Reviews on Crop Yield Prediction

Prof. Kalyani[1], Snehitha Srinivas[2], Aswin Manoj[3], Ayesha Javeriya[4], Ashwanth Dasari[5], Athwika Gade[6], B.V.S. Sayeesh sai. A[7]
[1]*Assistant professor, Malla Reddy university*
[2,3,4,5,6,7]*AI &ML, Malla Reddy University, Hyderabad, India*

Abstract: **Time series analysis is a technique used to analyze statistical data over a period of time. When predicting future events over an extended period of time, this method is trustworthy and scientific. Time series analysis could always forecast the probability of production almost exactly. The focus of prediction in this work is food production. In this research, logistic regressions and Random Forests are the most popular classification techniques. This research projects crop production over an extended period of time using the proposed ensemble model. The Random Forest and Logistic Regression techniques are contrasted with this ensemble model. The accuracy and the classification error are the two factors that are employed independently for the output prediction.**

Keywords: Crop yield prediction, Decision trees, Systematic literature review Machine learning, Deep learning

I. INTRODUCTION

Machine learning (ML) techniques are applied in a variety of domains, such as predicting customers' phone usage (Witten et al., 2016) and evaluating customer behavior in supermarkets (Ayodele, 2010).

Agriculture has been using machine learning for a number of years (McQueen et al., 1995). One of the difficult issues in precision agriculture is crop production prediction, for which numerous models have been developed and proven thus far. Since crop production varies on a wide range of parameters, including soil, weather, fertilizer use, and seed variety, this problem necessitates the use of several datasets (Xu et al., 2019). This suggests that predicting agricultural production involves a number of intricate procedures and is not a simple operation. These days, methods for predicting crop production can calculate the reasonable actual yield, although improved yield prediction performance is still preferred (Filippi et al., 2019a).

As a practical technique, machine learning—a subfield of artificial intelligence (AI) focused on learning—can improve yield prediction by taking into account many features. From datasets, machine learning (ML) may extract information by identifying patterns and connections. Datasets with outcomes represented by prior experience are required for training the models. Since the predictive model is constructed utilizing a number of features, historical data is used during the training phase to define the model's parameters. A portion of the previous data that hasn't been used for training is used for performance evaluation during the testing phase. The descriptive and predictive capabilities of an ML model vary based on the study problem and research objectives. Predictive models are used to forecast future events, whereas descriptive models are used to learn from the data gathered and describe what has occurred (Alpaydin, 2010). Several obstacles must be overcome in machine learning research in order to develop a high-performing predictive model. The correct algorithms must be chosen in order to tackle the issue at hand, and the underlying platforms and algorithms must be able to handle the volume of data.

We conducted a systematic literature review (SLR) to obtain an overview of the work that has been done on the application of ML in agricultural production prediction. A comprehensive review of the literature, or SLR, reveals any possible gaps. Provides research on a specific problem area and serves as a roadmap for practitioners and researchers who want to conduct additional research on that issue. All pertinent studies are retrieved from electronic databases using an SLR technique, combined, and presented to address the research questions outlined in the study. An SLR study opens up new viewpoints and aids in the understanding of the state-of-the-art by novice researchers in the subject.

II. LITERATURE REVIEW

Machine learning (ML) techniques are applied in a variety of domains, from the evaluation of consumer behavior in supermarkets (Ayodele, 2010) to the prediction of phone use by clients (Witten et al., 2016). Agriculture has been using machine learning for a number of years (McQueen et al., 1995). One of the more difficult issues in precision agriculture is predicting crop yield, and numerous models have been put forth and proven effective thus far. Given that crop output is dependent on a wide range of variables, including soil, weather, fertilizer use, and seed variety, this problem necessitates the use of several datasets (Xu et al., 2019). This suggests that predicting agricultural production involves a number of intricate procedures and is not a simple operation. These days, algorithms for predicting crop yield can estimate the actual yield fairly, but improved yield prediction performance is still preferred (Filippi et al., 2019a).

Machine learning, a subfield of artificial intelligence (AI) that focuses on learning, is a useful strategy that can improve yield prediction depending on a number of factors. From datasets, machine learning (ML) may extract knowledge and identify trends and connections. Datasets with outcomes represented by prior experience are required to train the models. Since the predictive model is constructed using a number of features, its parameters are established during the training phase by utilizing historical data. A portion of the historical data that was not used for training is used for performance evaluation during the testing phase. The descriptive and predictive qualities of an ML model vary based on the study problem and research objectives. Predictive models are used to forecast future events, whereas descriptive models are used to learn from the data gathered and describe what has occurred (Alpaydin, 2010). Several obstacles must be overcome in machine learning research in order to develop a high-performing predictive model. The correct algorithms must be chosen in order to tackle the issue at hand, and both the underlying platforms and the algorithms must be able to handle the volume of data.

III. PROBLEM STATEMENT

Statistical Models: These utilize historical crop yield data and associated factors (like weather, soil, etc.) to develop mathematical equations that predict future yields based on correlations found in the data.

Machine Learning Algorithms: ML algorithms, such as neural networks, random forests, or support vector machines, learn patterns from historical data and make predictions based on those patterns. They can handle large datasets and complex relationships among variables.

Remote Sensing: This involves using satellite imagery and remote sensors to gather information about crop health, growth, and conditions. Advanced analytics extract insights that contribute to yield predictions.

Weather Data Integration: Incorporating weather data (temperature, precipitation, humidity, etc.) is crucial for yield predictions. Models factor in how different weather patterns affect crop growth at various stages.

Crop Simulation Models: These simulate crop growth and development by integrating data on soil properties, weather, management practices, and crop genetics. They predict yields by simulating the crop's response to various conditions.

Each methodology has its strengths and weaknesses, often complementing each other when combined to improve the accuracy of crop yield predictions.

IV.EXPERIMENTAL STEPS

Experimental steps includes the procedure we have followed to meet the desired outcome. It includes all the methods and the steps to follow. The following output screens shows the libraries used and its working.

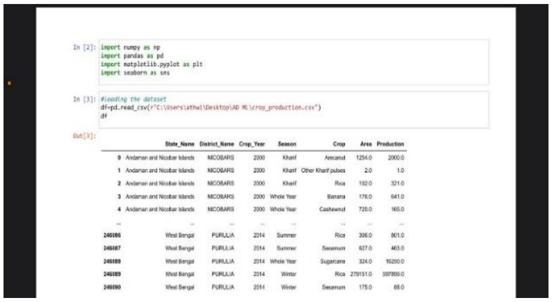
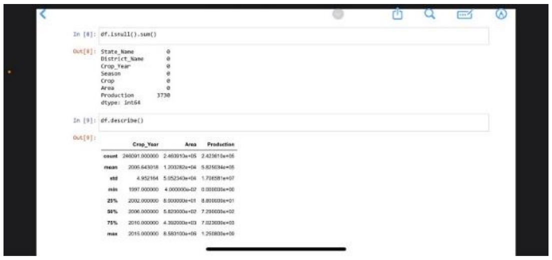
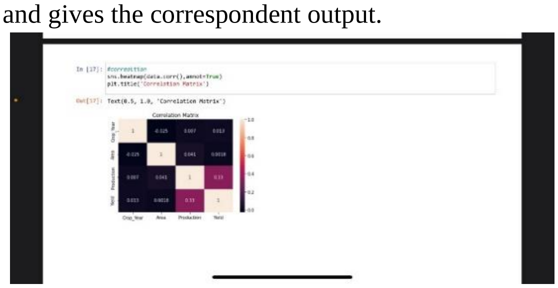


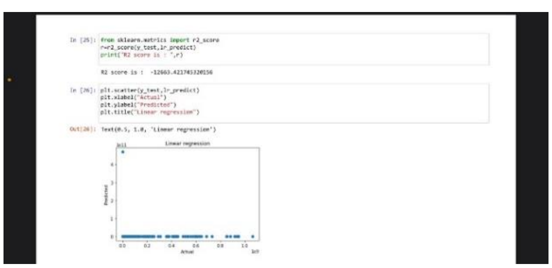
Fig1. Showing libraries

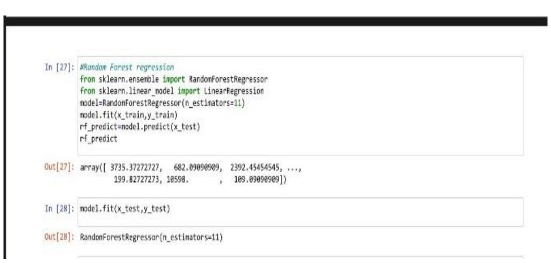


V.EXPERIMENTAL RESULTS

In the beginning, we need to import all the libraries from the command prompt, and later in Jupyter write and run the code. In this result, the assistant asks for the voice as an input, interprets the command given, and gives the correspondent output.







VI.CONCLUSION

It is possible to use machine learning algorithms to predict crop yields. There are several factors that can influence crop yields, including weather patterns, soil quality, and pest infestations.

By training a machine learning model on a dataset that includes these and other relevant factors, it may be possible to predict crop yields with a high degree of accuracy.

However, the accuracy of the prediction will depend on the quality of the training data and the effectiveness of the chosen machine learning algorithm.

It is also important to consider that crop yields can be affected by unpredictable events such as natural disasters, which may not be captured in the training data and could impact the accuracy of the prediction.

FUTURE WORK

In coming years, can try applying data independent system. That is whatever be the format our system should work with same accuracy. Integrating soil details to the system is an advantage, as for the selection of crops knowledge on soil is also a parameter. Proper irrigation is also a needed feature crop cultivation. In reference to rainfall can depict whether extra water availability is needed or not. This research work can be enhanced to higher level by availing it to whole India.

REFERENCE

Cheng Zhou, Boris Cule, Bart Goethals “Pattern Based Sequence Classification”, IEEE Transactions on Knowledge and Data Engineering, Vol. 28, No.5, 2016, pp.1285-1298.

B. Tang, H. He, P. M. Baggenstoss and S. Kay “A Bayesian Classification Approach Using Class-Specific Features for Text Categorization”, IEEE Transactions on Knowledge and Data Engineering, Vol.28, No.6, 2016, pp.1602-1606.

P.Samuel Quinan, Miriah Meyer “Visually Comparing Weather Features in Forecasts”, IEEE Transactions on Visualization and Computer Graphics, Vol. 22, No.1, 2016, pp. 389-398.

Y. Chen and Y. Li "Entropy-Based Combining Prediction of Grey Time Series and Its

Application", IEEE International Conference on Intelligent Computation Technology and Automation (ICICTA), 2009, pp. 37-40.
G. Chen, X. Xu, G. Wang and H. Chen "The corn output in a time series prediction model", IEEE International Conference on World Automation Congress (WAC), 2010, pp. 283-286.
A. Dehzangi, K. Paliwal, A. Sharma, O. Dehzangi and A. Sattar "A Combination of Feature Extraction Methods with an Ensemble of Different Classifiers for Protein Structural Class Prediction Problem", IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 10, No.3, 2013, pp. 564-575.
R. Kumar, M. P. Singh, P. Kumar and J. P. Singh "Crop Selection Method to maximize crop yield rate using machine learning technique" IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 2015, pp. 138-145.