, temperature, humidity and soil moisture parameters. To measure the distance between points in a feature space, various distance functions can be used, in which the Euclidean distance function is the most widely used one[18]. Let p and q are represented as feature vectors. To calculate the distance between p and q, the Euclidean metric is generally used by if a=(a1, a2) and b=(b1,b2) then the distance is given by

$$d(a, b) = \sqrt{(b1 - a1)^2 + (b2 - a2)^2}$$

III.RESULTS AND DISCUSSION

The main purpose of the implementation of KNN is to show how it helps in predicting the crop yield when the necessary input parameters are given. A slice of the training data is shown in the table1 below.

TABLE I TRAINING DATA

s.no	rainfall	soil moisture	temperature	Humidity
1	0.8	0	40.4	29.4
2	0	62.67	42	27.6
3	0.5	2.36	44.5	31.7
4	1	3.7	43.2	26.7
5	0.3	15.7	43.3	16.2
6	0.1	25.31	41.8	26.7





Fig 2: Chart representing how new sample can be predicted from the training data.

Let us consider a simple example as shown in the above chart. “K” is the KNN algorithm, from which we wish to consider vote from the nearest neighbors. In the fig:2 above soil moisture, temperature & humidity are considered. The similarity is calculated based on distance function[19]. There are many distance metrics that can be used to calculate the distance, but Euclidean distance is mostly preferred[20]. We use distance metric to calculate the similarity between new sample and training set cases to find the closest values to the new value considered.

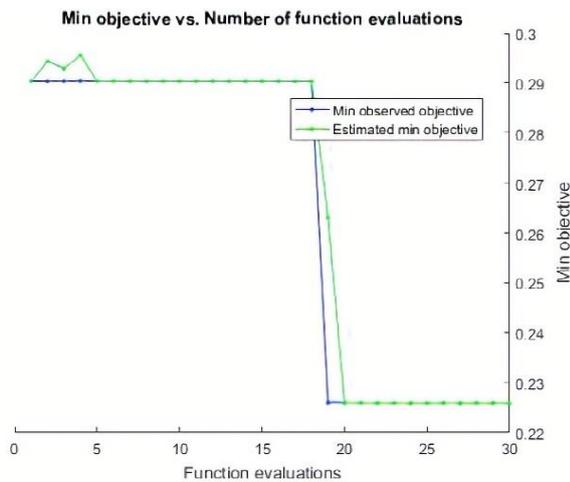


Fig.3 Performance Results

In the fig 3 shown we can observe the estimated crop yield and observed crop yield in green and blue lines respectively. The difference between them is negligible and can be used to predict the crop yield as the accuracy level is high.

IV. CONCLUSION

A model of machine learning to predict plant yield is proposed and giving reasonable crop yield suggestion for particular districts in Telangana. The research has been done on soil moisture, temperature, humidity and rainfall datasets of all the districts of Telangana State. By applying machine learning algorithm the model has been tested. K-NN suggests suitable accuracy in crop yield prediction. The well-constructed data set and the machine learning algorithm supports the proposed model. In future, providing other

factors that greatly influence the crop yield is our concern, also more data of all these parameters of different seasons in the state will be added to make this model more accurate and efficient.

REFERENCES

1. Statistical yearbook 2017, Directorate of Economics and Statistics, Government of Telangana.
2. Nivetha, R. Yamini, and C.Dhaya, Developing a Prediction Model for Stock Analysis, Technical Advancements in Computers and Communications (ICTACC), 2017 International Conference on IEEE, 2017.
3. Hemantkumar Wani, Nilima Ashtankar, An Appropriate Model Predicting Pest/Diseases of Crops Using Machine Learning Algorithms, 2017 International Conference on Advanced Computing and Communication Systems (ICACCS -2017), Jan. 06 – 07, 2017, Coimbatore, INDIA.
4. Haiguang Wang, Guanlin Li, ZhanhongMa, Xiaolong Li, “Image Recognition of Plant Diseases Based on Backpropagation Networks”, 5th International Congress on Image and Signal Processing (CISP 2012)201.
5. Introduction to Data Mining, P.N. Tan, M. Steinbach, V. Kumar, Addison Wesley.
6. Shambhavi Dighe , Anuja Jagdale , Amar Chadchankar, Crop Production Prediction System, International Journal of Research in Engineering, Science and Management Volume-1, Issue-12, December-2018 www.ijresm.com | ISSN (Online): 2581-5792.
7. An efficient k-means clustering algorithm: Analysis and implementation, T. Kanungo, D. M. Mount, N. Netanyahu, C.Piatko, R. Silverman, and A. Y. Wu, IEEE Trans. PatternAnalysis and Machine Intelligence, 24 (2002), 881-892.
8. Hema Geetha N, A survey on application of data mining techniques to analyze the soil for agricultural purpose, Computing for Sustainable Global Development (INDIAcom), 3rd International Conference on IEEE, 2016.
9. Chandrasegar Thirumalai, IEEE Member, M Lakshmi Deepak, Heuristic Prediction of Rainfall Using Machine Learning Techniques, International Conference on Trends in Electronics and Informatics ICEI 2017.
10. S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh, Machine learning approach for forecasting crop yield based on climatic parameters,2014 International Conference on Computer Communication and Informatics (ICCCI -2014), Jan. 03 –05, 2014, Coimbatore, INDIA.
11. Ramesh, V. and Ramar, K., 2011. Classification of agricultural land soils: a data mining approach. *Agricultural Journal*, 6(3), pp.82-86.
12. Kentaro Kuwata and Ryosuke Shibasaki, Estimating crop yields with deep learning and remotely sensing data, IIS, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, JAPAN.
13. Rani Pagariya and MahipBartere, “Review paper on identification of plant diseases using image processing technique”, 2014.
14. Gautam Kaushal1, Rajni Bala2, GLCM and KNN based Algorithm for Plant Disease Detection, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 6, Issue 7, July 2017.
15. Mr A Suresh, Dr. P. Ganesh Kumar, Dr.M.Ramalatha, Prediction of major crop yields of Tamilnadu using K-means and Modified KNN, Proceedings of the International Conference on Communication and Electronics Systems (ICES 2018) IEEE Xplore Part Number:CFP18AWO-ART; ISBN:978-1-5386-47653.
16. Yan, Xiaozhen, Hong Xie, and Wang Tong, “A multiple linear regression data predicting method using correlation analysis for wireless sensor networks,” Cross strait quad-regional radio science and wireless technology conference, 2011. Vol. 2. IEEE, 2011.
17. Sk Al Zaminur Rahman, Kaushik Chandra Mitra, Soil Classification using Machine Learning Methods and Crop Suggestion Based on Soil Series,2018 21st International Conference of Computer and Information Technology (ICCIT), 21-23 December 2018.
18. HAI-YANG JIA1, JUAN CHEN1, HE-LONG YU1,2, DA-YOU LIU1, Soil fertility grading with Bayesian network transfer learning, Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao, 11-14 July 2010.

19. K.R.SriPreethaa,,S.Nishanthini,D.SanthiyaK.VaniShree ,“CropYield Prediction”,International Journal On Engineering Technology and Sciences– IJETS™ISSN(P): 2349-3968, ISSN (O):2349-3976 Volume III,Issue III, March- 2016.
20. Ana Laura Diedrichs , Facundo Bromberg, Diego Dujovne, Keoma Brun-Laguna,and Thomas Watteyne, Senior Member, IEEE, Prediction of Frost Events Using MachineLearning and IoT Sensing Devices, IEEE INTERNETOF THINGS JOURNAL, VOL. 5, NO. 6, DECEMBER 2018.

AUTHORS PROFILE



Ms. Pavani Sunkara completed her Bachelor of Engineering in the stream of Electronics Engineering from Mallareddy Group Of Institutions, Hyderabad, Telangana and her Masters in the field of Digital Systems & Computer Electronics from Sreenidhi Institute of Technology, Hyderabad, Telangana. She has worked as VLSI project engineer in The Blackbucks Engineer & Company and as Assistant Professor with St. Mary's Group of Institution in Electronics & communication Department. She is pursuing her Ph.D at Vellore Institute of Technology, Chennai and her area of interests includes VLSI, IoT and Artificial Intelligence.



Augusta Sophy Beulet P., Being a lecturer for more than 25 years, Dr.P.Augusta Sophy, has developed a passion towards teaching Engineering subjects and guiding projects. The passion has naturally been extended towards developing the students too. Have handled more than 20 different subjects in Electronics and Communication Engineering and guided more than 50 projects in UG and PG levels. She has published 24 research papers in International Conferences and Journals. She did her research work in the field of VLSI Signal Processing in Anna University, in 2015. Received her M.E degree from Anna University and finished her B.E degree in Government College of Engineering, Tirunelveli. Now she is working as Associate Professor in VIT University, Chennai.