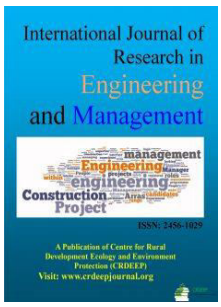
Vol. 4. No. 4. 2021

©Copyright by CRDEEP Journals. All Rights Reserved.

Contents available at:
www.crdeepjournal.org

International Journal of Research in Engineering & Management (ISSN: 2456-1029)(SJIF: 2.228)
A Peer Reviewed Journal



Full Length Research Paper

Business Analytics Application for Crop Prediction-An Empirical Study

B. Lavanya[1] and M. Hem Kumar Reddy[2*]

[1-] *Assistant Professor, SMS- Chaitanya Bharathi Institute of Technology (Autonomous), Hyderabad, India.*

[2-] *MBA II Year Student, SMS- Chaitanya Bharathi Institute of Technology (Autonomous), Hyderabad, India.*

ARTICLE INFORMATION

Corresponding Author:
M. Hem Kumar Reddy

Article history:
Received: 14-06-2021
Accepted: 18-06-2021
Published: 23-06-2021

Key words:
Crop Prediction ,Clustering , Logistic Regression.

ABSTRACT

Agriculture is one of the most essential occupations as it meets the daily requirement of us and widely practiced occupations in India. It has a vital role in the development of our country. Around 60 percent of the total land in India is used for agriculture to meet the needs of 1.2 billion people, so improving the crop production is an important need for the country .Production of crop depends on various factors like soil type, temperature, humidity, pH value, etc. The objectives of the study are to develop a Machine Learning Model for crop prediction by using Python Coding, various Statistical Concepts and to predict the best crop to be grown based on Soil Composition, Temperature, Ph-Value and Rainfall. The proposed project will increase the productivity, profits and reduce loss. Our project is a recommendation system which make uses of different machine learning Methodologies such as Clustering, Logistic Regression that it recommends the suitable crops based on the input soil parameters and climatic conditions. This entire process is known as "Crop Prediction". This system thus reduces the financial losses faced by the farmers by planting the incorrect crops and it also helps the farmers to churn to plant new crops if needed. Finally, I conclude that a machine Learning Model has been developed, which predicts the correct crop to be grown. The model has been developed by using secondary data. Real-time data can also be then revenue for the farmer's increases, decreases farmer's suicides,etc.

Introduction

Data analytics is the science of analyzing raw data in order to make conclusions about that information. Many of the techniques and processes of data analytics have been automated into mechanical processes and algorithms that work over raw data for human consumption. As the process of analyzing raw data to find trends and answer questions, the definition of data analytics captures its broad scope of the field. However, it includes many techniques with many different goals. The data analytics process has some components that can help a variety of initiatives. By combining these components, a successful data analytics initiative will provide a clear picture of where you are, where you have been and where you should go. The Significance of this research paper is to get more insights related to application of Data Analytics in Agriculture Industry and to develop a Machine Learning model to predict the most suitable crop for the land based on the soil composition, Climatic conditions and assure good yield and profits to the farmers. There are numerous advantages using Data Analytics Techniques for Crop Prediction.

They are as follows:

a) Receiving Useful Data to Help Fight Food Scarcity and Empower Small Farmers.
Ex: - Paddy, cotton in few districts of Telangana (Jowar, wheat)

b) Managing Crop Diseases and Pests.
Ex:- Crop images were analyzed with Advanced ML Algo.

c) To Make Yield Predictions. So that,Farmers can choose the suitable crop to the field.
d) To Cope with Climate Change

The objectives of the study are as follows

1. To develop a Machine Learning Model for crop prediction by using Python Coding and various Statistical Concepts.
2. To predict the best crop to be grown based on Soil Composition, Temperature, Ph- Value and Rainfall

Review of Literature

Pavan Patil et al., [2020] in their study emphasized that decision tree shows poor performance when dataset is having more

variations but naïve bayes provides better result than decision tree for such datasets. The combination classification algorithm like naïve bayes and decision tree classifier are better performing than use of single classifier model. Islam Sarker.M.N , M. Wu, Chanthamith et al., [2019] concluded that "Big data analytics which is often the complex process of examining large and varied data sets, or big data, to uncover information – such as hidden patterns, unknown correlations, market trends and customer preferences which help organizations make informed business decisions".

McCarthy et al., [2019] mentioned from their study that Crop prediction is a technique used in big data analytics which can predict the future outcome based on gathered or historic data. It is one of the very few technologies that have the potential to remold any business positively. It majorly consists of three important techniques also called as the "Big 3" techniques: which are regression analysis, decision trees, and neural networks. Peixeiro [2019] found that "In a time series analytic model, formerly observed values are used to forecast the output. It uses the concept of time series which is a series of data points arranged in time. However, here time is considered as an independent variable with an ambition to forecast the future. It is used in forecasting crops, vegetation price movement and price fluctuation in market.

Team.D & Data Flair [2019] concluded their study saying though, the field of agriculture is less impacted by the recent technological advancements. However, agriculturalists are rapidly moving towards working with modern tools and technologies. One such up to the minute technology is Big Data analytics. Big data has been introduced to almost every other sector even agriculture is not outdistanced from it.Veenadhari.S et al. [2019]This research is to study the various Machine learning approaches for forecasting crop yield based on climatic parameters. This study has reported a correlation between usage of chemicals and crop yield.

Viviliya.B & Vaidhehi.V [2019, In this study the attributes in the dataset are the soil type, groundwater level, rainfall, water availability, temperature of one dataset and the other dataset included the potassium, phosphorus, and nitrogen values, fertilizers, soil pH and organic carbon value. The final recommendation was done using association rules based on the results obtained from the classifiers. The model was trained using 10-cross validation. The testing was done based on different metrics like the Accuracy, ROC Area, Recall, Precision, F-Measure etc.Barun Basnet&Junho Bang[2018], In this study it tells the use of data analytics has shifted agriculture from input-intensive to knowledge-intensive as a large amount of agricultural data can be stored, shared, and analyzed to create information.

Cunha.L.F, et al., [2018], This research gave insights to users with the capability to perform strategy changes, like choosing more robust genetic variation before planting or even changing the crop type, in order to accommodate for extreme climatic variations further ahead in the crop cycle. This research focuses on Scalable Machine Learning System for Pre-Season Agriculture Yield Forecast.Charvat.K [2018]This research tells "An accurate understanding from the huge amount of data can be interpreted in the form of overviews, summary, verifiable models etc. However, most of the interpretation and visualization is organized in an Ad-hoc manner. The models are basically visualized in the form of tables, graphs, spreadsheets etc'.Doshi.Z et al., [2018], opined that the accuracy of a machine learning algorithm may depend on the number of parameters used and to the extent of correctness of the dataset. This research focuses on Agro Consultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms.

Kuma P et al., [2018], from their research found out that a complex and massive collection of data which is hard to process by applying traditional data processing techniques or on-hand database management tools is referred as 'big data". It reveals agriculture is highly depended on population and weather which produces data drastically. Olakunieet al.,[2018] mentioned that the surge in global population is compelling a shift toward smart agriculture practices. This coupled with the diminishing natural resources, limited availability of arable land, increase in unpredictable weather conditions makes food security a major concern for most countries.

Shivi Sharma et al., [2018]In this work, soil and environment features i.e. average temperature, average humidity, total rainfall and production yield are used in predicting two classes namely: good yield and bad yield. Acharjya D P & Kauser Ahmed P[2017], This research has identified the growth of big data was doubled at the end of 2016. Also, big data is characterized by 3Vs to 4Vs which is meant by Volume, Variety, Velocity and Veracity. Through this research, we came to know that Map Reduce, Storm and Apache Spark are emerging tools in big data analytics. Most batch processing tools follow Apache Hadoop which is an open-source software which is reliable and very fast.

Dakshayini Patil, Dr.& M .S Shirdhonkar [2017]. This surveys the technical achievements in the field of Rice crop yield prediction. It also Discussed methodology, comprehensive survey of many proposed approaches to predict rice crop yield and applications.Rising better plans to foresee crop productivity in various climatic conditions can help farmer and different partners in essential basic leadership as far as agronomy and product decision.

Dhivya B H et al., [2017] found that Farmers also grow non-food items like rubber, cotton, jute etc. More than 70% of the household in the rural area depend on agriculture. This domain provides employment to more than 60% of the total population and has a contribution to GDP also (about 17%).

Edwin T [2017] expressed that Our country had undergone several fluctuations in the price of onions last year. The price of the onions increased from Rs. 26 per kg to Rs. 50 per kg in the month of August. So most of the farmers decided to cultivate onion in their fields seeing this huge increase in price so that they could make large profits from their land. It is understood that prediction is done here based on price fluctuations. Francesc X. et al [2017] this research has revealed that "Tools for better demand and yield prediction should be developed". Among the suggested future directions, agricultural robots which are self-operated to identify weeds and remove them, tools with high performances, programming language for big data analysis can

be identified. It is revealed that data privacy is an issue in big data analysis in agriculture because misusing of data.

Jharna Majumdar et al., [2017], The findings of this study are "The estimation of production of crop helps these companies in planning supply chain decision like production scheduling. The industries such as fertilizers, seed, agrochemicals and agricultural machinery plan production and activities like marketing based on the estimates of crop yield".

Mukesh kumar & Prof.Mayura Nagar [2017] through their research stated that Precision agriculture gives not only great challenges but also great opportunities for Computer Scientists specially who are working in the field of data analysis. Most batch processing tools follow Apache Hadoop which is an open-source software which is reliable and very fast. Also, Hadoop framework has been used for taking decision on storage and distribution of the products in an efficient way without wastage.

Rajeswari.S, et al.,[2017] stated that Cloud based big data analytics and the IoT technology performs an important roleinthefeasibilitystudyofsmartagriculture.Smartorprecisionagricultural systems are estimated to play an essential role in improving agriculture activities.

Swarupa Rani [2017], This research has discussed the application of mathematical model like fuzzy logic designs in optimization of the crop yield, artificial neural networks in validation studies, genetic algorithms designs in accessing the fitness of the model applied, decision trees, and support vector machines to study soil, climate conditions and water regimes related to crop growth and pest management in agriculture.

Md. Tahmid Shakooret al., [2017], This paper helped us for selecting various attributes like land capability classification, soil depth, slope, drainage, texture, erosion and permeability. This research focuses on Agricultural Production Output Prediction Using Supervised Machine Learning (SML) Techniques.

Van-Quyet Nguyen et al., [2017], This research has also revealed that "Apache Hadoop and Apache Spark most popular among big data analytics applications". In order to minimize the restrictions regarding data collecting, data storage and data analyzing it is proposed to use deep learning techniques in near future.

Lekhaa T.R [2016],The paper hypothesizes analysis of Explorative Data and considers the design of different types of predictive models. A data set is taken as a sample data set, and different regression techniques are tried to recognize and examine each property. Specific regression methods discussed here are Multiple Linear, Linear, Non-Linear, Polynomial, Ridge regression and Logistic. Using this article, we obtain a comparative study of the different algorithms in data analytics. This helped in determining which algorithm is most appropriate to the proposed system.

Majumdar.J & Ankalaki S [2016], This research helps in Knowing what crops has been grown, and how much area of it had been shown historically, combined with the prices at which it could have been sold at the nearest market-place provides the income-growth profile of the farmer.

Pudumalar S et al., [2016] In their research The crops that were considered in the model for prediction include coriander, pulses, cotton, paddy, sorghum, groundnut, sugarcane, banana and vegetables. Different attributes of the soil were considered in order to predict the crop, which included pH, depth, erosion, permeability, texture, drainage, dater holding and soil color. The technique used was assembling, which combined the power of using two or more different models for better prediction. The assembling technique used was called the Majority Voting Technique.

Kumar.R et al., [2015] In this research the crops were inspected and graded depending on an examination to estimate crop yielding. This categorization is found from different data mining algorithms. This paper provides a perception into various grouping rules, such as K-Nearest Neighbour and Naive Bayes.

Materials and Methods

The current study relied on secondary data which is taken from 'Kaggle' website.

Sampling Technique adopted for this research is convenient sampling method. Sampling size of 2,200 fields Data is collected.

Data Analysis Tools include Python coding, Clustering and Logistic Regression.

Data Analytics Software used for this study is Anaconda Software, Jupyter Notebook and Python Coding Language

Results and discussion

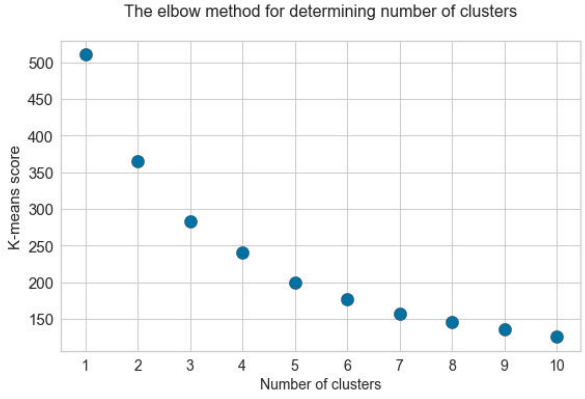
First objective was to develop a Machine Learning Model for crop prediction by using Python Coding and various Statistical Concepts. Here comes the discussion about this.

a) *Machine Learning Models for crop prediction:* - Different Machine learning algorithms are being used in order to make comparisons. The different algorithms used are as follows:

a) K-Nearest Neighbor
b) Decision Tree
c) Naive Bayes

Statistical Concepts

In this project, Clustering and Logistic Regression were the two statistical concepts used. In clustering also, specifically K-Means clustering was used.



The above diagram is called "Elbow Method".The plotting is done between SSE (Sum of squared Errors) and number of clusters. SSE values are represented on X-axis and Number of clusters on Y-axis. In K-means algorithm, the value of 'k' is determined by using this Elbow method.This method is used to represent 'optimal number of clusters'.The Elbow point is '4' as the point after which SSE (or) inertia stars decreasing in alinear fashion.

Second objective was to predict the best crop to be grown based on Soil Composition, Temperature, Ph- Value and Rainfall. Following is the discussion for this. A Machine Learning has been developed so that "The model will suggest the crop to be grown based on the input values given for Nitrogen, Phosphorus, Potassium, Temperature, Humidity, pH-value, Rainfall.

```
In [40]: # Lets do some Real time Predictions
         prediction = model.predict((np.array([[20,
                                                30,
                                                10,
                                                15,
                                                90,
                                                7.5,
                                                100]])))
         print("The Suggested Crop for Given Climatic Condition is :", prediction)

         The Suggested Crop for Given Climatic Condition is : ['orange']
```

Results

1.Data Scientists and Analysts use data analytics techniques, statistical tools to make proper and informed businesses decisions. Data analysis can help companies to better understand their customer's preferences, evaluate the performance of their ad campaigns, to reach customers with their personalized content.

2.In this project, we have developed a Machine Learning Model which can predict the best crop to be grown based on the soil conditions, Climatic conditions. The model has also performed well on 'Test Data'.

3.By applying some Deep Learning concepts, disease of the crop can also be detected by scanning the infected Image of the crop.

4.This type of model can be used to solve the real-time problems and improve the agricultural productivity, revenue of the sector. There is a scope that government (or) a private company can come forward and use this model with real-time data and help farmers.

5.The required soil and climatic conditions for each and every crop is different. It varies from crop-to-crop.for example, Cotton requires high amount of Nitrogen. Grapes and apple require high amount of phosphorous and potassium content in the soil.

6. Crops can also be segregated based on the season that *"They are suitable to be grown"*. For example, moth beans and Mango can be grown in summer only. Rainy season is more comfortable to grow rice, papaya and coconut.

Conclusion

Analytics is playing a major role in many fields such as retail industry, Marketing, Healthcare. It even entered into Agriculture sector. Here, in this project a Machine Learning model is built on the Dataset. The variables in the dataset are Soil conditions such as Nitrogen, Potassium, Phosphorus, pH value and Climatic Conditions such as Temperature, Rainfall. The model is also tested on the 'Test data' and it predicted correctly. So, The model is ready to use and by using this model profits are assured, decrease in farmer's suicide and Doubling the farmer's income can be made true. Here, In this project I have built the Machine Learning Model based on the past environmental conditions and soil conditions collected from 'Kaggle' website. If currently sensed data from the environment, where the crop is grown and comparing against the ideal condition for a good yield. A model can be trained for this purpose and can solve the farmer's real problem.

References:-

1. Basnet.B and Bang.J.(2018).The state-of-the-art of knowledge-intensive agriculture: a review on applied sensing systems and data analytics. Journal of Sensors, vol. 2018, Article ID 3528296, 13 pages.
2. Charvat.K et al.(2018) Advanced Visualisation of Big Data for Agriculture as Part of Databio Development. IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 415–418. DoI: 10.1109/IGARSS.2018.8517556.
3. Cunha, Renato & Silva, Bruno &Netto, Marco. (2018). A Scalable Machine Learning System for Pre-Season Agriculture Yield Forecast. 423-430. 10.1109/eScience.2018.00131.
4. Dakshayini Patil, Dr. Shirdhonkar.M.S.(2017). Rice Crop Yield Prediction using Data Mining Techniques: An Overview. International Journal of Advanced Research in Computer Science and Software Engineering, Volume 7(Issue 5).ISSN: 2277 128X, 2017.
5. Dhivya B H, Manjula R, Siva Bharathi S, Madhumathi R. (2017). A Survey on Crop Yield Prediction based on Agricultural Data. International Journal of Innovative Research in Science, Engineering and Technology. 2017.Volume no: 6(3).
6. Doshi.Z, Nadkarni.S, Agrawal.R and Shah.N. (2018). Agro Consultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms. Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-6. DoI: 10.1109/ICCUBEA.2018.8697349.
7. Acharjya.D.P , KauserAhmed.P. (2017). A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools. International Journal of Advanced Computer Science and Applications (IJACSA).vol. 7(Issue no. 2),page no. 9.
8. Francesc X, Prenafeta-Boldu, Andreas Kamilaris, Andreas Kartakoullis. (2017). A review on the practice of big data analysis in agriculture. Computers and Electronics in Agriculture.vol. 143. page no. 14.
9. Islam Sarker.M.N , M. Wu, Chanthamith.B, Yusufzada.S, D. Li,andZhang.J (2019). Big Data Driven Smart Agriculture: Pathway for Sustainable Development. 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD), 2019, page no.60–65, DoI:10.1109/ICAIBD.2019.8836982.
10. Jharna Majumdar, Sneha Naraseeyappa, Shilpa Ankalaki. (2017). Analysis of agriculture data using data mining techniques: application of big data. Journal of Big data. 2017.
11. Kumar. R, Singh.M.P, Kumar.P and Singh.J.P. (2015). Crop Selection Method to maximize crop yield rate using machine learning technique. International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM). Chennai, 2015, pp.

138-145. DoI: 10.1109/ICSTM.2015.72254.
12. Majumdar J, Ankalaki. S. (2016). Comparison of clustering algorithms using quality metrics with invariant features extracted from plant leaves. International Conference on Computational Science and Engineering.
13. McCarthy.R.V, McCarthy.M.M, Ceccucci.W, and L. Halawi. (2019). Introduction to Predictive Analytics. Applying Predictive Analytics, Cham:Springer International Publishing, pp. 1–25.
14. Md. Tahmid Shakoor, Karishma Rahman, SumaiyaNasrinRayta, Amitabha Chakrabarty. (2017). Agricultural Production Output Prediction Using Supervised Machine Learning Techniques"978-1-5386-3831-6/17/$31.00 ©2017 IEEE.
15. Mukesh kumar, Prof.Mayura Nagar. (2017).Big Data analytics in agriculture and distribution channel. IEEE International Conference on Computing Methodologies and Communication, 2017.
16. Olakunie Elijah, Tharek Abdul Rahman, IgbafeOrikumhia. (2018). An Overview of Internet of Things (IoT) and Data Analytics in Agriculture: Benefits and Challenges. IEEE INTERNET OF THINGS JOURNAL, Volume no. 5.
17. Pavan Patil, Virendra Panpatil, Professor Shrikant Kokate. (2020). Crop Prediction System using Machine Learning Algorithms. International Research Journal of Engineering and Technology (IRJET), Volume: 07(Issue no:02 c), e-ISSN: 2395-0056.
18. Peixeiro.M. (2019).The Complete Guide to Time Series Analysis and Forecasting.Medium.[Online]. Available:https://towardsdatascience.com/the-complete-guide-to-time-seriesanalysis-and-forecasting-70d476bfe775. [Accessed: 06-Jan-2020].
19. Kumar.P, Kumar.A, Panwar.S, Dash.S, Sinha.K , Chaudhary.V.K, Ray.M. (2018).Role of big data in agriculture-A statistical prospective. Indian Agricultural Statistics Research Institute (ICAR), vol. 39, page no. 5.
20. Pudumalar.S, Ramanujam.E, Rajashree.R.H, Kavya.C, Kiruthika.T and Nisha.J. (2016). Crop recommendation system for precision agriculture. Eighth International Conference on Advanced Computing (ICoAC), Chennai, 2017, pp. 32-36. DoI: 10.1109/ICoAC.2017.7951740.
21. Shivi Sharma, Geetanjali Rathee and Hemraj Saini. (2018). Big Data Analytics for Crop Prediction Mode Using Optimization Technique. 5th IEEE International Conference on Parallel, Distributed and Grid Computing(PDGC-2018).
22. Rajeswari.S, Suthendran.K, and Rajakumar.K .(2017). A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics. International Conference on Intelligent Computing and Control (I2C2), 2017, pp. 1–5, DoI:10.1109/I2C2.2017.8321902.
23. Swarupa Rani.(2017).The Impact of Data Analytics in Crop Management based on Weather Conditions. International Journal of Engineering Technology Science and Research. 2017; 4(5):299-308.
24. Team.D. (2019).Data Science in Agriculture - Advancing Together and Benefiting Farmers,"DataFlair, 27-Sep-2019. [Online]. Available: urlhttps://dataflair.training/blogs/data-science-in-agriculture/. [Accessed:02-Dec-2019].
25. T. Edwin. (2017).Onion, tomato price spike: season not the only reason .Nov. 13, 2017. [Online].Available:https://www.thehindubusinessline.com/economy/agri-business/onion-tomato-price-spike-season-not-the only reason / article9957255.ece#.[Accessed Feb. 22, 2020].
26. Van-Quyet Nguyen, Giang-Truong Nguyen, Sinh-Ngoc Nguyen, Jintae Choi, Kyungbaek Kim.(2017).Big Data Analytics and Visualization Techniques: A Case Study from Agriculture Domain. in KISM Fall Conference, 2017.
27. T.R.Lekhaa(2016). Efficient Crop Yield and Pesticide Prediction for ImprovingAgricultural Economy using Data Mining Techniques. International Journal of Modern Trends in Engineering and Science (IJMTES), 2016, Volume 03(Issue 10. D).
28. Veenadhari.S, Dr. Bharat Misra , Dr. Singh.C.D.(2019). Machine learning approach for forecasting crop yield based on climatic parameters. International Conference one Computer Communication and Informatics (ICCCI -2014), Jan. 03 – 05, 2019, Coimbatore, INDIA.
29. Viviliya, B. and Vaidhehi, V.,(2019) "The Design of Hybrid Crop Recommendation System using Machine Learning Algorithms"S. International Journal of Innovative Technology and Exploring Engineering, 2019(2), pp.4305-4311.

Website sources:-

1. https://www.investopedia.com/terms/d/data-analytics.asp
2. https://www.mastersindatascience.org/learning/what-is-data-analytics/
3. https://www.scribd.com/document/508064692/Complete-Project-Work
4. https://www.lotame.com/what-is-data-analytics/
5. https://vitalflux.com/k-means-elbow-point-method-sse-inertia-plot-python/