Review on Weather-Based Crop Yield Prediction Using Big Data Analytics

Prof. Pravin Kharat[1], Mr. Nikhil Agbattanwar[2], Miss. Pranjali Patil[3], Miss. Rohini Tayade[4], Mr. Rohan Shinde[5]
[1, 2, 3, 4, 5] Dept of Computer Engineering
[1, 2, 3, 4, 5] Padm. Dr. V. B. Kolte College of Engineering, Malkapur

Abstract- *Agriculture is the Indian economy's backbone. Big data analytics are becoming more precise and feasible in agricultural research. Current water scarcity, uncontrollable costs owing to demand-supply imbalances, and weather instability need farmers to be prepared with smart farming techniques. Crop yields must be addressed due to unknown climate changes, limited irrigation infrastructure, soil fertility decrease, and conventional agricultural approaches. Weather-based crop yield prediction is a critical area of agricultural research, providing valuable insights to enhance food security and optimize resource management. This paper explores the integration of big data analytics to predict crop yields based on weather patterns. With the growing availability of weather-related data from multiple sources, such as satellite imagery, weather stations, and IoT sensors, advanced machine learning algorithms and data mining techniques can be employed to analyze and predict the impact of weather variables (temperature, rainfall, humidity, etc.) on crop production. The study highlights the use of big data tools like Hadoop, Spark, and various data modeling techniques to process vast amounts of environmental and agricultural data. The predictive models developed from these data provide farmers, policymakers, and stakeholders with actionable insights, allowing them to make informed decisions on irrigation, fertilization, planting schedules, and crop selection. This research demonstrates how the fusion of weather data and big data analytics can significantly improve crop yield forecasting, ultimately contributing to better agricultural planning, sustainability, and economic growth.*

Keywords- Agriculture, Big Data Analysis, Graphical Visualization, K-Means Clustering, Map Reduce, Recommendation System etc

I. INTRODUCTION

Nowadays, many people are unaware of the importance of cultivating crops at the proper time and location. Seasonal climatic conditions are also altered as a result of these cultivating practices, putting key assets like as land, water, and air at risk, resulting in food insecurity Climate change has had a detrimental effect on the performance of the majority of agricultural crops in India during the previous two decades. Predicting crop yields in advance of harvest enables policymakers and farmers to take proper marketing and storage strategies

The agricultural sector plays a crucial role in global food security, economic development, and environmental sustainability. However, crop yield prediction remains a complex and challenging task due to the variability of weather patterns, soil conditions, pests, and other environmental factors. Weather, in particular, has a profound influence on crop production, making it essential for farmers and stakeholders to accurately predict crop yields to mitigate risks associated with climate change and natural disasters. Traditional methods of yield prediction, which often rely on historical data or simplified models, have limitations in terms of accuracy and scalability, especially in the face of rapidly changing climatic conditions.

The advent of big data analytics has revolutionized various industries, and agriculture is no exception. With the increased availability of large-scale data from diverse sources such as satellite imagery, weather stations, IoT sensors, and climate models there is now an unprecedented opportunity to harness these data streams for more precise crop yield forecasting. By integrating weather data with crop-specific characteristics, soil conditions, and other variables, big data analytics enables the development of more robust predictive models that can accurately forecast crop yields based on real-time and historical weather patterns. In this context, the application of advanced machine learning techniques, such as deep learning, decision trees, and regression analysis, can help identify complex relationships between weather variables (e.g., temperature, precipitation, humidity) and crop performance. Additionally, big data frameworks like Hadoop and Apache Spark provide the computational power needed to process and analyze vast datasets efficiently, leading to faster and more reliable predictions.

This paper explores the use of big data analytics in weather-based crop yield prediction, examining how advanced data-driven techniques can provide farmers and agricultural stakeholders with actionable insights. The integration of weather data with machine learning models offers the potential to enhance decision-making processes, optimize agricultural practices, and increase the resilience of crop production systems in the face of climate variability. Ultimately, leveraging big data analytics for crop yield prediction has the potential to drive agricultural innovation, improve resource management, and ensure sustainable food production in an increasingly unpredictable climate.

The advent of big data analytics presents a transformative opportunity to address these challenges. Big data refers to the vast volume of data that can now be analyzed to uncover patterns, trends, and insights. By leveraging this technology, farmers can analyze weather data, soil conditions, and crop yields more effectively, allowing for better planning and management of agricultural activities. This project, "Weather-Based Crop Prediction Using Big Data Analytics," aims to use big data tools and machine learning algorithms to predict crop yields based on weather forecasts. By doing so, it enables farmers to make informed decisions about when to plant, grow, and harvest crops, thereby improving productivity and minimizing the risks associated with climate unpredictability.

The system proposed in this project will analyze weather variables like temperature, rainfall, humidity, and seasonal changes in combination with historical crop data to predict the best times for planting and harvesting specific crops. The goal is to provide accurate, localized predictions that will help farmers optimize crop yields and reduce losses due to adverse weather conditions. By incorporating big data analytics into agricultural practices, the project aims to improve resource management, reduce waste, and foster more sustainable farming practices. Ultimately, the project will contribute to more efficient, data-driven farming, which is essential for India’s food security and the continued development of the agricultural sector.

II. LITERATURE REVIEW

The exploration of crop prediction using weather data has seen substantial advancements, driven by innovative methodologies and technologies.

- **M. Chen et al. (2019):** "Crop Prediction Using Weather Data: An Overview" serves as a foundational study in this domain. Chen and colleagues provide a comprehensive overview of various predictive models and their reliance on weather data. The authors outline the integration of statistical methods with climatic data to forecast crop yields accurately. Their work highlights the utility of historical weather patterns in enhancing prediction accuracy. However, the study also notes limitations related to the variability of weather conditions and the challenges in extrapolating models across different geographic regions. These limitations underscore the necessity for adaptive models that can account for regional climatic differences.
- **A. Gupta and R. Patel et al. (2017):** In "Application of Machine Learning in Predicting Crop Yields Based on Weather Data the focus shifts to the implementation of machine learning algorithms to refine yield predictions. Gupta and Patel leverage techniques such as neural networks and support vector machines to process complex weather datasets. Their approach demonstrates a significant improvement in prediction accuracy compared to traditional statistical methods. The authors emphasize that machine learning models can capture non-linear relationships between weather variables and crop yields, thus offering more precise forecasts. Nonetheless, the paper acknowledges the challenge of model interpretability and the need for extensive datasets to train robust models effectively.
- **L. Brown and T. Anderson et.al. (2020):** "Weather Data Integration in Agricultural Forecasting: Challenges and Opportunities" examines the integration of weather data into agricultural forecasting systems. Brown and Anderson explore various approaches to incorporating real-time weather information into predictive models. They discuss the advantages of using high-resolution weather data to improve forecast accuracy. However, the paper highlights several challenges, including data integration issues and the need for real-time data processing capabilities. The authors suggest that advancements in data assimilation technologies could address these challenges and enhance the effectiveness of weather-driven forecasting systems.
- **J. Wilson and S. Lee et. al. (2020):** "Enhancing Crop Yield Predictions with Advanced Weather Models" presents a detailed analysis of advanced weather models and their impact on crop yield predictions. Wilson and Lee investigate the use of sophisticated weather forecasting models, such as ensemble forecasts, to improve yield predictions. Their findings indicate that incorporating advanced weather models can significantly enhance prediction accuracy by providing more precise weather forecasts. The paper also discusses the importance of model calibration and validation to ensure the reliability of predictions. Despite the improvements offered by advanced models, the authors note that high computational requirements and the complexity of model integration remain key challenges.
- **K. Martinez and D. Rodriguez et. al. (2018):** The paper "Utilizing Weather Data for Sustainable Agricultural Practices" explores the application of weather data in

promoting sustainable agricultural practices. Martinez and Rodriguez focus on how weather data can be used to optimize irrigation and fertilization practices, leading to more sustainable crop management. Their research highlights the benefits of weather-informed decision-making in reducing resource waste and improving crop efficiency. The authors also address the limitations of current weather data applications, such as the need for localized data and the integration of weather information with other agricultural practices. Their study suggests that continued advancements in weather data technologies could further support sustainable agriculture efforts.

- **S. Kumar and P. Sharma et. al. (2020):** "Development of an IoT-based System for Crop Monitoring and Prediction" presents an innovative approach to crop monitoring using Internet of Things (IoT) technology. Kumar and Sharma's research highlights the integration of IoT devices to collect real-time weather and soil data, which are then used to predict crop conditions and yield. The authors discuss how IoT-based systems offer continuous monitoring and data collection, which improves the accuracy of crop predictions. However, they also address challenges such as data security and the need for robust data management systems. The study underscores the potential of IoT technology in enhancing agricultural practices but also calls for addressing these technical challenges to fully realize its benefits.
- **R. Singh and A. Patel et. al. (2020):** In "Forecasting Crop Yields with Remote Sensing and Weather Data" by the authors explore the use of remote sensing technologies in conjunction with weather data for crop yield forecasting. Singh and Patel's study demonstrates that remote sensing provides valuable spatial and temporal data that can complement weather data to enhance yield predictions. The paper highlights the advantages of integrating satellite imagery with weather information to create detailed crop models. Despite these benefits, the authors acknowledge limitations related to the resolution of remote sensing data and the high costs associated with satellite technology. Their research suggests that combining remote sensing with weather data could offer more comprehensive insights into crop health and yield potential.
- **J. Lee and M. Thompson et. al. (2014):** "Advancements in Weather-Based Crop Prediction Models: A Review" by offers a thorough review of recent advancements in weather-based crop prediction models. Lee and Thompson examine various modelling approaches, including statistical, machine learning, and hybrid models, and their applications in crop prediction. The paper provides a detailed analysis of the strengths and weaknesses of each modelling approach. The authors emphasize the growing importance of integrating diverse data sources and improving model accuracy. While advancements in weather prediction models have led to better forecasts, the paper also notes ongoing challenges, such as model over fitting and the need for continuous model updates to accommodate changing climate conditions.
- **P. Garcia and L. Robinson et. al. (2021):** The paper "Optimizing Crop Yield Predictions with Ensemble Learning Techniques" by explores the application of ensemble learning methods to improve crop yield predictions. Garcia and Robinson investigate various ensemble techniques, such as bagging and boosting, to enhance prediction performance. Their findings indicate that ensemble learning methods can significantly reduce prediction errors by combining multiple models to achieve more reliable forecasts. The authors highlight the benefits of ensemble approaches in capturing diverse data patterns and improving model robustness. However, they also point out the computational complexity associated with ensemble methods and the need for efficient algorithms to handle large datasets.
- **H. Kim and N. Evans et. al. (2020):** "Integrating Weather Data with Precision Agriculture: Benefits and Challenges" examines the integration of weather data into precision agriculture practices. Kim and Evans discuss how precision agriculture techniques, combined with weather data, can optimize resource use and improve crop management. Their study highlights the benefits of using weather-informed precision agriculture strategies, such as targeted irrigation and fertilization. However, the authors also address challenges related to data integration, technology adoption, and the need for tailored solutions for different crops and regions. Their research suggests that ongoing advancements in precision agriculture and weather data integration could enhance overall agricultural efficiency and sustainability.

III. METHODOLOGY

The major objective would be to analyze the data using MapReduce and write a Python recommender algorithm to extract output depending on seasonal conditions and geography, then perform IKM clustering and determine the mean production per area provided by a group of crops in a certain location. We picked our system's temperature, rainfall, wind speed, humidity, soil type, and seed type as decision factors based on past work in other journals. Python will be used to collect and preprocess the raw data. The preprocessed data is then utilized for Hadoop's MapReduce framework to process the data. MapReduce is a programming technique that uses a parallel, distributed method to handle massive volumes of data. Following that, we suggest combining all of the MapReduce datasets for the various parameters into a single final/super dataset and developing a recommendation system. The month, area, and state may all be entered by the user. After that, we'll establish numerous agricultural seasons. After that, we assign the month supplied by the user to the proper season/seasons. For instance, if November is specified as the month, we may assign Rabi/Winter crops. We then analyze the data to determine the three crops that yield the most during that season and the crops that yield the most throughout the year in that state and location. These would be stored in two separate data frames, one for each season and another for the whole year. Additionally, we output the temperature, rainfall, wind speed, and humidity values previously given by the crop. Additionally, we list the seed variety for each soil type for crops grown in different areas since seed availability varies by region. The recommendation function's output will be shown on a graphical user interface (website) created using Flask and Python. The user may submit pertinent data and get a response from the system. Following that, we will use an IKM clustering technique. First, we will construct an elbow graph to establish the cluster count and the required K value.

This will be accomplished via the usage of the Scikit-learn package. Following that, the fit predicts approach will be utilized to get cluster values. This will be accomplished via an array, where numbers starting with 0 indicate the values of a single cluster. The clusters will then be plotted using the scatter method provided by the Matplotlib toolkit. Each cluster centric will be exhibited to reflect the average value of the cluster in which each crop is plotted, and a separate color will represent each cluster.

3.1 DATA COLLECTION:

The production of crops. All of the states and their districts were included in the CSV file. It offers information on 125 crops, their productivity, and the region in which they were planted from 2000 to 2014 for six seasons: Kharif, Rabi, Summer, Winter, Autumn, and the whole year. The daily average temperature of the cities from 1995 to 2020 is included—India's rainfall from 1901 to 2015. From 1901 to 2015, CSV contains the average monthly rainfall of several subdivisions. The soil, seed, humidity, and wind speed data were manually gathered from several sources such as agricoop, the Department of Agriculture's weather online, and climateorg.

3.2 Preprocessing

This stage included merging and cleaning the collected datasets. We uploaded our datasets to the Colab notebook and utilized panda's data frames to eliminate superfluous columns while retaining the important ones. We used the NumPy and SciPy libraries to do our calculations. A few index columns were added for future calculations. Interpolation, as shown, was used statistically to get the estimated value for the dataset's missing value.

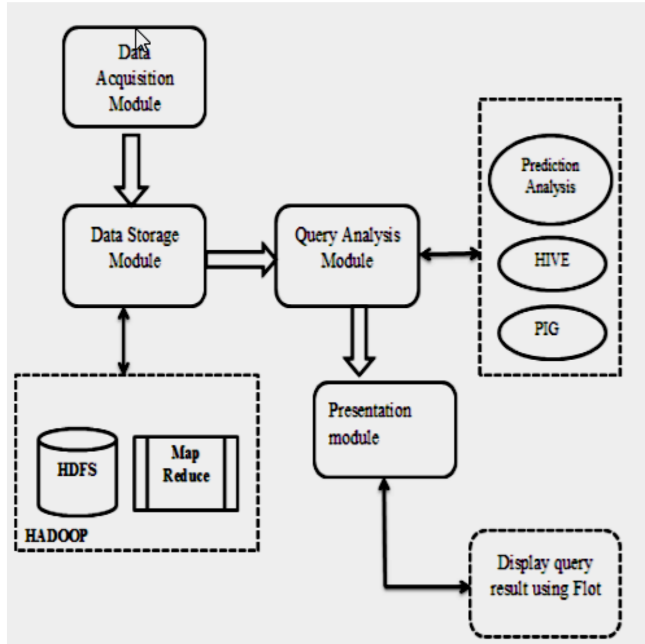

Figure 3.1 System Architecture

3.3 Data Acquisition Module

This module collects sensor data, weather forecasts, social media, and market trends.

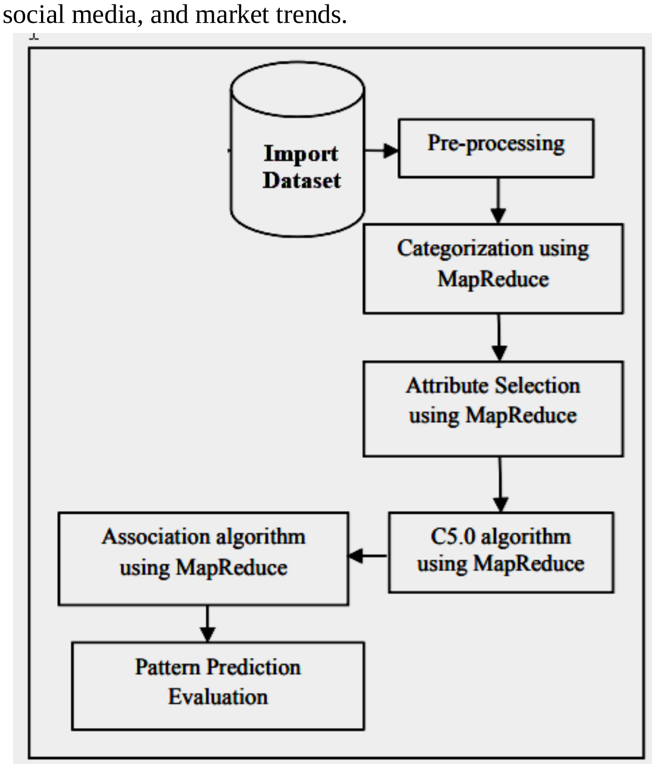

Figure 3.2 Level -0 Data Flow Diagram

These meteorological data can be issued manually or acquired using meteorological data acquisition equipment; small received data are initially stored in an Oracle database; once a sufficient number of small data has been gathered, small data are transferred to the storage module; transferred data are automatically deleted.

3.4 Data storage module:

It is responsible for storing metadata and data sets in duplicate, acting as a backup facility. HDFS is a data-agnostic storage container. Small data collected in the data collection module will be saved in the storage module regularly after accumulating to a certain amount

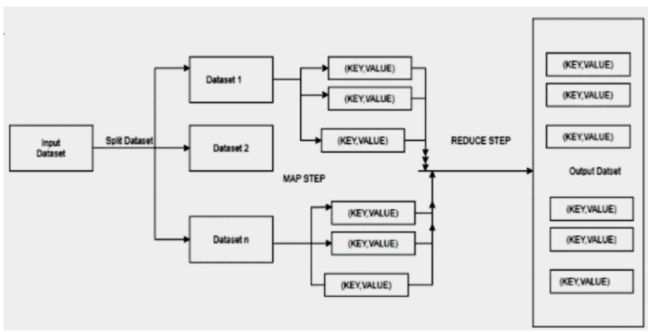

Figure 3.3 Level -1 Data FlowDiagram

3.5 Query Analysis:

This module's processing phase consists of two parts: data reading/analysis and forecast outcomes. Hive is primarily responsible for data reading. Hive is a Hadoop based data warehouse architecture. It was designed to allow analysts with good SQL expertise to operate on the massive amounts of data contained in HDFS. Hive is a workstation application that turns SQL queries into a sequence of MapReduce tasks executing a Hadoop cluster. MapReduce is an execution engine designed for big data processing that greatly reduces query response time. The second section includes a prediction function that uses the IKM cluster technique to generate forecast data. We utilize apache mahout in this case. It is an open-source, scalable machine learning library. Mahout is a fast technique to create unsupervised machine learning algorithms. The data from the last several years are utilized to forecast the future.

3.6 Presentation:

This module will show the findings produced from the query analysis module. The ability to represent complicated data using charts and graphs is critical for data analysis. Figure 3 depicts the suggested system process in the first phase, whether datasets acquired from different sources are further preprocessed to enter the prediction algorithm effectively. After cleaning the data, put it into HDFS and run a Hive query to analyze it. We can also execute a Py script to analyze the data, and the output is sent to flotend to generate the graph for examination. We utilize the logistic regression approach for prediction, which can be implemented using the Apache Mahout ML package.

3.7 An algorithm in Hadoop Theme:

We proposed a technique in the study that assists us in predicting crop production by recommending the optimal crop. It also focuses on the soil type, which helps predict which crop has to be planted in the field to increase productivity.

Model: Soil types are important in terms of agricultural productivity. The soil information may be acquired by considering the previous year's weather. It assists us in predicting which crops are suited for certain climatic circumstances. Crop quality may also be improved by using weather and disease data sets. A substantial value of the data set may be used to forecast the crop. An Agro algorithm can handle a vast number of data sets. It is implementable on the Hadoop platform and uses the Hadoop framework to manage massive data sets.

3.8 Observation:

According to the research discussed above, varied soil samples obtained from various sites may be assessed utilizing portable NPK sensors with a short response time. The NPK sensor, on the other hand, will detect just nitrogen, potassium, and phosphate in the soil. Numerous authors' work has been combined in one area to aid researchers in comprehending the current state of agriculture. The strategy allows us to make comparisons between agricultural approaches and crop compositions. The Agro algorithm supports farmers in making crop selection decisions. Because climatic conditions are subject to change, trustworthy conclusions are impossible. It assists the farmer in planting the proper crop accurately and quickly. The forecast is based on the atmosphere, incorrect since climatic circumstances might change.

3.9 Improved K-Means Clustering for Crop Yield Prediction

Improved K- Means clustering is one of the most effective methods for predicting crop productivity. The key benefits of employing this approach are that it may have numerous parameters and, if the K value is modest, it may be computationally closer to hierarchical clustering. IKM may hierarchical clustering, especially when the clusters are rotund. IKM characteristics are always K clusters. Each cluster contains at least one item. The clusters are nonhierarchical and expand beyond the boundaries. Because nearness does not entail the 'center' of clusters, each cluster member is closer to its cluster than another. The IKM cluster technique partitions the dataset into K clusters. The data points are randomly assigned to the clusters, resulting in nearly the same data points. Put it down if the point is closer to its cluster for each data point. If the data point is close to a cluster, it should be moved to the closest cluster. Repeat the procedures until the sum of the data points surpasses the sum, resulting in no point influencing one cluster over another. At this point, the clusters are consistent, and the processes have halted.

3.10 Improved K-means Clustering Algorithm

The IKM clustering algorithm is a partitioning technique that varies in determining the number of clusters and choosing the first cluster centroid. The following are IKM algorithms' processing of multiple clusters and the initial cluster centroid. The IKM technique is designed to identify initial cluster centers and many clusters. The approach begins by selecting the attribute with the least value in the X dataset. After deducting the minimum

IV. PROPOSED SYSTEM

After studying the previous work done, the main aim would be to process the data using MapReduce and frame a recommender algorithm in Python to extract output according to the seasonal conditions and region followed by executing k-means clustering and finding the mean produce per area a group of crops will give in a particular region.

Keeping in mind the previous work done in other papers, we have taken temperature, rainfall, wind speed, humidity, soil type and seed type as the deciding parameters of our system. Firstly, the raw data will be collected and pre-processed in a Python environment. Then this pre-processed data is used as an input for the MapReduce framework of Hadoop to process the data. MapReduce is a programming model for processing large amounts of data with a parallel, distributed algorithm. This model is implemented on the collected datasets for faster processing. In this work, each dataset will be processed differently. In the MapReduce model, the dataset will be divided into key and column pairs as shown in Fig. where the different parameters will be individually taken to perform a MapReduce. The year and region will be stored in the key, and the respective parameter for all the months will be taken as the value for the Map Function. In the Reduce function, these parameters will be calculated and assigned to crop seasons. For the Map function of the crop dataset, the region, year, season and crop will be assigned as the key and the produce and area will be taken as the value. Then the Reduce function will calculate the produce per area for each row of data where the region, year, season and crop will form the key and produce per area will become the value. Next, we propose to combine all the map reduced datasets of the different parameters to form one final/super dataset and make a recommendation algorithm. Three parameters can be taken as user input: the month, region and state. Next, we will initialise the different agricultural seasons. Then depending upon the user input of the month, we assign it the respective season/s. For example, if the user input is November, then we could assign it Rabi/ Winter crops. Next, we parse through the data and select the three crops that give the best yield in that particular season and also the crops that give the best yield throughout the whole year in that particular state and region. These would be in two different data frames, one for the particular season and one for the whole year. Along with this, we output the temperature, rainfall, wind speed and humidity in which the crop had previously given the same output. We also mention the seed type to be used for each kind of soil for the crops in different regions as the availability of seed and soil type varies from region to region. The output of this recommendation function will be displayed on a Graphical user interface (website) designed on Flask using Python, where the user could input the required data and get the output from the system. Next, we will use a K-means clustering model. Firstly, we will make an elbow graph to calculate the number of clusters/optimal value of K that is required. We will be utilizing the Scikit-learn library for this purpose. Then we will be using the fit predict method to get the values of clusters. This will be done in the form of an array where numbers starting from 0 will represent the values of one cluster. Then the clusters will be plotted using the scatter method of the Matplotlib library. Each cluster centroid will be shown which would represent the average value of a cluster about which each crop would be plotted, and every cluster would be represented by a different colour. To study the relationship between the produce per area, crops and the respective parameters, we would create several 3D graphs and scatter plots. The produce per area could be taken as the Y-axis and crops as the X-axis and we could change the Z-axis according to the parameters taken (here temperature, rainfall, humidity and wind speed). We will also study the relationship between soil and seed type using 2D bar graphs and scatter plots using the Matplotlib, seaborn and mpl_toolkits in Python by taking soil as the X-axis and the number of crops it supports growth for in Y-axis. Also, a graph could be made by taking Crops as the X-axis and Produce per area on the Y-axis to study which crop gives the maximum produce per area in the particular region.

VI. APPLICATION AND BENEFITS

The integration of big data analytics into weather-based crop prediction systems has the potential to transform agricultural practices and provide numerous applications and benefits for farmers, stakeholders, and the agricultural ecosystem as a whole. This chapter explores the key applications and the various advantages that arise from the adoption of this technology in precision agriculture.

6.1 Applications of the Weather-Based Crop Prediction System

The weather-based crop prediction system serves as a cornerstone for precision agriculture, where real-time data analysis enables farmers to make informed decisions regarding crop management. By accurately predicting optimal planting and harvesting periods based on weather conditions, the system ensures that crops are grown under the most favorable environmental conditions. This application helps reduce the risks associated with crop failure due to climate variations and leads to better resource utilization. With climate change causing increasingly unpredictable weather patterns, farmers face growing challenges in planning for adverse conditions such as droughts, floods, and storms. The weather-based crop prediction system acts as an early warning mechanism, alerting farmers to potential climate risks. By offering timely predictions, the system enables farmers to adjust their practices, such as altering planting schedules or using climate-resilient crops, reducing the likelihood of crop loss due to extreme weather. Water management is a critical issue in agriculture, particularly in regions facing water scarcity. The system provides insights into the optimal times for irrigation based on weather forecasts, soil moisture, and crop growth stages. By avoiding over-irrigation or under-irrigation, farmers can conserve water, reduce energy costs, and improve water use efficiency, leading to more sustainable agricultural practices. Accurate weather data is essential for predicting pest and disease outbreaks. Temperature, humidity, and precipitation influence the development of pests and plant diseases.

6.2 Benefits of the Weather-Based Crop Prediction System

By optimizing the timing of planting, harvesting, and irrigation, the system helps farmers maximize crop productivity. This leads to higher yields and, in turn, greater profitability. With accurate predictions of crop performance, farmers can plan their activities efficiently, reducing waste and increasing the return on investment (ROI). Efficient resource management is crucial for reducing costs and ensuring sustainability. The weather-based crop prediction system optimizes the use of resources such as water, fertilizers, and pesticides by predicting the ideal application times. This not only reduces input costs for farmers but also minimizes environmental impact, fostering sustainable farming practices. Climate-induced crop losses are a significant concern for farmers, especially in regions with erratic weather patterns. The system's ability to predict weather anomalies and crop growth conditions allows farmers to take proactive steps to mitigate crop losses. Early warnings of droughts, frosts, or excessive rainfall enable farmers to adjust planting schedules, choose climate-resilient crops, or take protective measures, thus reducing overall crop failure rates. Farmers in regions prone to extreme weather events can benefit from the risk mitigation capabilities of the system. By providing early warnings of extreme weather, such as storms or floods, the system helps farmers implement preventive measures to protect their crops. Additionally, by recommending climate-resilient crop varieties, the system assists in building long-term resilience against climate change, ensuring that farmers can continue to grow crops in challenging conditions. Sustainability is at the core of modern agricultural practices.

VII. CONCLUSIONS

We demonstrated the WCP technique, incorporating a crop recommendation system, MapReduce, and IKM clustering to provide computationally efficient results. The model considers various crops and their products according to location, soil type, and seed types depending on the varieties grown in a particular area. The mean production of a set of crops may be calculated using IKM clustering visualization graphs. The recommender function and IKM Clustering algorithms are accessible on https://github.com/ oorjagarg/WB-CPI. The link between elements (such as the optimal temperature, seasonal rainfall, wind speed, humidity, soil availability, and seed types required), crop, and the area has also been researched and shown using 2D and 3D graphs. The strategy is scalable, and in the same manner that the methodology is discussed, it may be used to determine the proposed crops for various states

REFERENCES

[1] Muangprathub, J., Boonnam, N., Kajornkasirata, S., Lekbangpong, N.,Wanichsombat, A. and Nillaor, P. (2019) IoT and Agriculture DataAnalysis for Smart Farm. Computers and Electronics in Agriculture,156, 467-474. https://doi.org/10.1016/j.compag.2018.12.011

[2] Thorat, A., Kumari, S. and Valakunde, N.D. (2017) An IoT BasedSmart Solution for Leaf Disease Detection. 2017 International Conference on Big Data, IoT and Data Science (BID), Pune, 20-22 December 2017, 193-198. https://doi.org/10.1109/BID.2017.8336597
[3] Katarya, R., Raturi, A., Mehndiratta, A. and Thapper, A. (2020) Impactof Machine Learning Techniques in Precision Agriculture. 2020 3rdInternational Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE),Jaipur, 7-8 February 2020, 1-6 https://doi.org/10.1109/BID.2017.8336597
[4] Patil, B., Maske, R., Nakhate, R., Nikam, R. and Javheri, P.S. (2020) Soil Fertility Detection and Plant Prediction Using IoT and MachineLearning Approach. International Journal of Innovative Research inScience, Engineering and Technology, 9, 3329-3334.
[5] Girish, L., Gangadhar, S., Bharath, TR, Balaji, KS and Abhishek, K.T.(2018) Crop Yield and Rainfall Prediction in Tumakuru District UsingMachine Learning. National Conference on Technology for RuralDevelopment, 61-65.
[6] Nischitha, K. and Dhanush, V. (2020) Crop Prediction Using MachineLearning Approaches. International Journal of Engineering Researchand Technology, 9, 23-26.
[7] Ms. Fathima, Ms. Sowmya K, Ms. Sunita Barker, Dr. Sanjeev Kulkarni(2020), "Analysis of Crop yield Prediction using Data MiningTechnique" International Research Journal of Engineering andTechnology (IRJET), Volume: 07 Issue: 05.
[8] A. Gonzalez-Sanchez, J. Frausto-Solis, and W. Ojeda-Bustamante(2014): "Predictive ability of machine learning methods for massive crop yield prediction," Spanish Journal of Agricultural Research, vol. 12, no. 2, pp. 313–328.
[9] D. K, R. M, S. V, P. N, and I. A. Jayaraj(2021)," Meta-Learning BasedAdaptive Crop Price Prediction for Agriculture Application," in 2021IEEE, 5th International Conference on Electronics, Communication,and Aerospace Technology (ICECA), pp. 396-402, DOI:10.1109/ICECA52323.2021.9675891
[10] S. Jain and D. Ramesh, "Machine Learning convergence for weather based crop selection," 2020 IEEE International Students' Conferenceon Electrical,Electronics and Computer Science (SCEECS), Bhopal,India, 2020, pp. 1-6, doi: 10.1109/SCEECS48394.2020.75.
[11] R. Gupta et al., "WB-CPI: Weather Based Crop Prediction in IndiaUsing Big Data Analytics," in IEEE Access, vol. 9, pp. 137869-137885, 2021, doi: 10.1109/ACCESS.2021.3117247.