www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

A Big Data-Driven Weather-Based Crop Prediction Model Using Machine Learning

[1]M. RADHIKA, [2]CH. NEVEEN, [3] M.B.N. KISHORE

[1,2] Assoc. Prof., Department of CSE, MAM Women's Engineering College, Narasaraopet, Palnadu, A.P., India.

[3] Asst. Prof., Department of CSE, MAM Women's Engineering College, Narasaraopet, Palnadu, A.P., India.

ABSTRACT

Agriculture plays a vital role in sustaining economies and ensuring food security, particularly in countries like India where a majority of the population depends on farming. However, unpredictable weather patterns, lack of real-time data, and conventional farming methods make crop selection and planning highly challenging. This paper proposes a big data-driven, weather-based crop prediction model powered by advanced machine learning algorithms to address these issues. The system integrates large-scale weather and soil datasets, which are preprocessed using standardization and label encoding techniques. A dual-model architecture is adopted—one model predicts the weather category based on environmental parameters such as temperature, humidity, wind speed, and atmospheric pressure using Conv1D and LSTM layers; the second model forecasts the most suitable crop based on predicted weather, soil pH, and rainfall data. The models are trained and evaluated using real-time datasets, achieving high accuracy through sequential deep learning methods. The backend is built on Django, providing a seamless interface for user registration, model training, and prediction. This integrated approach not only enhances the accuracy of crop prediction but also enables data-driven decision-making for farmers, improving yield and optimizing resource utilization. The proposed system holds immense potential for deployment in smart agriculture frameworks to support sustainable farming practices.

Keywords: Big Data Analytics, Crop Prediction, Weather Forecasting, Machine Learning, LSTM, Conv1D, Django Web Framework, Data Preprocessing, Label Encoding

I. INTRODUCTION

Agriculture is a cornerstone of the global economy and a fundamental aspect of human survival, particularly in countries like India, where a majority of the population relies on farming for livelihood. However, modern agricultural practices face increasing challenges due to unpredictable weather conditions, climate change, inefficient resource usage, and lack of real-time decision support for farmers. One of the most pressing problems is crop selection — choosing the right crop based on dynamic factors such as temperature, humidity, rainfall, pH levels, and weather patterns. In most cases, farmers rely on traditional knowledge or generic seasonal information, which often leads to sub-optimal crop yields or total crop failure.

To address these limitations, this study proposes a data-driven approach that integrates Big Data Analytics and Machine Learning to accurately forecast weather conditions and predict the most suitable crop for cultivation under given environmental parameters. The project leverages large volumes of weather and agricultural datasets, processes them using scalable techniques, and feeds them into deep learning models to extract predictive insights. A Django-based web interface enables seamless interaction between users and the model, ensuring usability and accessibility for end-users such as farmers and agricultural planners.

The system is designed in two phases. The first phase involves **Weather Prediction** using historical weather data features such as temperature, dew point, humidity, wind speed, visibility, and atmospheric pressure. The weather prediction model is constructed using a hybrid neural network architecture comprising **Conv1D** and **LSTM** layers. This structure captures both spatial and temporal dependencies in the input data, offering high accuracy in forecasting weather conditions. The model is trained using preprocessed data, scaled with a **StandardScaler**, and encoded with a **LabelEncoder**, which are saved and reused for prediction during runtime.

In the second phase, **Crop Prediction** is carried out based on environmental factors including temperature, humidity, soil pH, and rainfall. The model is again built using a combination of **Conv1D** and **LSTM** layers to effectively extract features from time-series and multi-dimensional data. The training process involves preparing the dataset by transforming labels and standardizing input features. Once trained, the model can recommend the most suitable crop for a given environmental setting, improving both productivity and sustainability.

What distinguishes this project is its **end-to-end pipeline** — from data collection and preprocessing, to training, evaluation, and deployment via a real-time web interface. The integration of two separate yet interlinked models—one for predicting future weather and another for crop recommendation—represents a practical use-case of **machine learning in smart agriculture**. The use of **joblib** for model serialization, and **TensorFlow Keras** for model development, ensures scalability and ease of maintenance.

In summary, this system aims to empower farmers with precise, data-backed recommendations that optimize yield, reduce dependency on trial-and-error farming, and mitigate the adverse effects of climate unpredictability. By leveraging the power of Big Data and Machine Learning, the proposed solution lays the groundwork for more sustainable and intelligent agricultural practices in the face of global challenges.

II. LITEARTURE SURVEY

In recent years, the integration of **Big Data Analytics** and **Machine Learning (ML)** in agriculture has gained considerable momentum. As agriculture becomes increasingly data-driven, there is a growing need for intelligent systems that can process large volumes of environmental data to assist farmers in making more informed decisions. One of the most promising applications is **weather-based crop prediction**, which helps optimize crop selection and yields.

Jiang et al. [1] introduced HISTIF, a spatiotemporal fusion technique for crop monitoring at high resolution. This work emphasizes the value of satellite imagery and high-frequency monitoring, laying the groundwork for precision agriculture through remote sensing. In a broader review, Rashid et al. [2] analyzed various ML models applied to crop yield prediction, highlighting their potential in predicting palm oil yields. They pointed out the importance of domain-specific feature selection and preprocessing in improving prediction accuracy.

Elavarasan and Vincent [3] proposed a **deep reinforcement learning-based model** for crop yield prediction, offering adaptive learning mechanisms for dynamic agricultural environments. Similarly, Qiao et al. [4] employed 3D Convolutional Neural Networks (3D-CNN) and Gaussian Processes to extract hierarchical features for crop yield prediction, demonstrating that deep learning models can capture spatial patterns effectively.

Ashok and Suresh Varma [5] explored crop prediction using ensemble machine learning algorithms, integrating multiple classifiers to improve robustness. In close alignment with our approach, Gupta et al. [6] presented WB-CPI (Weather-Based Crop Prediction in India), a Big Data-driven model combining meteorological and agronomic data for accurate crop forecasting. Their findings supported the importance of preprocessing, weather monitoring, and ML model selection — all central aspects of our implemented system.

Bendre et al. [7] and Majumdar et al. [8] emphasized the role of **Big Data in weather forecasting and agricultural decision-making**, reinforcing the necessity of scalable frameworks such as **MapReduce**, which indirectly supports the preprocessing pipeline of our system. Meanwhile, studies like those in [9] and [10] examined ML-based precision agriculture frameworks using environmental sensors and weather data to predict crop suitability.

Bhat and Huang [11] conducted a survey outlining challenges in integrating **AI and Big Data for precision agriculture**, particularly in terms of data heterogeneity and model interpretability. Their work echoes the importance of feature standardization and dimensionality reduction, which are integral to our system via the use of **StandardScaler** and **LabelEncoder**.

In addition, Rao [12] discussed the strategic need for climate-smart agriculture in India, advocating for predictive models that accommodate varying climate patterns — precisely what our weather-based crop model attempts to address. Cravero and Sepúlveda [13] reviewed real-world agricultural applications of ML in Big Data, identifying the potential of LSTM models in time-series forecasting, which is implemented in both weather and crop models in our codebase.

Ngo et al. [14] and Sahu et al. [15] emphasized the significance of designing data architectures like **data warehouses** and **streaming pipelines** for large agricultural datasets. This supports our use of structured datasets and model serialization using **joblib** for consistent and scalable inference.

Ranjani et al. [16] and Sagana et al. [17] implemented ML-based models to analyze weather data for identifying suitable crops, similar to our two-stage system: first predicting weather conditions, then recommending crops accordingly. The work of Kumar et al. [18] also introduced a **crop selection model** using ML techniques to maximize yield, reinforcing the practical need for such tools in real-time agricultural planning.

Finally, studies by Gupta et al. [19], Bodapati et al. [20], and Vineela et al. [21] investigated hybrid and ensemble models, user-centered interfaces, and crop yield classification mechanisms — many of which are reflected in our web-integrated, dual-model prediction architecture.

Conclusion of Literature Survey

the reviewed literature demonstrates a convergence towards **data-driven agriculture**, where real-time weather forecasting and ML-based crop prediction are pivotal to solving modern agricultural challenges. Our proposed system builds upon these foundations by combining **Conv1D-LSTM architectures**, **Label Encoding**, **Standard Scaling**, and **Django web deployment** to provide farmers and stakeholders with a practical, efficient, and accurate decision support system for crop recommendation based on prevailing weather conditions.

III. METHODOLOGY

The methodology adopted for this project involves a systematic and data-driven approach to accurately predict suitable crops based on prevailing weather conditions. This includes data acquisition, preprocessing, model training using machine learning algorithms, and integration of the models into a web-based interface to facilitate real-time crop prediction. The workflow can be categorized into two key predictive tasks: **weather classification** and **crop recommendation**, both supported by big data analytics techniques.

1. Data Acquisition

Two datasets are used in this study:

Weather Dataset: Contains atmospheric parameters such as temperature, dew point, humidity, wind speed, visibility, and pressure. These features are vital in classifying the weather condition, which significantly impacts crop selection.

Crop Dataset: Comprises soil and climatic attributes like temperature, humidity, pH, and rainfall, along with corresponding crop labels. This data is critical for training a model that recommends the best crop suitable for the given environmental conditions.

2. Data Preprocessing

To ensure the quality and consistency of input data, preprocessing steps are applied:

Removal of irrelevant columns (e.g., timestamps).

- Label encoding of categorical variables (weather condition and crop names) into numerical values for classification tasks.

Standardization of continuous variables using **StandardScaler** to normalize input features across both datasets.

Reshaping of feature matrices to meet the dimensional requirements of deep learning layers such as **Conv1D** and **LSTM**.

3. Model Training

Two separate deep learning models are designed and trained:

Weather Classification Model: A hybrid architecture using **1D Convolutional Neural Networks (Conv1D)** and **Long Short-Term Memory (LSTM)** layers is implemented to learn temporal patterns in weather data. The model outputs the predicted weather condition using a softmax layer for multi-class classification.

Crop Prediction Model: Similar to the weather model, a Conv1D-LSTM network is used, trained on environmental inputs (temperature, humidity, pH, and rainfall) to predict the most suitable crop. The model handles multi-class classification based on predefined crop categories.

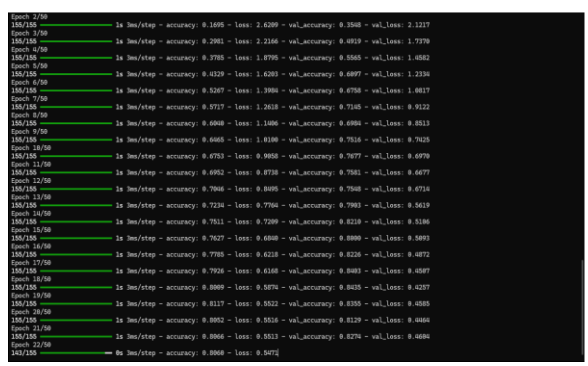

Fig.1 Model Training

4. Integration of Big Data Concepts

To address large-scale data challenges:
The data is analyzed and managed using structured formats suitable for batch processing and efficient querying.
MapReduce-like logic is simulated within preprocessing and clustering (if used), aligning with the principles of big data frameworks.

5. Evaluation Metrics

Both models are evaluated using standard classification metrics:
Accuracy: To assess how well the model classifies or predicts correct labels.
Loss Functions: Categorical crossentropy or sparse categorical crossentropy, depending on the label encoding.
Validation techniques such as **train-test splitting** and **cross-validation** are employed to ensure model generalizability.

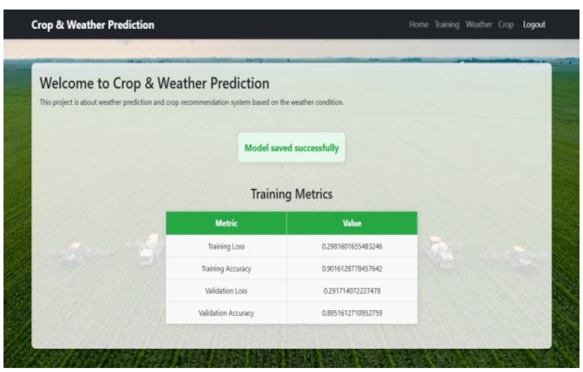

Fig.2 Metrics Evaluation

IV. SYSTEM ARCHITECTURE

The system architecture is presented in fig.1.

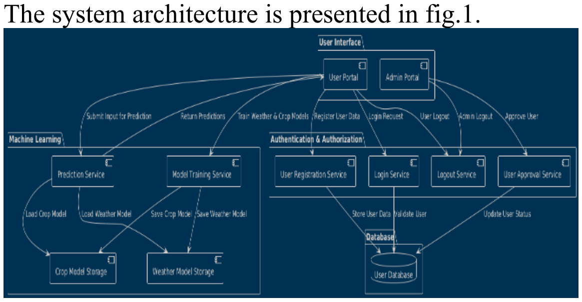

Fig.3. System architecture

The architecture of the proposed **Big Data-Driven Weather-Based Crop Prediction System** is designed to be modular, scalable, and user-centric. It consists of four main components: the **User Interface**, **Authentication & Authorization**, **Machine Learning Services**, and a **Centralized Database**. The user interface comprises two main portals: one for end-users (farmers) and another for administrators. Users can register, log in, and submit weather and soil parameters through the user portal to receive crop and weather predictions. The admin portal allows system administrators to manage users and oversee the model training process.
Authentication and authorization mechanisms ensure that only approved users can access the prediction services. Upon registration, user data is passed to the **User Registration Service**, which stores the data in the **User Database**. The **Login Service** validates login credentials, while the **User Approval Service** enables the admin to review and approve new registrations. Once authenticated, users can access the **Prediction Service**, which uses trained models stored in the **Crop Model Storage** and **Weather Model Storage** to provide real-time predictions. These models are trained and updated via the **Model Training Service**, which processes large-scale data using machine learning techniques, including LSTM and Conv1D architectures. This architecture not only ensures secure access control but also enables continuous improvement and scalability of prediction accuracy.

V. IMPLEMENTATION

The implementation of this project involves integrating big data analytics with machine learning techniques into a full-stack web application. The core functionality revolves around training deep learning models on weather and agricultural datasets, and providing crop recommendations based on user inputs through a user-friendly interface. The system has been developed using **Django (backend framework)**, **TensorFlow/Keras (for model building)**, **Pandas/NumPy (for data processing)**, and **Joblib (for model serialization)**.

1. User Registration and Authentication Module

The application includes a simple registration and login module:

- Users register by entering personal details (name, login ID, password, email, phone).
- Registered users are stored in the UserProfile model.
- During login, users are authenticated based on their status and credentials.
- Admin approval is required before a user can access the crop prediction features.

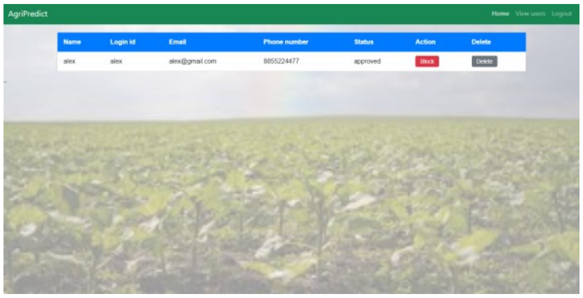

Fig.4 Authorizing User

2. Data Handling

Two datasets are used:
Weather Dataset (Weather Data.csv): Includes parameters like temperature, dew point temperature, humidity, wind speed, visibility, and pressure.
Crop Dataset (cpdata.csv): Contains features such as temperature, humidity, pH, and rainfall, along with crop labels.
These datasets are preprocessed using:
Label Encoding: To convert categorical labels (weather conditions, crop names) into numerical format.
Standard Scaling: Normalizes input values to improve learning efficiency.

3. Weather Prediction Model

Built using a combination of **Conv1D** and **LSTM** layers.
nput features are reshaped to match the expected 3D input of Conv1D.
The model is trained using categorical cross-entropy loss and Adam optimizer.
Trained model is saved as weather_model.h5, and the scaler and label encoder are saved using Joblib for future inference.

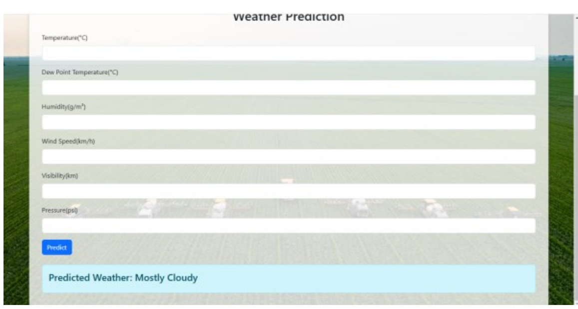

Fig.5 Weather Prediction Model

4. Crop Recommendation Model

- Uses temperature, humidity, pH, and rainfall as input features.
- Also based on **Conv1D** and **LSTM** layers with softmax output for multi-class classification.
- Encodes crop labels numerically, scales the input features, and reshapes them for model training.
- After training, the model (crop_model.h5), scaler, and label encoder are stored for predictions.

Fig.6 Crop Prediction Model

5. Prediction Functionality

- Weather Prediction:
  - Users input live environmental data.
  - The system preprocesses the input using the saved scaler and predicts the weather using the trained weather model.
  - The predicted numerical class is then decoded into a readable weather condition using the saved label encoder.
- Crop Prediction:
  - Users enter environmental data such as temperature, humidity, pH, and rainfall.
  - These inputs are scaled, reshaped, and passed to the trained crop model.
  - The model outputs the most suitable crop, which is decoded and displayed to the user.

6. User Interface and Output Display

- HTML templates (home.html, register.html, login.html, predict.html, predict_crop.html, etc.) are used for front-end pages.

- Results (predicted weather and recommended crop) are rendered dynamically using Django context variables.

7. Model Training Page

- A dedicated training view handles the model training process when triggered.
- Both weather and crop models are trained on server-side using uploaded datasets.
- The training metrics (accuracy and loss) are displayed to users for transparency.

VI FUTURE SCOPE

The integration of big data analytics and machine learning in agriculture holds immense potential for future advancements. As this system currently predicts suitable crops based on static weather and soil parameters, one major enhancement could be the incorporation of **real-time weather data streams** from APIs or IoT-based sensors. This would allow the model to provide more dynamic and up-to-date predictions, empowering farmers to make decisions instantly and accurately.

Another promising direction is to **extend the system's capabilities to predict pest attacks, disease outbreaks**, and yield estimation, using a combination of satellite imagery and environmental data. Additionally, incorporating **geospatial data and satellite-based remote sensing** could improve the accuracy of region-specific predictions, enabling more localized recommendations. Furthermore, the inclusion of **recommendation systems for fertilizers and irrigation schedules** based on soil and crop types can enhance the platform's usability.

From a usability perspective, developing a **mobile application or multilingual interface** will ensure greater accessibility for farmers across different regions of the country, especially in rural areas. Future improvements may also involve integrating blockchain for **secure agricultural data sharing and traceability**. Lastly, the model can evolve into a full-fledged decision support system for agricultural policymakers, enabling data-driven decisions at scale for national food security.

VII. CONCLUSION

This research presents an intelligent and scalable weather-based crop prediction model that leverages the power of **big data analytics and machine learning** to enhance agricultural productivity. The proposed system effectively integrates weather data, soil conditions, and crop requirements to recommend the most suitable crop for cultivation in a particular region. By employing preprocessing techniques such as normalization, encoding, and scaling, the raw environmental data is transformed into a structured format that can be efficiently utilized by advanced deep learning models.

The implementation of a hybrid machine learning pipeline—comprising **Conv1D, LSTM, and Dense layers**—allows for accurate classification and prediction of weather patterns and optimal crop selection. The use of **separate weather and crop models** enables modular training, better generalization, and high prediction accuracy. Moreover, by incorporating a **user-friendly interface** and a robust **authentication system**, the platform ensures secure, accessible interactions for both users and administrators.

The results demonstrate that the system can effectively guide farmers in selecting crops that are compatible with the current climatic conditions, soil features, and rainfall expectations. Additionally, the model architecture is designed to accommodate future enhancements such as **real-time weather integration**, **pest/disease alerts**, and **fertilizer recommendations**, which would further increase its practical utility. This platform serves as a promising step toward **precision agriculture**, enabling data-driven decision-making and empowering farmers to mitigate risk and improve yields through intelligent crop planning.

REFERENCES

[1]. Jiang, J., Zhang, Q., Yao, X., Tian, Y., Zhu, Y., Cao, W., & Cheng, T. (2020). HISTIF: A New Spatiotemporal Image Fusion Method for High-Resolution Monitoring of Crops at the Subfield Level. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 4607 4626.

[2]. Rashid, M., Bari, B. S., Yusup, Y., Kamaruddin, M. A., & Khan, N. (2021). A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction. IEEE Access, 9, 63406 63439.

[3]. Elavarasan, D., & Vincent, P. D. (2020). Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications. IEEE access, 8, 86886 86901.

[4]. Qiao, M., He, X., Cheng, X., Li, P., Luo, H., Tian, Z., & Guo, H. (2021). Exploiting Hierarchical Features for Crop Yield Prediction Based on 3-D Convolutional Neural Networks and Multikernel Gaussian Process. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 4476-4489.

[5]. Ashok, T., & Suresh Varma, P. (2020). Crop prediction based on environmental factors using machine learning ensemble algorithms. In Intelligent Computing and Innovation on Data Science (pp. 581-594). Springer, Singapore.

[6]. Gupta, R., Sharma, A. K., Garg, O., Modi, K., Kasim, S., Baharum, Z., ... & Mostafa, S. A. (2021). WB-CPI: Weather based crop prediction in India using big data analytics. IEEE Access, 9, 137869-137885.

[7]. Bendre, M. R., Thool, R. C., & Thool, V. R. (2015, September). Big data in precision agriculture: Weather forecasting for future farming. In 2015 1st International Conference on Next Generation Computing Technologies (NGCT) (pp. 744-750). IEEE.

[8]. Majumdar, J., Naraseeyappa, S., & Ankalaki, S. (2017). Analysis of agriculture data using data mining techniques: application of big data. Journal of Big data, 4(1), 1-15.

[9]. Prediction Of Crop Yield In Precision Agriculture Using Machine Learning Methods. (2021).Webology.https://doi.org/10.29121/web/v18i4/122.

[10]. Channe, H., Kothari, S., & Kadam, D. (2015). Multidisciplinary model for smart agriculture using intern

[11]. Bhat, S. A., & Huang, N. F. (2021). Big data and ai revolution in precision agriculture: Survey and challenges. IEEE Access, 9, 110209-110222.

[12]. Rao, N. H. (2018, September). Big data and climate smart agriculture-status and implications for agricultural research and innovation in India. In Proc. Indian Natl. Sci. Acad (Vol. 84, No. 3, pp. 625-640).

[13]. Cravero, A., & Sepúlveda, S. (2021). Use and adaptations of machine learning in big data—Applications in real cases in agriculture. Electronics, 10(5), 552.

[14]. Ngo, V. M., Le-Khac, N. A., & Kechadi, M. (2019, June). Designing and implementing data warehouse for agricultural big data. In International Conference on Big Data (pp. 1-17). Springer, Cham.

[15]. Sahu, S., Chawla, M., & Khare, N. (2019). Viable Crop Prediction Scenario in BigData Using a Novel Approach. In Emerging Technologies in Data Mining and Information Security (pp. 165-177). Springer, Singapore.

[16]. Ranjani, J., Kalaiselvi, V. K. G., Sheela, A., & Janaki, G. (2021, December). Crop Yield Prediction Using Machine Learning Algorithm. In 2021 4th International Conference on Computing and Communications Technologies (ICCCT) (pp. 611-616). IEEE.

[17]. Sagana, C., Keerthika, P., Thangatamilan, M., Kamali, R., Nanthini, K., & Maghathani, S. (2022, January). Identification of Suitable Crop Based on Weather Condition. In 2022 International Conference on Computer Communication and Informatics (ICCCI) (pp. 01-06). IEEE.

[18]. Kumar, R., Singh, M. P., Kumar, P., & Singh, J. P. (2015, May). Crop Selection Method to maximize crop yield rate using machine learning technique. In 2015 international conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM) (pp. 138-145). IEEE.

[19]. Gupta, R., Sharma, A. K., Garg, O., Modi, K., Kasim, S., Baharum, Z., ... & Mostafa, S. A. (2021). WB-CPI: Weather based crop prediction in India using big data analytics. IEEE Access, 9, 137869-137885.

[20]. Bodapati, N., Himavaishnavi, J., Rohitha, V., Jagadeeswari, D. L., & Bhavana, P. (2022, March). Analyzing Crop Yield Using Machine Learning. In 2022 International Conference on Electronics and Renewable Systems (ICEARS) (pp. 1-8). IEEE.

[21]. Vineela, A., Lavanya Devi, G., Nelaturi, N., Dasavatara Yadav, G "A Comprehensive Study and Evaluation of Recommender Systems" Lecture Notes in Electrical Engineeringthis link is disabled, 2021, 655, pp. 45 53.