International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-12 Issue-2, December 2022

Agriculture Crop Yield Analysis and Prediction using Feature Selection based Machine Learning Techniques

T. V. Rajinikanth, Burma Kavya, Narameta Thanuja Sri, Alley Yashwanth Saikrishna

Abstract: *Agriculture is being the world's largest industry; it plays a major role in maintaining the economic stability of developing countries. Because of the responsibilities that this sector bears, it is critical to find the precision of production in making profitable decisions in agricultural sector. Machine learning is the most effective tool for making decisions. Machine learning techniques with correct optimizations have been utilized in conjunction with the use of multiple algorithms and create an accurate model for predicting production and also in guiding to improve crop cultivation for enhanced output. The elements like cost of cultivation, cost of production, and yield are utilized to predict the crop yield during the analysis. In this study, the necessary data was acquired, and the methodologies and features employed in agricultural yield analysis were studied. During the literature survey more than 50 articles were referred for analysis. Relevant topics were collected from electronic databases and found useful machine learning approaches with which desired model was developed. Along with Random Forest, Decision Trees, and Support Vector Machine, Gaussian Nave Bayes, and Ada Boost machine learning techniques, Carl Pearson Correlation, Mutual Information, and Chi Square Feature* Selection *techniques were applied. The accuracy percentage for different algorithms was calculated crop yield prediction with and without feature selection approaches. We also used time complexities to figure out which method is the most efficient and accurate.*

Keywords: *Yield Analysis, Decision Tree, Feature Selection, Random Forest, Crop Selection.*

I. INTRODUCTION

Even for big datasets, machine learning is a promising tool for both descriptive and predictive answers; it was crucial in our prediction effort. Deep learning approaches use a practical approach to make better predictions by analyzing the data presented based on various features.

Manuscript received on 28 November 2022 | Revised Manuscript received on 03 December 2022 | Manuscript Accepted on 15 December 2022 | Manuscript published on 30 December 2022.

*Correspondence Author (s)

Dr. T. V. Rajini Kanth, Professor & Head, Department of Computer Science Engineering- AI & ML, SNIST, Hyderabad (Telangana), India.

Burma Kavya*, B.Tech Students, Department of Computer Science and Engineering, Sreenidhi Institute of Science and Technology, Hyderabad (Telangana), India. E-mail: burma.kavya1312@gmail.com

Narameta Thanuja Sri, B.Tech Students, Department of Computer Science and Engineering, Sreenidhi Institute of Science and Technology, Hyderabad (Telangana), India.

Alley Yashwanth Saikrishna, B.Tech Students, Department of Computer Science and Engineering, Sreenidhi Institute of Science and Technology, Hyderabad (Telangana), India.

Machine Learning, a part of Artificial Intelligence, will entail finding patterns and correlations and uncovering information, which can then be used to create a modelfor an application using data sets and algorithms to train and test the model. Because the agriculture business is growing at an exponential rate in tandem with the population, finding an accurate yield prediction based on various criteria has become difficult. Agriculture production can be forecastin a variety of methods utilizing a variety of criteria such as crop requirements, geographical and climate data, and so on. In addition, in this study, gathered data on state, crop, cost of cultivation, cost of cultivation C2, cost of production, and yield in order to generate accurate predictions, which aids in yield prediction. Machine learning approaches were used to compare the accuracy of algorithms including Support Vector Machine types Polynomial, RBF, Linear, Decision Tree, Gaussian Nave Bayes, Random Forest and Ada Boost in order to discover the best accurate predictions for agricultural productivity.

The identification was made in this investigation, and applied a feature selection strategy to improve the model's performance in a later deep analysis. To choose features use the knowledge from research sources and methodologies like Pearson Correlation, Mutual Information, and Chi Square. The model runs, trains, and tests based on the data provided. We use feature selection to remove unnecessary and non-used data before applying the algorithms to create the best model possible. Data improvement leads to an effective model, so feature selection is used to remove unnecessary and un-used data before applying the algorithms to create the best model possible. This makes it easier for the model to learn and interpret the data set so that it can make predictions or provide solutions. We also divided the best working algorithms based on their time complexity in order to narrow down the best algorithms and their functionality among the available options, as well as to make our study useful for future research. This was employed in our study to establish the best successful method out of all test cases both before andafter feature selection.

II. LITERATURE REVIEW

S. Veenadhari et al. [1] paper is about the existence of the correlation between the climatic factors and the Soyabean crop productivity using decision tree algorithm and also they stated that the algorithm is helpful in prediction of the conditions for the Soyabean crop production with particular climatic parameters.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Avat Shekoofa et al. [2] reviewed that the decision tree models play crucial role in understanding relationships among the physiological and agronomic features in selecting the important traits for predicting the maize yield. Manoj et al. [3] stated that the data from the dataset they chose being highly variable because of which decision tree repressor is used for further prediction and analysis in the research. Mayank Champaneri et al. [4] provided with the research on predicting the crop yield using random forest algorithm based on the climatic input parameters. Yvette Everingham et al. [5] Mentioned about the search space predictor variables. It stated that when the search space pf predictor variables are large then the algorithm random forest can cope by generating model. Mayank Champaneri et al. [6] explained about the random forests functionality. It states that random forest being the most popular and effective machine learning algorithm, it has the capability to both regression and classification tasks. They evaluated the construction of the random forest is done by the multiple decision trees. Sujal A et al. [7] reviewed that the predictions made be using the machine leaning algorithms helps farmers in deciding the crop to choose for the cultivation to get the maximum yield by considering factors like crop name, temperature, soil quality, rainfall etc. SuYing-xue et al. [8] paper's about the Support vector machine-based open crop model (SBOCM), which contains information about its development and yield prediction. W. Mupangwa et al. [9] reviewed that the yield prediction models play a vital role in exploring the technologies involving multiple experiments. They also mentioned that they found the KNN algorithm to be the most comparable one during their study of yield prediction on maize and that SVM was the worst algorithm. Rui Yan et al. [10] stated that an application of the face recognition with the Naïve Bayes algorithm has given an effective result and performs very better when compared with some other algorithms. Eibe Frank et al. [11] stated that Naïve Bayes algorithm performs well in classification tasks but it also works well when it predicts the numerical values for any dataset. S.R.Shankar Gowda et al. [12] stated that using Naïve Bayes algorithm we can classify the tweets on social media based on the where the sentiment is positive, negative, neutral. Pritika Bahad et al. [13] stated that ensemble machine learning techniques like boosting gives a very good accuracy prediction when compared with general machine learning techniques. Erico N DeSouza et al. [14] stated that the idea to differ the weak base classifier using classifier techniques can be dealt using resampling techniques. V Karthikeyan et al. [15] stated that the hybrid RF Adaboost algorithm gives a better accuracy of the predicted values when compared with the algorithms like KNN, SVM, and KNN-Adaboost. Subadhra Mishra et al. [16] mentioned that accurate crop production factors are important for economic and statistical decisions. They added, however, that earlier production predictions are not required because it requires a lot of descriptive assessment based on many criteria. Jig Han Jeong et al. [17] compared the Random Forest machine-learning algorithm's capacity to forecast agricultural yield based on geographical and climatic characteristics to the Multi Linear Regression. They claimed that of all the Multi-Linear Regression algorithms, the Random Forest was sure to be the most competent. Vaishali Pandith et al. [18] have done the mustard crop yield analysis based on the soil analysis from the different districts with the involvement of machine learning techniques K-Nearest Neighbour (KNN), Artificial Neural Network (ANN), Nave Bayes, Random Forest, and Multinomial Logistic Regression. They stated that among the applied algorithms, KNN and ANN can be the most accurate algorithms to give the prediction. Diego Gomez et al. [19] said that they used satellite remote sensing to forecast potato yields using nine machine-learning approaches. They claim that using the feature selection step is critical for reducing multicollinearity among predictors and resulting in a better model. P. Surya et al. [20] conducted research on predicting and identifying the most yield-giving crops in the Indian state of Tamil Nadu using various regression techniques. It also stated that to predict the agricultural crop production in tons, the predictor formula is very useful.

III. MOTIVATION

Agriculture is the world's largest industry, finding it increasingly difficult to forecast yield as the world's population grows exponentially. Our country, India, has over 70 percent of its population dependent on agriculture. Farmers may be unable to achieve the target yield owing to poor yield building instructions, resulting in losses. Suicides are being committed because of their inability to thrive in the business on which their families rely. In India, the fatality rate due to crop failure will have grown to 18% by 2020. All of this loss is the result of a faulty decision-making process when selecting a crop to cultivate. We sought to make a difference by using this cause as inspiration. Research was done with a hope that it may be useful to the agriculture industry and for future research.

IV. METHODOLOGY

Before beginning of our research, research instructions were reviewed on agricultural production prediction utilizing predictive analytic approaches published by Dr. T.V. Rajinikanth [21] et al. (2016). The questions are created at the beginning of the study, while following the rules, in order to get the study started quickly. The study's blueprint has been finished based on the report created by the questions, which gathered information from many sources on agricultural forecast analysis. The materials were acquired and evaluated by extracting relevant research and comparing them to a set of exclusion and quality criteria. The sources for applying algorithms and feature choices are obtained and then synthesized based on the research's goal.

4.1 Data Set Retrieval:

Every machine model is built on the foundation of data sets, which must be gathered and pre-processed using data pre-processing techniques to ensure that the data set is in the best possible state for the model's effective operation. The data set features were essentially thought out based on the project goal and work, and a search was undertaken to acquire them.

Retrieval Number:100.1/ijeat.B39421212222
DOI: 10.35940/ijeat.B3942.1212222
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The data set with the relevant attributes needed and has been extracted from Kaggle [21]. Crop, state, cost of cultivation, and cost of production are the attributes employed in the study to fulfil the goal of agricultural forecast. The data in the data set is mostly concerned with economic variables. The predictions of several agricultural analysis models are based on geographical and climatic factors. We did, however, make a prediction based on economic factors such as those indicated previously.

4.2 Algorithms building:

Five machine-learning algorithms make up the model in this study. The algorithms make predictions depending on the correctness of each individual's work. In some cases, the accuracy supplied by the algorithms may be identical, but incorporated extra features to forecast the best method in addition to the accuracy provided by the algorithms. Support Vector Machine with various kernel approaches (polynomial, RBF, and linear), Decision Tree, Random, Nave Bayes, and Ada Boost are the algorithms employed. To acquire the projected yield, we apply several algorithms to the dataset during algorithm development. We compare the yield predicted among different algorithms when feature selection is applied and when feature selection is not applied. When feature selection is

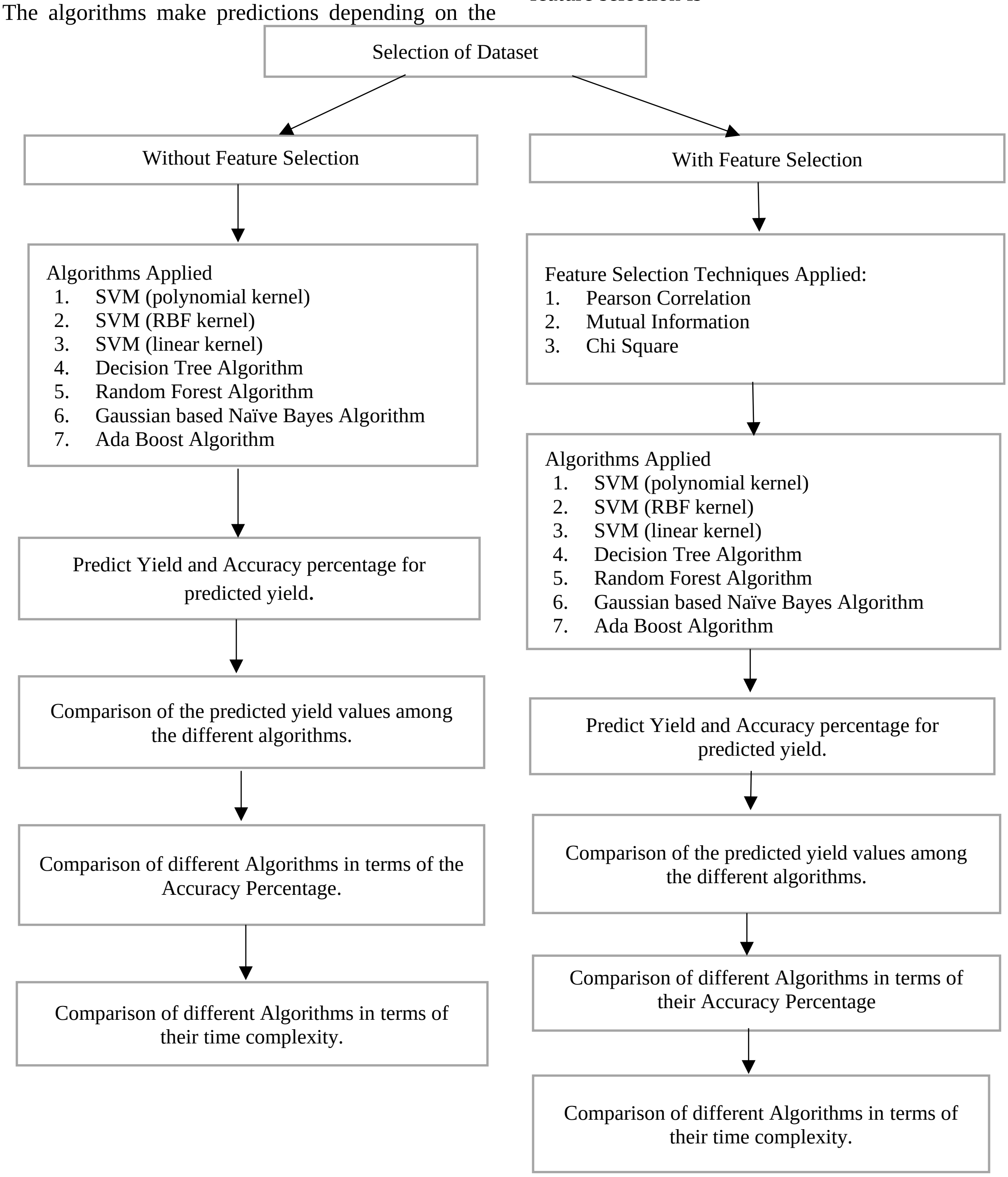

Figure 1: Architecture of Agriculture Crop Yield Analysis and Prediction.

Retrieval Number:100.1/ijeat.B39421212222
DOI: 10.35940/ijeat.B3942.1212222
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

applied to the dataset, few or one attribute is removed, which has the least preference among all the attributes. The Cost of Cultivation feature is deleted when using Pearson Correlation Feature Selection, while the Crop attribute is removed when using Mutual Information, Chi Square Feature Selection due to its lesser preference. The key benefit of adopting Feature Selection approaches is that they allow the model to run as quickly as possible with the least amount of time complexity while still providing the best output with the best-predicted values.

Fig.1 represents the Architecture of Agriculture Crop Yield Analysis and Prediction in which the flow of methodology shown. Two approaches were followed one without Feature selection and the other with Feature selection and comparisons were made between these two approaches.

4.3 Yield Prediction and Comparison

The yield prediction is done using different algorithms and depending on the attributes in the dataset. The yield can be forecasted in two ways: without Feature Selection and with various Feature Selection Techniques.

4.3.1 Without Feature Selection Techniques:

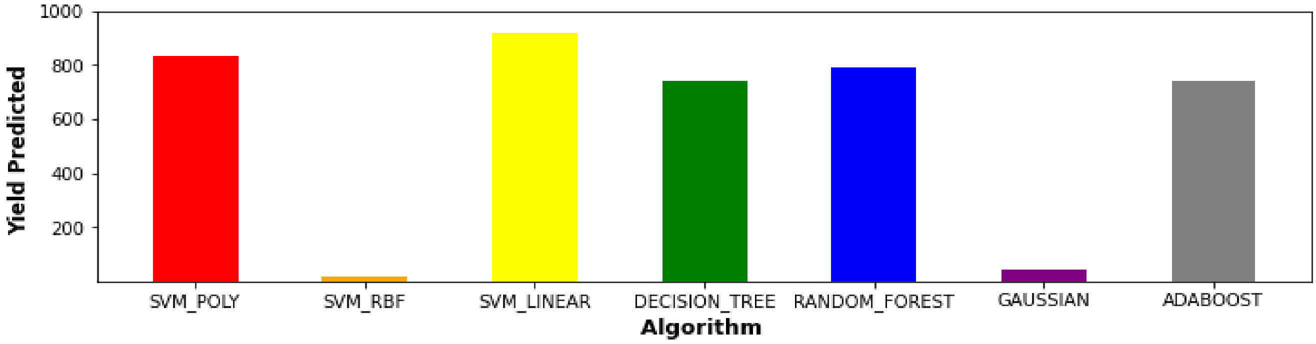

Figure 2: Predicted Yield Graph without Feature Selection.

The predicted yield with different algorithms are applied without any feature selection is represented graphically by taking different algorithms on the X-axis and predicted yield values on the Y-axis with a scale of 20 in the range of 0–100 in **Figure 2**.

4.3.2 With different Feature Selection Techniques:

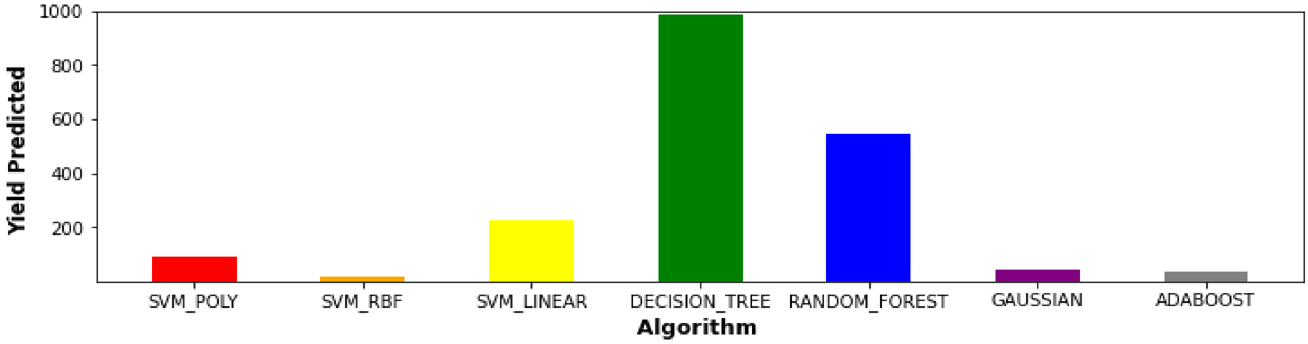

Figure 3: Predicted Yield Graph with Pearson Correlation Feature Selection.

The predicted yield when different algorithms are applied with Pearson Correlation Feature Selection is represented graphically with algorithm names on the X-axis and predicted yield values on the Y-axis with a scale of 20 in the range of 0–100 in **Figure 3**.

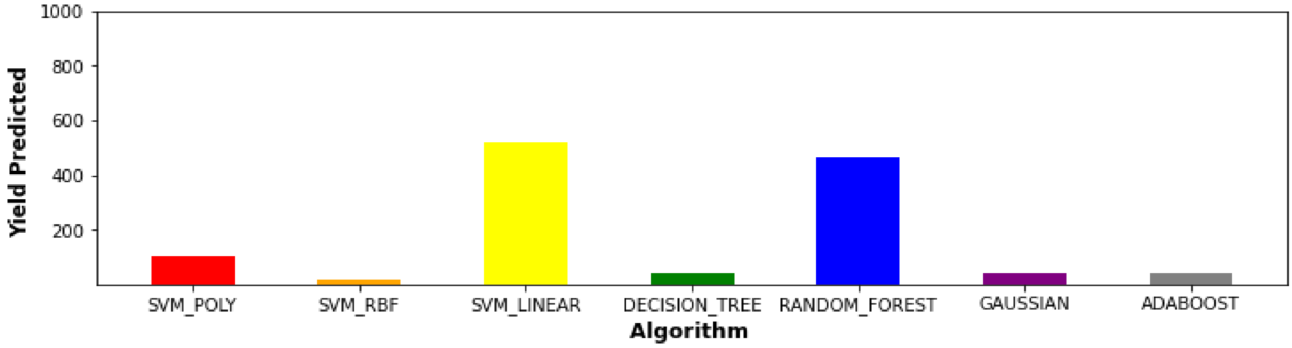

Figure 4: Predicted Yield Graph with Mutual Information Feature Selection.

Retrieval Number:100.1/ijeat.B39421212222
DOI: 10.35940/ijeat.B3942.1212222
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

The predicted yield when different algorithms are applied with Mutual Information Feature Selection is represented graphically with algorithm names on the X-axis and predicted yield values on the Y-axis with a scale of 20 in the range of 0–100 in **Figure 4**.

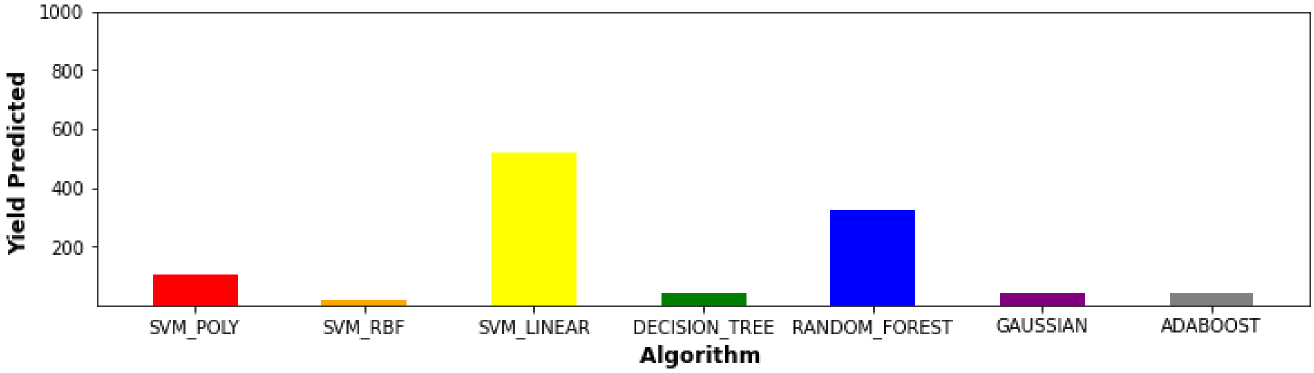

Figure 5: Predicted Yield Graph with Chi Square Feature Selection.

The predicted yield when different algorithms are applied with Chi Square Feature Selection is represented graphically with algorithm names on the X-axis and predicted yield values on the Y-axis with a scale of 20 in the range of 0–100 in **Figure 5**.

4.4 Accuracy Prediction and Comparison

The accuracy percentage for the predicted yield is done by applying confusion matrix to yield predicted using different algorithms. The predicted accuracy percentage is compared among different algorithms.

4.4.1 Without Feature Selection Techniques:

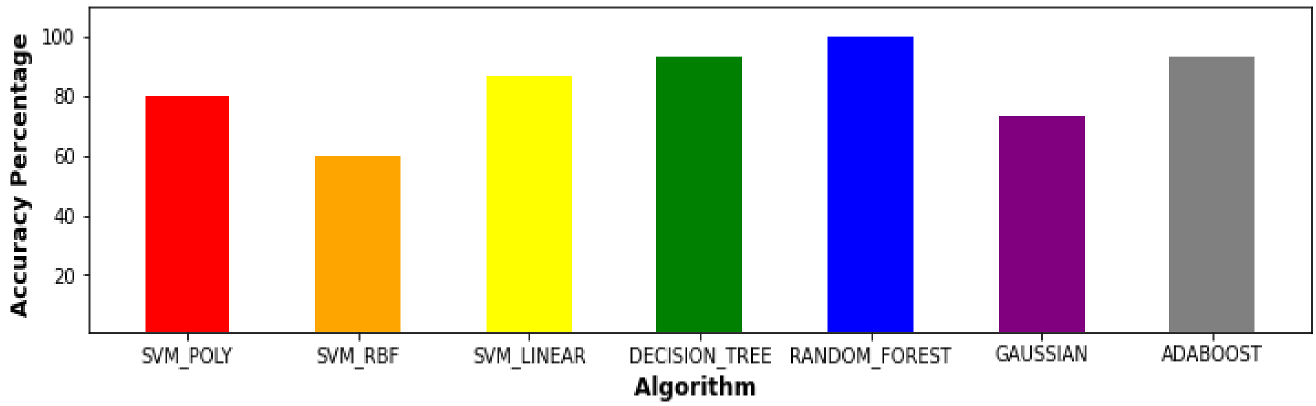

Figure 6: Accuracy Percentage Graph without Feature Selection

The accuracy comparison of the applied algorithms without feature selection is represented graphically with algorithm names on the X-axis and accuracy percentage on the Y-axis with a scale of 20 in the range of 0–100 in **Figure 6**. The bar plot shows different accuracy percentage values for different algorithms, and among all the algorithms, the Random Forest algorithm gives the best accuracy percentage, i.e., 100%.

4.4.2 With different Feature Selection Techniques:

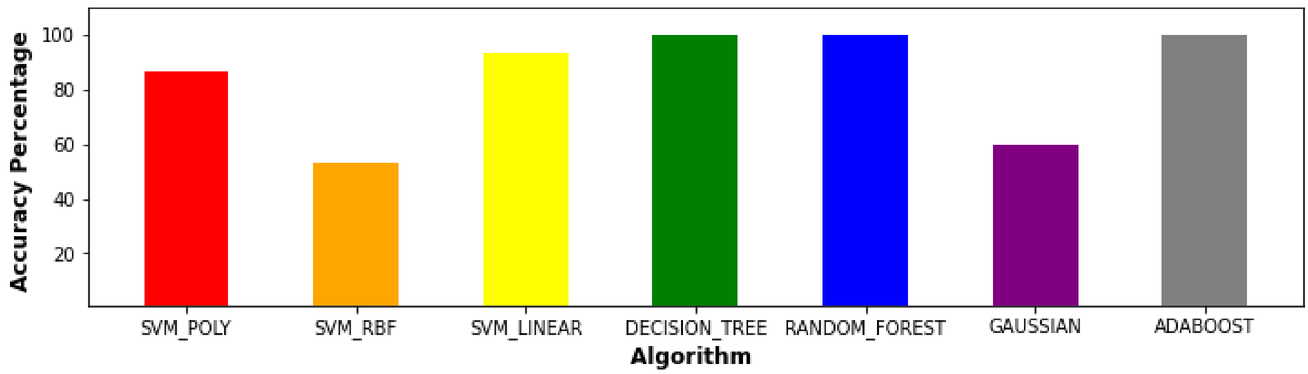

Figure 7: Accuracy Percentage Graph with Pearson Correlation Feature Selection

Retrieval Number:100.1/ijeat.B39421212222
DOI: 10.35940/ijeat.B3942.1212222
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

The accuracy comparison of the applied algorithms with Pearson Correlation Feature Selection is represented graphically with algorithm names on the X-axis and accuracy percentage on the Y-axis with a scale of 20 in the range of 0–100 in **Figure 7**. The bar plot shows different accuracy percentage values for different algorithms with Pearson Correlation Feature Selection and, among all the algorithms - Random Forest, Ada Boost, and Decision Tree algorithms give the best accuracy percentage, i.e., 100%.

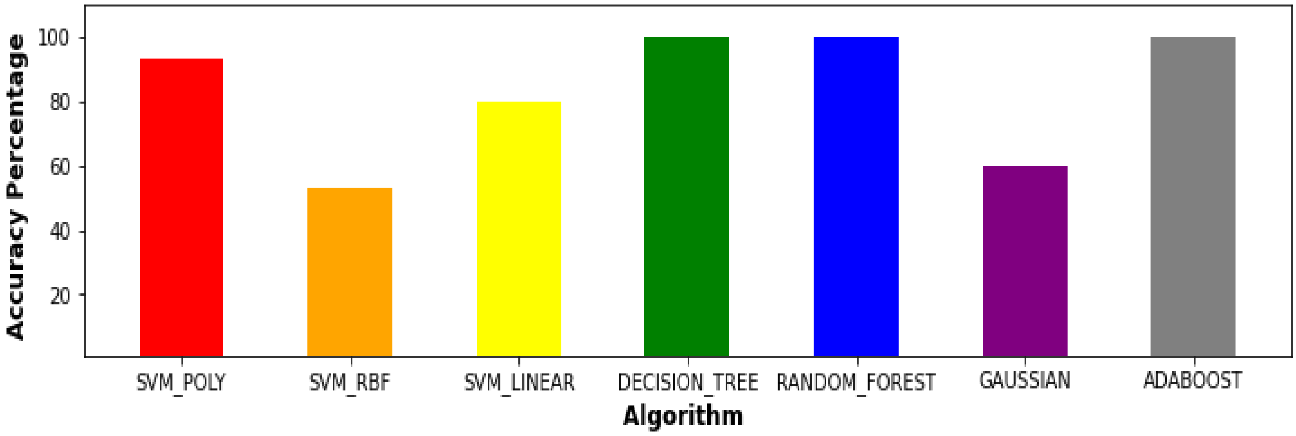

Figure 8: Accuracy Percentage Graph with Mutual Information Feature Selection

The accuracy comparison of the applied algorithms with Mutual Information feature selection is represented graphically with algorithm names on the X-axis and accuracy percentage on the Y-axis with a scale of 20 in the range of 0–100 in **Figure 8**. The bar plot shows different accuracy percentage values for different algorithms with Mutual Information Feature Selection and, among all the algorithms – Ada Boost, Decision Tree, and Random Forest algorithms give the best accuracy percentage, i.e., 100%.

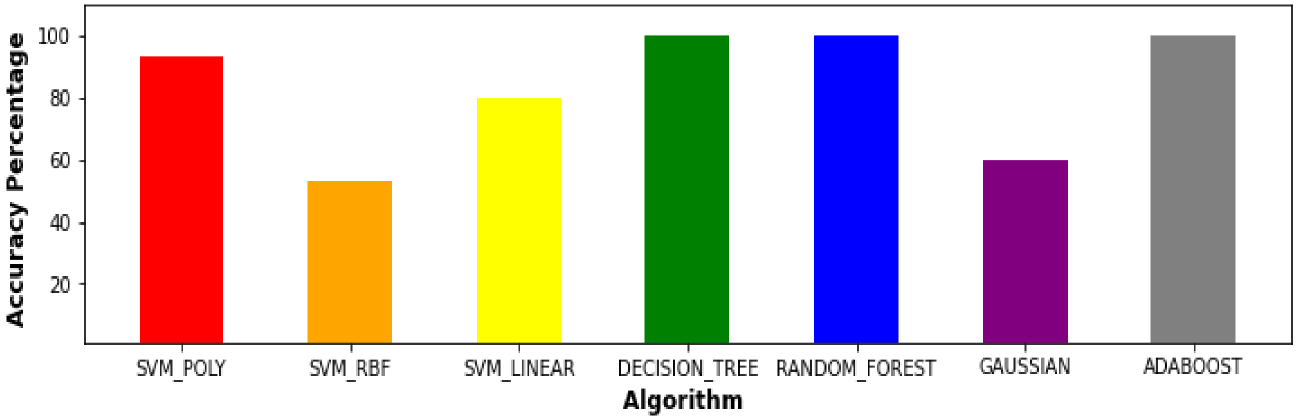

Figure 9: Accuracy Percentage Graph with Chi Square Feature Selection

The accuracy comparison of the applied algorithms with Chi Square feature selection is represented graphically with algorithm names on the X-axis and accuracy percentage on the Y-axis with a scale of 20 in the range of 0–100 in **Figure 9**. The bar plot shows different accuracy percentage values for different algorithms with Chi Square Feature Selection and, among all the algorithms – Decision Tree, Ada Boost, and Random Forest algorithms give the best accuracy percentage, i.e., 100%.

V. TIME COMPLEXITY AND COMPARISON

Time complexity is calculated by considering the time taken by each algorithm to complete its execution. Here we are going with time complexity because we have three algorithms with best accuracy so the algorithm, which takes less time when compared with other algorithms, is considered as the best algorithm for prediction of yield. The study includes the three levels of the filtrations for the differentiation of the best algorithm among the applied algorithms. In first step compared the accuracy produced by the algorithm mentioned in section two. In second step the feature selection techniques were applied on the algorithms mentioned in section 3. In the third step filtration includes the time complexity of the algorithms used for before and after feature selection techniques application. The comparisons of the complexities were given in the table [2]. Based on the last filtration of the study the effective algorithm can be mentioned to predict the yield of the crop and that is Random Forest. The graphical presentation of the time complexities comparison is made to make the future approaches easy for further development.

Retrieval Number:100.1/ijeat.B39421212222
DOI: 10.35940/ijeat.B3942.1212222
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

5.1 Without Feature Selection Techniques

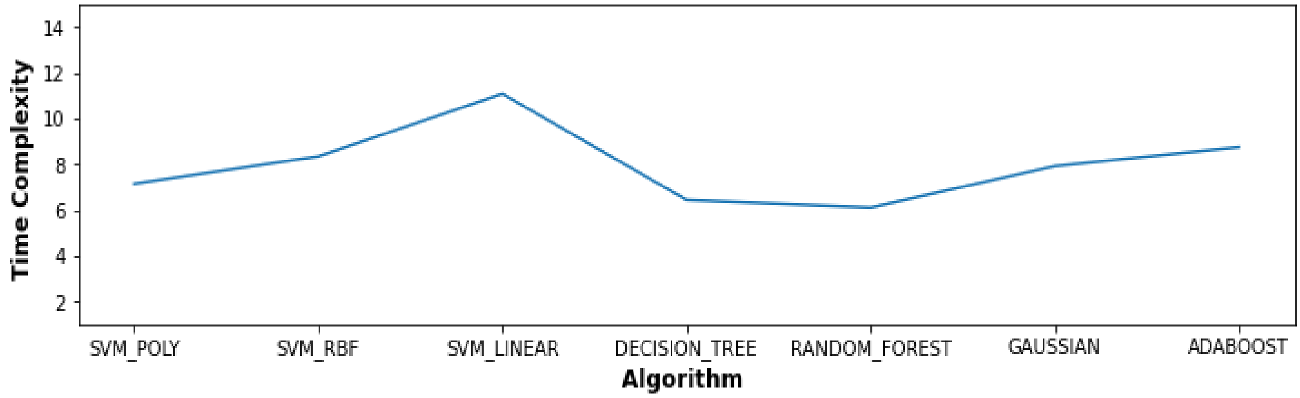

Figure 10: Time Complexity Graph without Feature Selection

The time complexity is taken for every algorithm, with the time taken to run on the X-axis and algorithm names on the Y-axis without any Feature Selection techniques in the **Figure 10**. The above graph shows the time taken by each algorithm without any feature selection technique, and among all of them, Random Forest takes the least time to complete the execution.

5.2 With Different Feature Selection Techniques

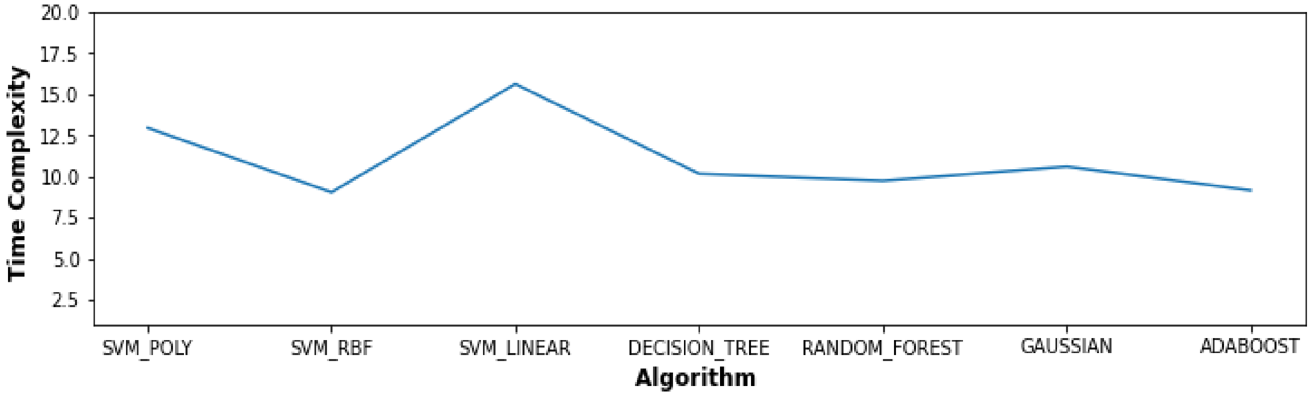

Figure 11: Time Complexity Graph with Pearson Correlation Feature Selection

The time complexity is taken for each algorithm, with the time taken to run on the X-axis and algorithm names on the Y-axis with the Pearson Correlation Feature Selection technique in **Figure 11**. The above graph shows the time taken by each and every algorithm with the Pearson Correlation feature selection technique, and among all of them, the time complexity comparison is done between only the algorithms which have 100% accuracy of yield, i.e., between Decision Tree, Random Forest, and Ada Boost algorithms. Therefore, between those three algorithms, the time complexity is less for the Ada Boost algorithm.

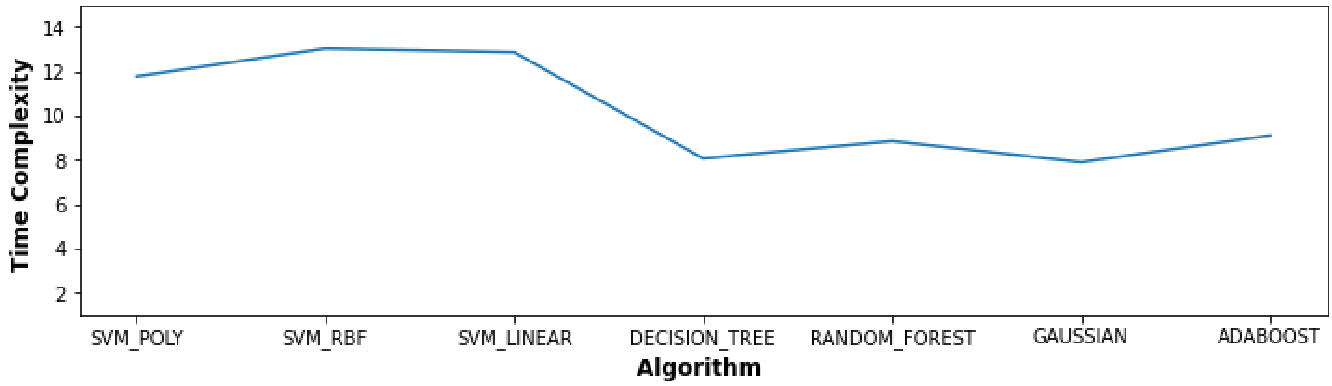

Figure 12: Time Complexity Graph with Mutual Information Feature Selection

The time complexity is considered for each algorithm, with the time taken to run on the X-axis and algorithm names on the Y-axis with the Mutual Information technique for Feature Selection in **Figure 12**.

Retrieval Number:100.1/ijeat.B39421212222
DOI: 10.35940/ijeat.B3942.1212222
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

From the above graph, it shows the time taken by each and every algorithm with the Mutual Information feature selection technique, and among all of them, the time complexity comparison is done only between the algorithms which have 100% accuracy of yield, i.e., between Random Forest, Decision Tree, and Ada Boost algorithms. Therefore, between those three algorithms, the time complexity is less for the Decision Tree algorithm.

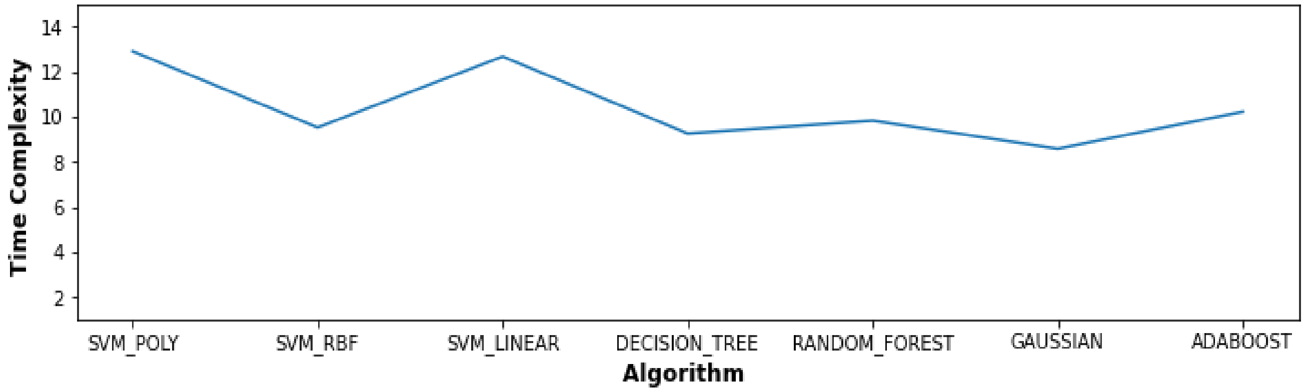

Figure 13: Time Complexity Graph with Chi Square Feature Selection

The time complexity is taken for each algorithm, with the time taken to run on the X-axis and algorithm names on the Y-axis with the Chi Square Feature Selection technique in **Figure 13**. The above graph shows the time taken by each and every algorithm with the Mutual Information feature selection technique, and among all of them, the time complexity comparison is done only between the algorithms which have 100% accuracy of yield, i.e., between Ada Boost, Random Forest, and Decision Tree algorithms. Therefore, between those three algorithms, the time complexity is less for the Decision Tree algorithm.

VI. RESULTS

The best algorithm for prediction of the yield of the crop and the accuracy of the crop can be found when we compare all the algorithms with each other based on the percentage of accuracy and based on the time complexity of the algorithms.

5.3 Accuracy Percentages Table:

The Table-1 below lists all of the accuracies that may be predicted using various algorithms and feature selection strategies. Three algorithms, which include Decision Tree algorithm, Random Forest algorithm, and Ada Boost algorithm, have a relatively high accuracy Percent among all the algorithms. By comparing the time complexity of these three algorithms, we can determine which one is the greatest fit. The percentage of yield prediction for several algorithms with feature selection and without feature selection strategies is shown in the accuracy table above. When yield accuracy is forecasted using multiple algorithms without using any feature selection techniques, Random Forest delivers the best yield accuracy of 100 percent.

Table -1: Accuracy Percentage for Different Algorithms with and without Feature Selection.

Algorithms	Without Feature Selection (Accuracy %)	Pearson Correlation Feature Selection Technique (Accuracy %)	Mutual Information Feature Selection Technique (Accuracy %)	Chi Square Feature Selection Technique (Accuracy %)
SVM using Polynomial Kernel	80	86.6	93.3	93.3
SVM using RBF Kernel	60	53.3	53.3	53.3
SVM using Linear Kernel	86.6	93.3	80	80
Decision Tree Algorithm	93.3	100	100	100
Random Forest Algorithm	100	100	100	100
Gaussian Naïve Bayes	73.3	60	60	60
Ada Boost Algorithm	93.3	100	100	100

When yield accuracy is projected using the Pearson Correlation feature selection approach and several algorithms, Ada Boost algorithm, Random Forest algorithm, and Decision Tree algorithm provide the highest yield accuracy of 100%. When yield accuracy is projected using the Mutual Information feature selection approach and several algorithms, Random Forest algorithm, Ada Boost algorithm, and Decision Tree algorithm provide the highest yield accuracy of 100%. When yield accuracy is projected using the Chi Square feature selection approach with several algorithms, Decision Tree algorithm, Ada Boost algorithm, and Random Forest algorithm provide the highest yield accuracy of 100%.

Retrieval Number:100.1/ijeat.B39421212222
DOI: 10.35940/ijeat.B3942.1212222
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

5.4 Time Complexity Table

The time complexity table [2] gives information about the time taken in seconds by each algorithm without Feature Selection technique and also by having different feature selection techniques.

Table 2: Time Complexity for Different Algorithms with and without Feature Selection.

Algorithms	Without Feature Selection Techniques	Pearson Correlation Feature Selection Technique	Mutual Information Feature Selection Technique	Chi Square Feature Selection Technique
SVM using Polynomial Kernel	7.15s	12.93s	11.78s	12.90s
SVM using RBF Kernel	8.34s	9.02s	13.03s	9.52s
SVM using Linear Kernel	11.08s	15.60s	12.86s	12.67s
Decision Tree Algorithm	6.44s	10.15s	8.06s	9.26s
Random Forest Algorithm	6.11s	9.72s	8.85s	9.82s
Gaussian Naïve Bayes	7.92s	10.58s	7.9s	8.58s
Ada Boost Algorithm	8.75s	9.14s	9.10s	10.22s

The Time complexities for several algorithms with and without feature selection strategies were shown in the table above. The time complexities of several algorithms were examined without using any feature selection techniques, it was found that the Random Forest approach has the best time complexity of 6.11 seconds. When the Pearson Correlation technique for feature selection is applied to multiple algorithms, the Ada Boost algorithm has very less time complexity, with 9.14 seconds, among three 100 percent accuracy predicted algorithms, namely Ada Boost algorithm, Decision Tree algorithm, and Random Forest algorithm. When the Mutual Information is used for feature selection to several algorithms, the Decision Tree has taken less time complexity of 8.06 seconds among three 100 percent accuracy predicted algorithms, namely, Decision Tree algorithm, Ada Boost algorithm, and Random Forest algorithm. When the Chi Square approach is used for Feature Selection to multiple algorithms, Decision Tree algorithm has the least time complexity of 9.26 seconds among three 100 percent accuracy predicted algorithms.

VII. CONCLUSIONS

Machine-learning approaches were used to anticipate crop production in our experiment, albeit with varied feature selection methods. The characteristics selected were based on the dataset's availability and the research's objectives. Several algorithms have been used in different studies. The findings show that there is no definitive algorithm that can be concluded as the only solution because the effectiveness of algorithms is dependent on the circumstances and conditions in which they are used, but it can be demonstrated that the improved one is better than the other algorithms in the conditions in which it succeeded. Here, Naive Bayes, SVM, Random Forest, Ada Boost, and Decision Tree models were employed. The bulk of the study compared a variety of machine learning models to determine which one could predict the most accurately. Because supervised algorithms are the most extensively used algorithms for agricultural production prediction, and wanted to see how widespread they were employed. The synthesized algorithms were extracted and used includes SVM, Decision Tree, Random Forest, Ada Boost, and Naive Bayes. However, there are varieties of algorithms that may be employed to tackle this issue. The project model became more effective because of the use of Feature Selection Techniques. Furthermore, the presence of time complexity was pivotal in the distinguishing the equal accuracy algorithms in order to select an effective one. This study can be used to pave the way for additional research on the crop yield prediction development problems by applying numerous characteristics.

REFERENCES

1. S. Veenadhari ,Dr. Bharat Mishra, Dr.CD Singh, "Soybean Productivity Modelling using Decision Tree Algorithms", International Journal of Computer Applications (0975 – 8887) Volume 27– No.7, August 2011. [CrossRef]
2. Avat Shekoofa, Yahya Emam, Navid Shekoufa, Mansour Ebrahimi, Esmaeil Ebrahimie, "Determining the Most Important Physiological and Agronomic Traits Contributing to Maize Grain Yield through Machine Learning Algorithms", May 15, 2014. [CrossRef]
3. Manoj G S, Prajwal G S, Ashoka U R, Prashant Krishna, Anitha P, "Prediction and Analysis of Crop Yield using Machine Learning Techniques", International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Published by, www.ijert.org NCAIT - 2020 Conference Proceedings.
4. Mayank Champaneri, Darpan Chachpara, Chaitanya Chandvidkar, Mansing Rathod, "Crop Yield Prediction Using Machine Learning", International Journal of Science and Research (IJSR) ISSN: 2319-7064 ResearchGate Impact Factor (2018): 0.28 | SJIF (2019): 7.583.
5. Yvette Everingham, Justin Sexton, Danielle Skocaj, Geoff Inman-Bamber , "Accurate prediction of sugarcane yield using a random forest algorithm", 19 April 2016. [CrossRef]
6. Mayank Champaneri, Darpan Chachpara, Chaitanya Chandvidkar, Mansing Rathod, "Crop Yield Prediction Using Machine Learning", International Journal of Science and Research (IJSR) 9(4 April 2020):2. [CrossRef]
7. Sujal A. Shelke, Nitin R. Chopde, "A Review on Crop Yield And Demand Predictive Using Machine Learning", 2021 IJCRT | Volume 9, Issue 2 February 2021 | ISSN: 2320-2882.
8. Su, Ying-xue, Xu, Huan, Yan, Li-jiao, "Support vector machine-based open crop model", Saudi Journal of Biological Sciences Volume 24, Issue 3, March 2017, Pages 537-547. [CrossRef]
9. W. Mupangwa, L. Chipindu, I. Nyagumbo, S. Mkuhlani, G. Sisito, "Evaluating machine learning algorithms for predicting maize yield under conservation agriculture in Eastern and Southern Africa", 24 April 2020. [CrossRef]
10. Rui Yan, Wen Jie, Jian Cao, Yong Xu, "An Optimized Naive Bayesian Method for Face Recognition", July 2017 Communications in Computer and Information Science.
11. Eibe Frank, Leonard Trigg, Geoffrey Holmes, Ian H. Witten, "Naive Bayes for Regression", October 2000.
12. S. R. Shankara Gowda, B. R. Archana, Praajna Shettigar, Kislay Kumar Satyarthi, "Sentiment Analysis of Twitter Data Using Naïve Bayes Classifier", 09 November 2021.
13. Pritika Bahad, Preeti Saxena, "Study of AdaBoost and Gradient Boosting Algorithms for Predictive Analytics", 20 December 2019.

14. Erico N. de Souza, Stan Matwin, "Improvements to AdaBoost Dynamic", Canadian Conference on Artificial Intelligence Lecture Notes in Computer Science book series (LNAI,volume 7310).
15. V Karthikeyan, S Suja Priyadharsini, "A strong hybrid AdaBoost classification algorithm for speaker recognition", 09 July 2021. [CrossRef]
16. Subhadra Mishra, Debahuti Mishra, Gour Hari Santra, "Applications of Machine Learning Techniques in Agricultural Crop Production", Indian Journal of Science and Technology, Vol 9(38), October 2016. [CrossRef]
17. Jig Han Jeong, Jonathan P. Resop, Nathaniel D. Mueller, David H. Fleisher, Kyungdahm Yun, Ethan E. Butler, Dennis J. Timlin, Kyo-Moon Shim, James S. Gerber, Vangimalla R. Reddy, Soo-Hyung Kim, "Random Forests for Global and Regional Crop Yield Predictions", June 3, 2016.
18. Vaishali Pandith, Haneet Kour, Surjeet Singh, Jatinder Manhas, Vinod Sharma, "Performance Evaluation of Machine Learning Techniques for Mustard Crop Yield Prediction from Soil Analysis", Volume 64, Issue 2, 2020 Journal of Scientific Research Institute of Science, Banaras Hindu University, Varanasi, India. [CrossRef]
19. Diego Gómez, Pablo Salvador, Julia Sanz, Jose Luis Casanova, "Potato Yield Prediction Using Machine Learning Techniques and Sentinel 2 Data", 24 July 2019. [CrossRef]
20. P.Surya, Dr. I.Laurence Aroquiaraj, "Crop Yield Prediction in Agriculture Using Data Mining Predictive Analytic Techniques", 2018 IJRAR December 2018, Volume 5, Issue 4.
21. Kaggle Data Set, Agriculture Crop Production in India https://www.kaggle.com/datasets/srinivas1/agricuture-crops-production-in-india?select=datafile+%281%29.csv

AUTHORS PROFILE

Dr. T. V. Rajini Kanth is currently working as Professor & Head, Department of CSE- AI & ML, SNIST, Hyderabad and has obtained his Ph.D. degree in C.S.E branch from Osmania University, Hyderabad in July, 2008 and M. Tech. (C.S.E.) degree from Osmania University, Hyderabad in January, 2001. His specialization area in research is "Spatial Data Mining". He obtained his PGDCS degree from HCU, Hyderabad in 1996. He received his M. Sc. (Applied mathematics) degree in the year 1989 from S.V. University, Tirupati as University Ranker. He is working as a Professor in CSE department, Sreenidhi Institute of Science and Technology (SNIST), Hyderabad. Before that, he worked as a Professor and Head, Department of IT, GRIET, Hyderabad. Before that, he worked as a Professor in CSE department, GRIET, Hyderabad since 2007 November. Prior to that, he worked as Associate Professor in VNRVJIET, Hyderabad. He joined in VNRVJIET in 1996. His total teaching experience is 22 years. His writings have appeared in numerous Professional conferences and Journals (International journals-68, national Journals-3, International conferences-42, national conferences-2) a total of 115. He was an author of a few books, i.e. Artificial Intelligence, etc. His current research area interests include Image processing, Data Warehousing & Mining, Spatial Data Mining, Web Mining, Text Mining and Robotic area, etc. Academic Society for computer systems and Information Technologies certified that his Paper titled "Design and Analysis of Novel Similarity Measure for Clustering and Classification of High Dimensional Text Documents" has been rewarded with the CRYSTAL CUBE PRIZE for best paper award of International Conference on Computer Systems and Technologies - CompSysTech'14, from Permissions@acm.org. CompSysTech'14, June27-28 2014, Ruse, Bulgaria. He has Produced 13 Ph.D. Research Scholars (5- JNTUH, 5-ANU, 3-JNTUA). He has received seminar grants from AICTE and DST organizations. He has conducted two International conferences, namely ICACT-08, and ICACM-11 at GRIET, Hyderabad and also acted as session chair for many conferences. He was the convener for Pragnya-08, a national level technical Fest at GRIET. He has also received UGC Project Grant of Rs1,10,000/- as a Principle Investigator and AICTE project grant of Rs.4,40,000/- as Co-Principle Investigator. Currently TEQUIP-III JNTUH Research Project grant of Rs3,00,000/- is going on in the Medical domain area. He is a Review Committee Member \ Editorial Board member for 18 International Journals namely IJAEGT, IJAC, JDECS, IJEAT, IJAENT, etc., also acted as session chair for many conferences like ICACT-08, ICACM-11, and ICACM-13 organized by GRIET, ICRSKT-2014, 66th CSI International Conference-2014 – Organized by JNTUH, ICCCT-2014 in association with IEEE – organized by Osmania University, 2nd World Conference on Applied Sciences, Engineering & Technology (WCSET 2013) organized by GITAM, Hyderabad. He is currently editor for two journals. He was called for about 50 AICTE sponsored workshops as resource person. In the Robo products display at 2nd Annual IT summit was appreciated by Hon. Minister Ponnala Lakshmaiah held at HICC, Hyderabad. He received seminar grants from AICTE and DST organizations. He has Life Membership in ISTE & CSI apart from a membership in IEEE. SD. He has received very prestigious award "Best Faculty of The Year 2018 Award" by CSI (Computer Society of India) at IIT Bombay. He received BEST ENGINEERING FACULTY of the Year Award 2018 across Telangana from ISTE. Apart from that he also got "Best Author of The Year Award 2017" and "Best Reviewer of the Year Award 2017" by MTRES, Delhi. He also got "Best Teacher Award" & "Best Researcher Award" from SNIST continuously for 3 years. Patent was filed and published in Official Journal of The patent Office and the title is "Autonomous Visual Tracking System and Method" and the publication date is 07/09/2018.

Burma Kavya is a BTech Graduate in Computer Science and Engineering from Sreenidhi Institute of Science and Technology in the year 2022 and also awarded the distinction. She has done project under Dr. T.V.Rajinikanth based on Machine Learning titled Wheat Crop Prediction using different machine learning algorithms. Also have a good knowledge on Artificial Intelligence based projects. Currently working at Cognizant Technology Solutions as a Programmer Analyst Trainee started in November. Also done few internships based on Machine Learning and Artificial Intelligence and also few certifications in the domain of Computer Science based programming languages like python, java and many more.

Narameta Thanuja Sri is a BTech Graduate in Computer Science and Engineering from Sreenidhi Institute of Science and Technology in the year 2022 and also awarded the distinction. She has done project under Dr. T.V.Rajinikanth based on Machine Learning titled Wheat Crop Prediction using different machine learning algorithms. Also have a good knowledge on Artificial Intelligence based projects. Currently doing Masters in Data Science at Illinois Institute of Technology first semester at Chicago. Also done few internships based on Machine Learning and Artificial Intelligence and also few certifications in the domain of Computer Science based programming languages like python, java and many more.

Alley Yashwanth Saikrishna is a BTech Graduate in Computer Science and Engineering from Sreenidhi Institute of Science and Technology in the year 2022 and also awarded the distinction. He has done project under Dr. T.V.Rajinikanth based on Machine Learning titled Wheat Crop Prediction using different machine learning algorithms. Also have a good knowledge on Artificial Intelligence based projects. Currently working at Accenture as Associate Software Engineer started in October. Also done few internships based on Machine Learning and Artificial Intelligence and also few certifications in the domain of Computer Science based programming languages like python, java and many more.

Retrieval Number:100.1/ijeat.B39421212222
DOI: 10.35940/ijeat.B3942.1212222
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)