COMBINING CROP MODELS AND REMOTE SENSING FOR YIELD PREDICTION USING KNN ALGORITHM OVER K-MEANS ALGORITHM

B. Bitchi Reddy[1], J. Chenni Kumaran[2*]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: The combination of crop models and remote sensing to yield prediction using KNN Algorithm over K-Means Algorithm.

Materials and Methods: When implementing an accurate prediction model it might not be sufficient to just consider one or two parameters. Data about Rainfall, temperature, humidity and various other factors are collected and analyzed. This analysis will be fed to the prediction model. Based on the Previous Collected Datasets, following KNN Algorithm, K-means Algorithm, the upcoming Crop Yield can be predicted with calculations. **Results:** By using Jupyter notebook and the previous datasets of crop yield prediction, after iterating the datasets using KNN Algorithm 93% accuracy is obtained and using K-means algorithm 86% accuracy is obtained on the other hand. Since the significance is around 0.023, there is statistically a significant difference among the study group with ($p < 0.05$).

Conclusion: After using iterations with KNN algorithm it yields 93%(0.93) and K-means gives 86%(0.86). So it can be said that by using KNN Algorithm more Accuracy is obtained than K-means algorithm.

Keywords: mean accuracy, Crop Yield Prediction, K-Nearest Neighbor, K-Means Algorithm, Machine Learning.

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences. Saveetha University, Chennai, Tamilnadu, India, Pincode:602105

[2*]Project Guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Science, Saveetha University, Chennai, Tamilnadu, India, Pincode 602105.

1. Introduction

Agriculture in India has a significant history. Nowadays, India is ranked 2d worldwide in farm output.Agriculture and allied sectors like forestry and fisheries accounted for sixteen.6% of the GDP 2009, approximately 50% of the whole group of workers (Squires and Gaur 2021). The monetary contribution of agriculture to India's GDP is regularly declining with the united states' huge-based economic boom. Agriculture hobby is a kind of business with chance. The manufacturing of vegetation depends on different factors like climatic, geographical, biological, political and economic elements. Accurate records about the nature of ancient yield of crops is crucial modeling center, which can be useful to farmers & authorities corporation for choice making method in setting up right policies related to next manufacturing (National Research Council et al. 1993). The advances in computing and factories have provided good sizes at maximum of statistics. The assignment has been to extract information from this raw statistics, statistics mining that could bridge the know-how of the data to the crop yield estimation.

This research aimed at information mining techniques and followed them to the diverse variables consisting within the database to set up if meaningful relationships can be determined and the use of fuzzy good judgment to find the condition of vegetation on diverse situations of rainfalls. K-means is one of the simplest unsupervised learning algorithms that solve the well-known problem i.e. average mean. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed apriority. The main idea is to define k centers, one for each cluster. These centers should be placed in a cunning way because different locations cause different results . ra in this period which are a reaction to many problems in terms of time, nice, coins or effort.Engineers at the moment are collaborating with farmers to create a technological solution for agricultural elements. As the growing boom in technological traits, but, there are problems as once the climatic preconditions alternate, all the conditions will be terrible for the farmers, leading to financial loss as it performs an important position (Uhrin and Supuka 2019).

Our institution is passionate about high quality evidence based research and has excelled in various domains (Vickram et al. 2022; Bharathiraja et al. 2022; Kale et al. 2022; Sumathy et al. 2022; Thanigaivel et al. 2022; Ram et al. 2022; Jothi et al. 2022; Anupong et al. 2022; Yaashikaa, Keerthana Devi, and Senthil Kumar 2022; Palanisamy et al. 2022). The research gap of existing systems for prediction of agricultural crop yield is not optimum and least accurate. Using the combination of Crop Models and Remote Sensing, the most accurate prediction is proposed using these algorithms. Research indicates that inspecting all the attributes together with soil, rainfall and so on, it will be expecting the type of crop that is appropriate for improvement in particular places. This information assists to enhance these crops within the destiny and might guide farmers (Hillmann et al. 2015). In a few elements of us a demand for a specific crop will also be more and the soil may be more likely to produce that unique crop, but due to some droughts and also lack of rain, it is able to also make the soil at suitable for growing that particular crop.analyzing and forecasting such vegetation is easier for such situations due to the fact they will know what unique crops may be produced (Huth, Robertson, and Poulton 2010). Difficult plantings and effort is not wasted either. Enhancing the satisfaction of the plant is the primary purpose of gaining a clean picture of the crop yield through the information mining techniques.Rentability and sustainability are the principal desires of precision agriculture. (Brownlee 2016)

2. Materials and Methods

Dataset Collection

When implementing an accurate prediction model it might not be sufficient to just consider one or two parameters. Data about rainfall, temperature, humidity and various other factors are collected and analyzed. This analysis will be fed to the prediction model. Following KNN Algorithm, K-means algorithm based on the previously collected datasets and it can be predicted the upcoming crop yield with results.

Agricultural Techniques

K-means is one of the simplest unsupervised learning algorithms that solve the well-known problem i.e. average mean. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed apriority. The main idea is to define k centers, one for each cluster. These centers should be placed in a cunning way because different locations cause different results.

K-means Algorithm

Mean = (Sum of all N numbers) / N
Step1: Start
Step2: Take N inputs
Step3: [Initialize] Sum := 0
Step4: Repeat for i := 1 to N by 1
 Sum := DATA[i] + Sum
 [End of loop]
Step5: mean := Sum / N

Step6: Write: 'Mean of N number is mean'
Step7: Stop

K-means based to find the prediction of production, the GUI is created for K-means based by using K-means and GUI coding where k is the number of years which lie between 1 to 15. suppose it is needed to find the crop prediction for the year 2015 then k=1 will predict the value on the basis of previous year data, similarly it will predict the rainfall and crop production for next year.
When K=1 then area is 17239 and production is 9359 and total annual rainfall will be 1039.1mm. This process works on the basis of average from data, prediction is done by last k year value, and average among the area, production, and annual rainfall, this can be found by applying formula.

Testing and Analysis
First it is needed to calculate the coordinate values with the previous datasets by using KNN algorithm and K-means. The purpose of the test was to find the workings of the KNN algorithm and how it will predict the yield when three parameters were given as input. Input data is given as follows: Location: Mangalore, Soil-Type: Coastal alluvial and Area: 1395 cents, Weather Conditions: temperature, humidity, datasets. The system predicted coconut and cocoa as two potential crops with the mean accuracy of 91.83%. While testing for Kodagu district the soil type is laterite soil and the area given was 1395 cents then the system predicted cardamom and Pepper as two potential crops and noted the mean accuracy of 86.76%. The gadget facilitates in heading off the use of sensors and reduces needless fees. This system results in an efficient utilization of time and fee. A key factor of crop prediction is to perceive a suitable crop quickly and suggest to the farmer as to which crop to grow. This system allows in collecting all vital facts and giving a version of output which is not handiest will increase current crop values then it gives pricey advantage but additionally safeguards destiny profitability. The accuracy part of the machine is cited as iteration by using previous historical data sets.

3. Results

By using Jupyter notebook and the previous datasets of crop yield prediction, after iterating the datasets KNN algorithm perform better then K - means algorithm,
Accuracy for Crop Yield Prediction by using KNN algorithm gives 91.83% mean accuracy and Decision Tree algorithm gives 86.76% mean accuracy based on iterations samples is shown in **Table 1.**

The **Table 2.** gives the overview accuracy of crop yield prediction by using KNN algorithm and K-means algorithm based on the final average iterations.

Table 3. depicts the Independent Sample T- test Result is applied for dataset fixing confidence interval as 95% and level of significance as 0.05 (KNN appears to perform significantly better than K-means algorithm).
Thus, using KNN yielding 93% and using the K-means Algorithm getting 86% mean accuracy.

4. Discussion

In this study the iteration based on the previous historical datasets are considered for getting the accurate value of crop yield prediction (Teufelberger 2019) using KNN algorithm and K-means on the basis of anaconda navigator (Jupituer notebook) and SPSS. After getting many iterations it compared all the selective algorithms and decided to get the accurate values of crop yield prediction. The implementation of the machine was to learn about vegetation and agriculture and locate a green way of harvesting (Food and Agriculture Organization of the United Nations 2018). Decision support models are broadly used to extract significant crop features for prediction. Precision agriculture focuses on monitoring (Matcham et al. 2022) (sensing technologies), management information systems, variable rate technologies, and responses to inter- and intra-variation in cropping systems. The benefits of precision agriculture involve increasing crop yield and crop quality, while reducing the environmental impact (Teufelberger 2019).
Have a look at focuses on the rural datasets obtained from various portals belonging to some districts of Karnataka in India (Pradeep et al. 2019). Datasets ordered in a nicely based manner. The KNN algorithm rules are used for the prediction model and crop yield prediction and its accuracy are received. Destiny is brilliant for the implementation of machine learning algorithms (Pradeep et al. 2019) inside the area of crop manufacturing (Shahrajabian, Cheng, and Sun 2022) and it is hoped to enforce more advanced algorithms so that the gadget turns into greater efficiency. It is still in hope to make device prediction extra strong and achieve excessive mean accuracy with the help of extra datasets and advanced algorithms (Toscano 2020).

5. Conclusion

After coming through so many iterations then KNN algorithm gets 93% (0.93) and K-means gets 86%

(0.86). So, KNN Algorithms yield more accuracy than K-means.

Declarations

Conflict of Interests

No conflict of interest in this manuscript.

Author Contribution

Author BBR was involved in data collection, data analysis, manuscript writing. Author JCK was involved in conceptualization, guidance and critical review of manuscript.

Acknowledgments

The authors would like to express their gratitude towards Saveetha School of engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

1. Indira Agrotech, Chennai.
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

6. References

Anupong, Wongchai, Lin Yi-Chia, Mukta Jagdish, Ravi Kumar, P. D. Selvam, R. Saravanakumar, and Dharmesh Dhabliya. 2022. "Hybrid Distributed Energy Sources Providing Climate Security to the Agriculture Environment and Enhancing the Yield." *Sustainable Energy Technologies and Assessments*. https://doi.org/10.1016/j.seta.2022.102142.

Bharathiraja, B., J. Jayamuthunagai, R. Sreejith, J. Iyyappan, and R. Praveenkumar. 2022. "Techno Economic Analysis of Malic Acid Production Using Crude Glycerol Derived from Waste Cooking Oil." *Bioresource Technology* 351 (May): 126956.

Brownlee, Jason. 2016. *Master Machine Learning Algorithms: Discover How They Work and Implement Them From Scratch*. Machine Learning Mastery.

Food and Agriculture Organization of the United Nations. 2018. *E-Agriculture in Action: Drones for Agriculture*. Food & Agriculture Org.

Hillmann, Felicitas, Marie Pahl, Birte Rafflenbeul, and Harald Sterly. 2015. *Environmental Change, Adaptation and Migration: Bringing in the Region*. Springer.

Huth, Neil I., Michael J. Robertson, and Perry L. Poulton. 2010. "Regional Differences in Tree - Crop Competition due to Soil, Climate and Management." *Crop and Pasture Science*. https://doi.org/10.1071/cp09254.

Jothi, K. Jeeva, K. Jeeva Jothi, S. Balachandran, K. Mohanraj, N. Prakash, A. Subhasri, P. Santhana Gopala Krishnan, and K. Palanivelu. 2022. "Fabrications of Hybrid Polyurethane-Pd Doped ZrO2 Smart Carriers for Self-Healing High Corrosion Protective Coatings." *Environmental Research*. https://doi.org/10.1016/j.envres.2022.113095.

Kale, Vaibhav Namdev, J. Rajesh, T. Maiyalagan, Chang Woo Lee, and R. M. Gnanamuthu. 2022. "Fabrication of Ni–Mg–Ag Alloy Electrodeposited Material on the Aluminium Surface Using Anodizing Technique and Their Enhanced Corrosion Resistance for Engineering Application." *Materials Chemistry and Physics*. https://doi.org/10.1016/j.matchemphys.2022.125900.

Matcham, F., E. Carr, K. M. White, D. Leightley, F. Lamers, S. Siddi, P. Annas, et al. 2022. "Predictors of Engagement with Remote Sensing Technologies for Symptom Measurement in Major Depressive Disorder." *Journal of Affective Disorders*, May. https://doi.org/10.1016/j.jad.2022.05.005.

National Research Council, Board on Science and Technology for International Development, Board on Agriculture, and Committee on Sustainable Agriculture and the Environment in the Humid Tropics. 1993. *Sustainable Agriculture and the Environment in the Humid Tropics*. National Academies Press.

Palanisamy, Rajkumar, Diwakar Karuppiah, Subadevi Rengapillai, Mozaffar Abdollahifar, Gnanamuthu Ramasamy, Fu-Ming Wang, Wei-Ren Liu, Kumar Ponnuchamy, Joongpyo Shim, and Sivakumar Marimuthu. 2022. "A Reign of Bio-Mass Derived Carbon with the Synergy of Energy Storage and Biomedical Applications." *Journal of Energy Storage*. https://doi.org/10.1016/j.est.2022.104422.

Pradeep, N., Sandeep Kautish, C. R. Nirmala, Vishal Goyal, and Sonia Abdellatif. 2019. *Modern Techniques for Agricultural Disease Management and Crop Yield Prediction*. IGI Global.

Ram, G. Dinesh, G. Dinesh Ram, S. Praveen Kumar, T. Yuvaraj, Thanikanti Sudhakar

Babu, and Karthik Balasubramanian. 2022. "Simulation and Investigation of MEMS Bilayer Solar Energy Harvester for Smart Wireless Sensor Applications." *Sustainable Energy Technologies and Assessments*. https://doi.org/10.1016/j.seta.2022.102102.

Shahrajabian, Mohamad Hesam, Qi Cheng, and Wenli Sun. 2022. "The Effects of Amino Acids, Phenols and Protein Hydrolysates as Biostimulants on Sustainable Crop Production and Alleviate Stresses." *Recent Patents on Biotechnology*, April. https://doi.org/10.2174/1872208316666220412133749.

Squires, Victor R., and Mahesh K. Gaur. 2021. *Food Security and Land Use Change under Conditions of Climatic Variability: A Multidimensional Perspective*. Springer.

Sumathy, B., Anand Kumar, D. Sungeetha, Arshad Hashmi, Ankur Saxena, Piyush Kumar Shukla, and Stephen Jeswinde Nuagah. 2022. "Machine Learning Technique to Detect and Classify Mental Illness on Social Media Using Lexicon-Based Recommender System." *Computational Intelligence and Neuroscience* 2022 (February): 5906797.

Teufelberger, Peter. 2019. *Crop Yield Prediction in Agriculture Based on Long Short-Term Memory*.

Thanigaivel, Sundaram, Sundaram Vickram, Nibedita Dey, Govindarajan Gulothungan, Ramasamy Subbaiya, Muthusamy Govarthanan, Natchimuthu Karmegam, and Woong Kim. 2022. "The Urge of Algal Biomass-Based Fuels for Environmental Sustainability against a Steady Tide of Biofuel Conflict Analysis: Is Third-Generation Algal Biorefinery a Boon?" *Fuel*. https://doi.org/10.1016/j.fuel.2022.123494.

Toscano, Piero. 2020. *Remote Sensing Applications for Agriculture and Crop Modelling*. MDPI.

Uhrin, Peter, and Ján Supuka. 2019. "Assessment of Urban Trees Adaptability in Changed Climatic Conditions in Cities Using by Visual Growing and Ecophysiological Traits." https://doi.org/10.15414/2019.9788055220116.

Vickram, Sundaram, Karunakaran Rohini, Krishnan Anbarasu, Nibedita Dey, Palanivelu Jeyanthi, Sundaram Thanigaivel, Praveen Kumar Issac, and Jesu Arockiaraj. 2022. "Semenogelin, a Coagulum Macromolecule Monitoring Factor Involved in the First Step of Fertilization: A Prospective Review." *International Journal of Biological Macromolecules* 209 (Pt A): 951–62.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Algal Biofuels: Technological Perspective on Cultivation, Fuel Extraction and Engineering Genetic Pathway for Enhancing Productivity." *Fuel*. https://doi.org/10.1016/j.fuel.2022.123814.

Tables snd Figures

Table 1. Accurate Table for Crop Yield Prediction By Using KNN algorithm gives 93.33% accuracy and K-means algorithm gives 86.76% accuracy based on iterations samples.

No.of iterations	KNN Algorithm	K-means Algorithm
1.	93.33	86.76
2.	91.20	82.10
3.	94.60	80.10
4.	93.30	86.90
5	93.00	84.00
6	90.20	79.40
7	92.60	78.40
8	95.20	86.80

9	93.50	85.55
10	93.56	85.67

Table 2. The overview accuracy of crop yield prediction by using KNN algorithm and K-means algorithm based on the final average iterations.

	Groups	**N**	**Mean**	**Std.Deviation**	**Std.Error Mean**
Accuracy	KNN Algorithm	10	92.9922	1.54468	0.51489
	K means algorithm	10	86.5333	2.96206	0.98735

Table 3. Independent Sample T- test Result is applied for dataset fixing confidence interval as 95% and level of significance as 0.05 (KNN appears to perform significantly better than K-means algorithm).

		Lavene's test for equality of variances		**T-test for Equality of Means**					**95% confidence interval of the difference**	
		F	**Sig**	**t**	**df**	**sig(2 tailed)**	**Mean diff**	**Std.error**	**Lower**	**Upper**
Accuracy	**Equal Variances assumed**	0.951	0.023	5.737	16	<0.01	<0.01	13.49111	2.35173	8.50566
	Equal Variances not assumed			5.737	12.304	<0.01	<0.01	13.49111	2.35173	8.38113

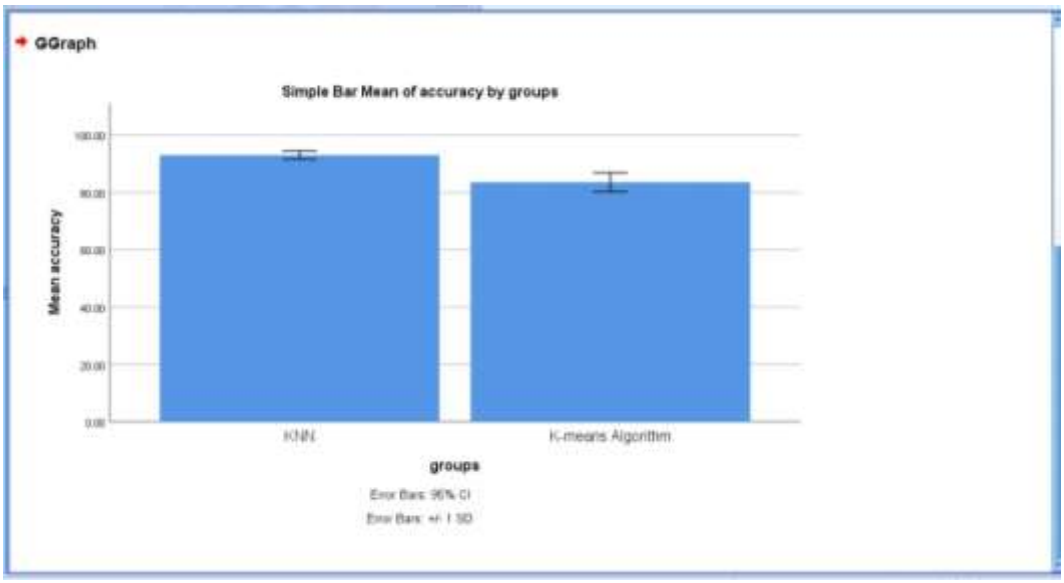

Fig. 1. The bar graph With Accurate values indication. It shows the Accuracy value of Testing algorithms (KNN and K-means algorithm).