International Journal of Multidisciplinary Research and Growth Evaluation.

Crop yield prediction using machine learning techniques

Krupashini S [1*], Mahalakshmi PR [2], Sangavi C [3], Dr. R Manivannan [4]
[1-3] Computer Science and Engineering, E.G.S. Pillay Engineering College, Nagapattinam, Tamil Nadu, India
[4] Ph.D., Computer Science and Engineering, E.G.S. Pillay Engineering College, Nagapattinam, Tamil Nadu, India

* Corresponding Author: **Krupashini S**

Article Info

ISSN (online): 2582-7138
Volume: 04
Issue: 03
May-June 2023
Received: 25-03-2023;
Accepted: 18-04-2023
Page No: 37-43

Abstract
India is an imperial nation, in which agriculture is the primitive profession. Indian economic stability depends on agricultural wealth. Crop yield prediction in agriculture is critical and is chiefly depend upon soil and environment conditions, including rainfall, humidity, and temperature. Ancient farmers were able to decide on the crop to be cultivated, monitor its growth, and determine when it could be harvested. Rapid changes in environmental conditions have made it difficult for the farming community to continue to do so. Thus, agriculture is highly dependent on the new technology for obtaining large profits. Machine learning techniques have taken over the task of prediction in recent times and this work has used several of these to determine crop yield. To ensure that a given machine learning (ML) model works at a high level of precision, it is mandatory to employ efficient feature selection methods to pre-process the raw data into an easily computable Machine Learning friendly dataset. To reduce redundancies and make the ML model more accurate, only data features that have a significant degree of relevance in determining the final output of the model must be employed. Optimal feature selection arises to ensure that only the most relevant features are accepted as a part of the model. The results depict that an ensemble technique offers better prediction accuracy.

Keywords: Crop yield prediction, environmental condition, ML- Machine Learning, feature selection

1. Introduction

Crop prediction in agriculture is a complex process and multiple models have been proposed and tested this to end. The Problem make use of assorted datasets, given that crop cultivation depends on biotic and abiotic factors [16]. Biotic factors include those elements of environment that can occur as a result of the impact of living organisms (micro-organisms, plants, animals, pests, parasites), directly or indirectly on other living organisms. This group also includes environmental changes caused or influenced by people (fertilization, air pollution, water pollution, soil, etc.). These factors may cause many changes in the yield of crops, cause internal defects, shape defects and change in the chemical composition of the plant yield. The shaping of the environment as well as the quality and growth of plant is influenced by biotic and abiotic factors. Abiotic factors can be divided into physical chemical and other. The recognized physical factors include: radiations, climatic conditions (atmospheric pressure, temperature, humidity, air direction, sunlight); soil type, topography, soil rockiness, atmosphere, and water chemistry, especially salinity. Abiotic factors also include bedrock, relief, climate, and water conditions which affect its properties. Soil-forming factors have a varied effect on the formation of soils and their agricultural value [12].

Predicting crops yield is not an easy task. The procedure for predicting the area under cultivation is, according to Myers *et al.* [8] and Muriithi [2], a set of statistical and mathematical techniques useful in an evolving and improving optimization process. It also has important uses in design, development, and formulation new as well as improving existing products. Presentation or performance of statistical analysis requires the possession of numerical data. Based on them, conclusions are drawn as to various phenomena and further, on this basis, binding economic decisions can be made.

According to Muriithi [2], the better you describe certain phenomena in terms of numbers, the more you can say about them, and with increasing data accuracy you can also obtain more accurate information and make more accurate decisions.

The biggest problem in the temperate climate zone is assessment of agroclimatic factors in terms of shaping the yield of winter plant species, especially cereals. The key factor influencing wintering yield, which provides access to days with a temperature over of 5◦C, their number and frequency, and the number of days in the wintering period with temperatures above 0◦C and 5◦C. A number of these can be estimated on the basis of public statistics and yield regression statistics in years. Developed models for checking the situation that assess whether they want to be a probation of state policy in the field of intervention in the cereal market. Efficient forecasting of productivity requires forecasting of agrometeorological factors. Aspects related to the variability of these factors may pose a particular problem [9]. Grabowska *et al.* [11] predicted narrow-leaf lupine yields for 2050-2060 using weather models and three climate change scenarios for Central Europe: E-GISS model, HadCM3 and GFDL. The fit of the models was assessed by means of the determination coefficient R2, corrected coefficient of determination R2adj, standard error of estimation and the coefficient of determination R2pred calculated using the Cross Validation procedure. The selected equation was used to forecast lupine yield under the conditions of doubling the CO2 content in the atmosphere. These authors stated that the influence of meteorological factors on the yield of narrow-leaved lupine varied depending on the location of the station. The temperature (maximum, average, minimum) at the beginning of the growing season, as well as rainfall during the flowering - technical maturity period, most often had a significant influence on the yield. It has been shown that the predicted climate changes will have a positive effect on the lupine yield. The simulated profitability was higher than that observed in 1990-2008, and HadCM3 was the most favorable scenario.

Li *et al*. [4] found that accurate, high-resolution yield maps are needed to identify spatial patterns of yield variability, to identify key factors influencing yield variability, and to provide detailed management information in precision farming. Varietal differences may significantly affect the forecasting of potato tuber yields with the use of remote sensing technologies. These authors argue that improving potato crop forecasting with remote sensing of Unmanned Aerial Vehicles (UAVs) by incorporating varietal information into machine learning methods that has the best chance at present. There are different challenges in this research area. At present, crop prediction [6] models generate actual results that are satisfactory, though they could perform better. This paper attempts to propose an improved crop prediction model that addresses these issues. The prediction process [10] depends on the two fundamental techniques of feature selection [FS] and classification. Prior to the application of FS techniques, sampling techniques are applied to balance an imbalanced dataset.

2. Related Works

Kodimalar Palanivel and Chellammal Suriyanarayanan [13], have been investigated on how various machine learning algorithms are useful in the prediction of crop yield and proposed an approach for prediction of crop yield using machine learning techniques in big data computing paradigm. An extensive study on the crop prediction is made. The literatures reveal diverse machine learning techniques adopted for prediction of crop yield. Further, the performance metrics of the machine learning algorithms such as root mean square error are studies. Along with machine learning algorithms for prediction, it is planned to study the impact of big data techniques in the prediction of crop yield. A conceptual approach is proposed for the same. Her proposed solution is much efficient to train the dataset.

Farhat Abbas, Hassan Afzaal *et al.* [14] experimented on the potential of four ML algorithms, namely Linear

Regression, Elastic Net, k-Nearest Neighbor and Support Vector Regression for the prediction of potato tuber yield from data of soil and crop properties collected through proximal sensing for datasets of six fields across Atlantic Canada. For the growing seasons of 2017 and 2018, the data about horizontal and vertical components of soil electrical conductivity, soil moisture content, field slope, soil pH, SOM, normalized difference vegetative index, and potato tuber yield were named as PE-2017, PE- 2018, NB-2017 and NB-2018 for Prince Edward Island and New Brunswick fields. Modeling techniques were employed to generate yield predictions with statistical parameters from the collected data. The performance of k- NN remained poor except for PE-2018. However, all ML algorithms worked well by explaining about 60% of the tuber yield from the soil properties mentioned above. The remaining 40% explanation may come from external factors, such as climate change and environment. Furthermore, larger datasets may generate precise and accurate results using either model. The information generated from this study will be needed for creating site- specific management zones for potatoes, which form a major component for food security initiatives across the globe. However, the proposed solution is adaptable and fault-tolerant but it doesn't work with the large datasets.

Mohsen Shahhossein, Isaiah Huber *et al.* [3] investigated whether the coupling crop modeling and machine learning improves corn yield in the US Corn Belt. Their objectives are to explore whether a hybrid approach would result in better predictions, investigate which combinations of hybrid models provide the most accurate predictions and examine the features from the crop modeling that are more effective to be integrated with ML for corn yield prediction. They have designed five ML models (LASSO, Linear Regression, Random Forest, LightGBM and XGBoost) and six ensemble models to address the research question. The results suggest the coupling model can decrease yield prediction. So, they demonstrated improvements in yield prediction accuracy across all designed ML models when additional inputs from a simulation cropping systems model (APSIM) are included. Among several crop model (APSIM in this study) variables that can be used as inputs to ML, analysis suggested that the most important ones were those related to soil water, and in particular growing season average drought stress, and average depth to water table. They concluded that inclusion of additional soil water related variables could further improve ML yield prediction in the central US Corn Belt.

Dilli Paudel, Hendrik Boogard, Allard de Wit *et al* [15] made a workflow for crop yield prediction using MCYFS data and they evaluated the workflow by predicting crop yield at NUTS2 or NUTS3 levels for five crops and three countries. They designed a modular and reusable ML workflow for the prediction of crop yield and tested the workflow on thirteen case studies. They found that explainable features designed using principles of crop modeling can be used to predict crop yield at sub-national level. For early season predictions, the ML baseline performed similar to MARS Crop Yield Forecasting System (MCYFS) in most cases. There was room for improvement as the season progressed. For crops and countries where regional data is reliable, sub-national yield prediction using machine learning is a promising approach going forward. Apart from addressing data quality issues, the baseline could be improved in three main ways: adding new data sources, designing more predictive features and evaluating different algorithms. The machine learning baseline serves as a starting point to explore the potential of machine learning for large-scale crop yield forecasting.

3. Working Methodology

Predicting crop yield will be very much useful for the farmers. The main objective of the proposed solution is to provide accurate prediction and to reduce the loss of cost. The crop yield is mainly depending upon the factors such as weather condition, soil type, temperature, rainfall and pesticides. This prediction is proportional to the accuracy on dataset provided. This proposed system predicts the crop yield with high accuracy and helps in reducing the loss.

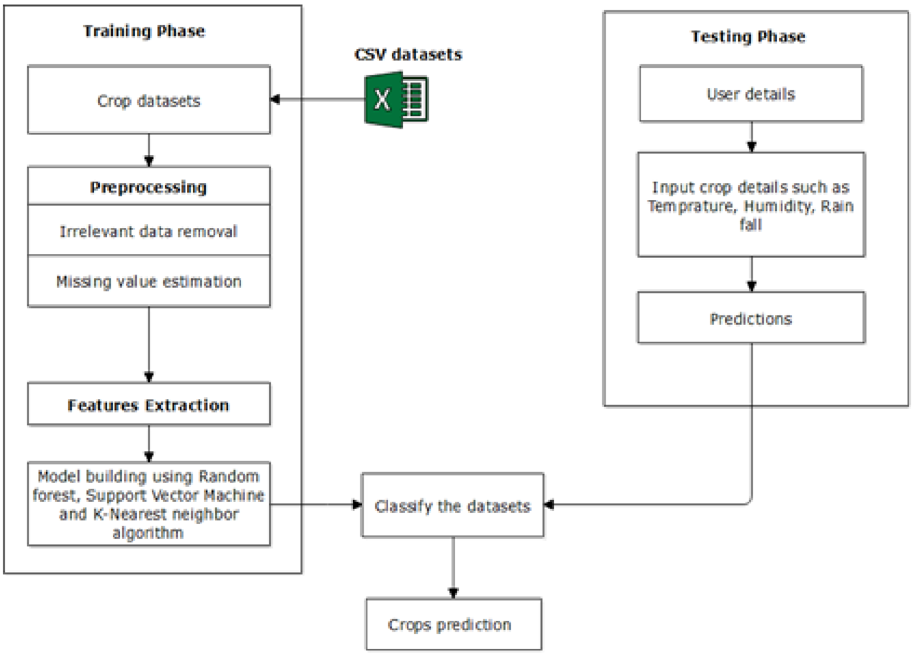


Fig 1: Outline of Proposed System

The forecasted system acts as experienced farmer. It provides high accuracy by considering many other factors. The more increase in accuracy results in more profit in crop yield. To increase accuracy the data has to be perfect. With all the provided information, the proposed system extracts necessary features from dataset and process all the data using ML techniques and predict the crop yield. With this prediction the farmers will be able to know their requirements and precautionary measures can be taken to prevent loss of crops.

4. Implementation

A. K-Nearest Neighbor Algorithm

KNN algorithm is one of the most commonly used non-parametric and supervised machine learning techniques, which is used in regression and classification problems. Supervised machine learning relies on labelled input and output training data. Supervised learning algorithms take the data and make models that predict the output data given suitable inputs.

The algorithm hangs on distances between points, which can be establish using one of a few methods. A key aspect for consideration is that the distance is always required to be either 0 or positive. This is done by squaring the distance or raising it to a certain power or taking the absolute values.

Methods to find distance include the following:

i. Manhattan Distance

$$\text{Manhattan Distance} = d(x,y) = \left(\sum_{i=1}^{m} |x_i - y_i|\right) \tag{1}$$

ii. Euclidean Distance

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (y_i - x_i)^2} \tag{2}$$

iii. Hamming Distance

$$\text{Hamming Distance} = D_H = \left(\sum_{i=1}^{k} |x_i - y_i|\right) \qquad \begin{matrix} x=y & D=0 \\ x\neq y & D\neq 1 \end{matrix} \tag{3}$$

iv. Minkowski Distance

$$\text{Minkowski Distance} = \left(\sum_{i=1}^{n} |x_i - y_i|\right)^{1/p} \tag{4}$$

KNN works on the premise of similar entities existing in close proximity. Related data fields would therefore occur nearby. This helps us in mapping similarities between datasets and a given query.

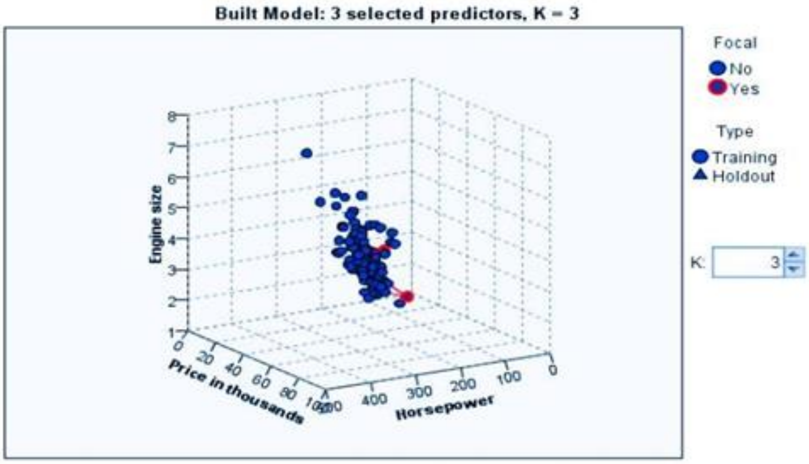


Fig 2: Projection of feature using KNN, when k=3 for Crop yield dataset

Before implementing the KNN algorithm, all the labeled data must be pre-processed. First of all the data must be normalized. Then, feature selection must be employed to delete the unrelated features as KNN doesn't work well with too many features. Missing values is not supported, thus in the case of missing values that particular row must be deleted. Moving towards the implementation of the KNN algorithm, data is loaded into the model. KNN method requires data to be loaded in labeled form. Then, K is declared according to the desired number of neighbours. Next, for every element in the dataset, the ''distance'' or ''relation'' with the query input is calculated by the machine learning algorithm. The distance between the element and the query input is then added to an ordered collection and is subsequently sorted in increasing order of the distances. Finally, the first K items of the collection are selected and the output, depending upon the model being a regression or a classification problem. Choosing k is an important factor as it heavily influences the result of our ML model. If the value of K is too low, the model will not be stable and the results become highly inaccurate. Conversely, an extremely high value of k will increase the number of errors in the model. Therefore, the value of k must be balanced between the two extremums. In the case of a model where a vote is required to get the output, K should be taken as an odd number to ensure a deciding game.

The chief advantages of the KNN algorithm is, it is simple and relatively easy to implement. The algorithm can serve multiple purposes, right from classification and regression to searching problems. Further, the algorithm can be improved by adding additional training data.

The main disadvantage of KNN is that the speed of the algorithm goes on decreasing for large dataset as the cost of computation keeps increasing. Therefore, it is not suitable in cases where immediate results are required. The value of k must be accurately determined to get appropriate results. This process can be difficult sometimes. Also, the KNN algorithm requires a large amount of memory to store large datasets. The KNN method can be used for predicting crop yield using a set of known factors such as rainfall, humidity, temperature and soil moisture. The value of crop yield is calculated by using the values of the nearest neighbors. KNN has yielded suitable accuracy in predicting crop yield. This model can be enhanced by adding additional features and more data from all the seasons. KNN has also been applied for predictive analysis of paddy production.

B. Support Vector Machine (SVM)

SVM is a supervised machine learning algorithm which is very useful technique for data classification. However, this learning algorithm can also be used for regression challenges. A classification task usually involves separating data into training and testing sets. Each instance in the training set contains one target value (i.e. the class labels) and several attributes (i.e. the features or observed variables).

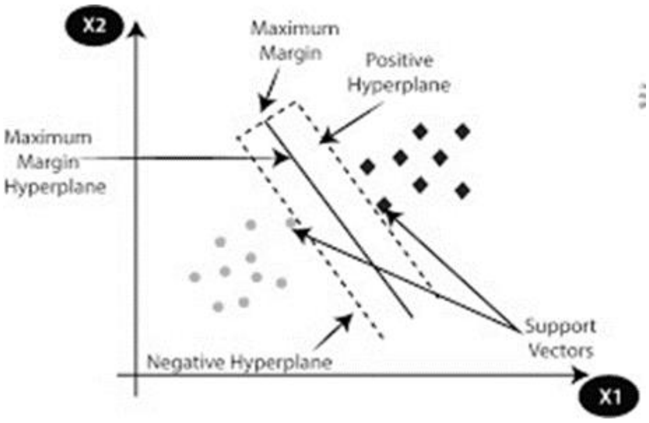


Fig 3: Concept of SVM

SVM works well on small data sets but is more efficient with large data sets. Given a dataset with n features, SVM initiates with plotting all points in the dataset in n- dimensional space, and each point is assigned a coordinate according to the value of its features. Hereon, the classification process is conducted by determining a suitable hyperplane which to the furthest extent, differentiates the points into two distinct classes. Support vectors are essentially the points that are located close to the hyperplane and determine its position and orientation. The distance between the support vectors and the hyperplane is called the margin and to generate the most accurate hyperplane, the margin needs to be maximized as far as possible.

The advantages of the SVM algorithm are, primarily it is very effective in analysing high dimensional datasets. It is of great use in cases where the number of dimensions is greater than the number of samples. SVM utilizes the support vectors for training and therefore consumes less memory.

Few disadvantages of the SVM algorithm are, it is not suitable for very large datasets as the time required to train the model increases. It also gives inaccuracies when the target classes overlap with each other. Moreover, the SVM algorithm cannot account for probability. SVM is used to classify agricultural data to allow for better decision- making. In a comparative study of classification techniques used for agricultural data, SVM was able to outperform Naïve Bayes and Artificial Neural Network methods.

To decide how to measure the importance of accuracy, as small residuals may be inevitable even need to avoid large ones. The loss function determines this measure. Support vector regression performs linear regression in the feature space using Ɛ-insensitive loss function.

$$L_\varepsilon(y, f(\mathbf{x}, \omega)) = \begin{cases} 0 & if\,|y - f(\mathbf{x}, \omega)| \le \varepsilon \\ |y - f(\mathbf{x}, \omega)| - \varepsilon & otherwise \end{cases}$$

The empirical risk is:

$$R_{emp}(\omega) = \frac{1}{n}\sum_{i=1}^{n} L_\varepsilon(y_i, f(\mathbf{x}_i, \omega)) \qquad (5)$$

It is well known that SVM generalization performance depends on a good setting of meta parameters C, Ɛ and the kernel parameters. Selecting a particular kernel type and kernel function parameters is usually based on application domain knowledge and also should reflect distribution of input (x) values of the training data. Parameter C determines the trade-off between the model complexity and the degree to which deviations larger than Ɛ are tolerated in optimization formulation.

C. Random Forest ML Algorithm

Random Forest (RF) is one of the most successful supervised machine learning algorithms. The RF algorithm embodies the essence of ensemble learning in that it links multiple classifiers to resolve a complicated problem, thereby enhancing the performance of the model. In this method, the ‘‘forest’’ that is built is a set of decision trees. Characteristics in the RF are randomly picked in each decision split. The correlation between trees is diminished by randomly picking features that promote prediction and result in greater efficiency. Random Forest is an ML classification algorithm that works by dividing the dataset into subsets or decision trees and computing the outputs of all the trees to produce the final output. Random Forest comes under the Bagging category of ensemble learning techniques. The Row and Feature samples from the main dataset are randomly selected and fed into the decision trees in the Random Forest Technique. The analyst chooses the number of decision trees for the model. Each decision tree works on the data and predicts a result based upon its calculation. Random Forest doesn’t take the result from any one of the decision trees but combines the outputs from all the decision trees. Random Forest takes the majority of the result or the mean/median of the result. Thus, a higher number of decision trees gives a more accurate result and circumvents the problem of overfitting. The number of trees is proportional to accuracy in prediction. The dataset includes factors like rainfall, perception, temperature and production. These factors in dataset is used for training. Only two-third of the dataset is considered. Remaining dataset is used for experimental basis [21]. The Random Forest technique provides many advantages. RF is simple and relatively easy to understand and is therefore extremely popular. It is capable of performing both classification and regression tasks. It is also suitable for handling large sets of data with high dimensionality and most importantly, it makes the model much more precise and resolves the overfitting issue. Random Forest cannot be used in case of extrapolation of data as it could produce inaccurate results. Also, it does not produce proper results when dealing with sparse data. Random Forest also needs more time for implementation and requires larger data and greater resources. In the presence of correlated predictors, Random Forest is known to produce inexact results. Random Forest can be used to predict pest attacks in cotton plants. Various factors were very considered and the Correlation filter selection method was used to select the most important features. Random Forest was then used to determine the number of trees to get a low error rate and important parameters were sighted out and used for clustering to determine the outcome. Optimized usage of water for farms is essential for reducing wastage as well as enhancing productivity. The use of a precision irrigation system for furnishing the optimal water supply needed by plants or crops will lead to better output. The amount of water required by plants can be expressed in terms of pH. The Random Forest algorithm is used to determine the pH level which in turn helps determine the water supply required by a piece of land.

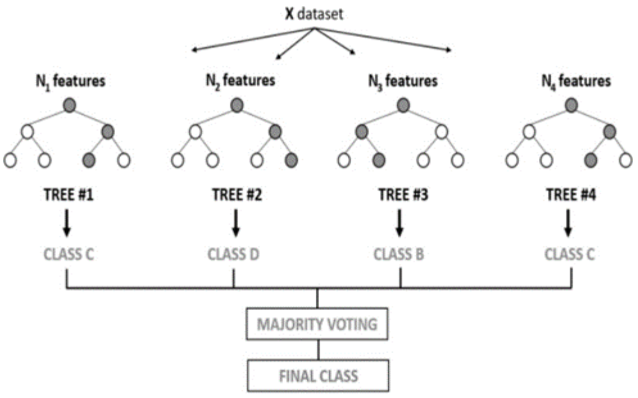


Fig 4: Concept of RF

Given that each bagged tree is identically disseminated, the expectation of an average of B trees is the equivalent of the expectation of each. Since this accounts for the bias of bagged trees being the same as that of individual trees, a change may only be affected through variance reduction. This contrasts with advancing, where the trees are grown adaptively to exclude bias, and hence are not identically distributed. An average of B identically distributed random variables has a variance of σ^2. If the variables are completely identically distributed, but with positive pairwise correlation ρ, the variance of the average is given as

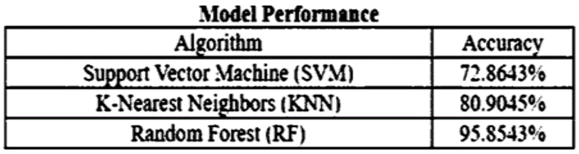
$$\rho\sigma^2 + \frac{1-\rho}{B}\sigma^2 \qquad (6)$$

It is observed that as B increases, the value of the second term shifts negligibly while that of the first term remains unchanged. Consequently, the size of the correlation of the bagged trees limits the benefits of averaging. The RF focuses on bagging variance minimization by cutting the correlation between the trees without increasing the variance excessively. The tree-growing process makes this procedure possible through picking input variables at random.

5. Result Analysis

The below table represents the accuracy rate provided by the Support Vector Machine, KNN and Random Forest algorithm.

Table 1



Model Performance

Algorithm	Accuracy
Support Vector Machine (SVM)	72.8643%
K-Nearest Neighbors (KNN)	80.9045%
Random Forest (RF)	95.8543%

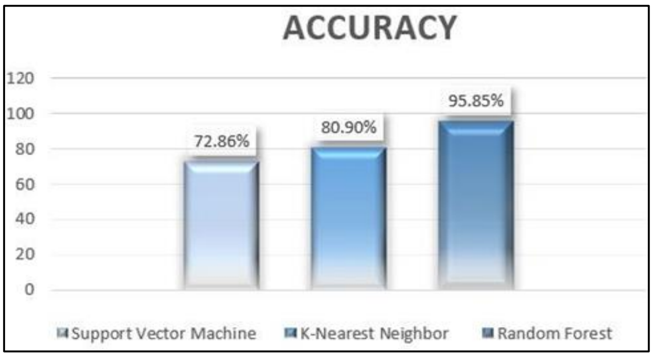


Fig 5

From the table, it is analyzed that Support Vector Machine produced least amount of 72% of accuracy among the three algorithms. The K-Nearest Neighbor algorithm produced 80% of accuracy which is comparatively acceptable as compared to SVM algorithm. It is analyzed that the Random Forest (RF), has produced around 96% of accuracy, which is highest among these three algorithms.

6. Conclusion

The proposed solution for predicting crop yield based on the factors such as soil type, rainfall, temperature, pesticides etc., uses various machine learning algorithms. Experiments were conducted on datasets taken from open source and it has been analyzed that the Random Forest Algorithm produces the highest yield prediction accuracy. It will greatly help farmers in maintaining the right crop supply to grow and trigger them to take precautionary measures to prevent from huge loss of crops, thus, it also helps in cost management.

7. Future Enhancement

The research work can be build up to the higher level by building a recommender system of agriculture production and distribution for farmer. Through this farmer can make their own decision like, for which season which crop should sow so that they can get better gain. The proposed system works for structured dataset or database. In upcoming years try applying data independent system also that mean as the format may be whatever, our system should work with same accuracy.

8. References

1. B Sawicka, AH Noaema, A Gáowacka. The predicting the size of the potato acreage as a raw material for bioethanol production, in Alternative Energy Sources, B. Zdunek, M. Olszáwka, Eds. Lublin, Poland: Wydawnictwo Naukowe TYGIEL, 2016, pp. 158–172.
2. K Muriithi. Application of response surface methodology for optimization of potato tuber yield,'' Amer. J Theor. Appl. Statist. 2015; 4(4):300-304. doi: 10.11648/j.ajtas.20150404.20.
3. Shahhosseini Mohsen, Hu Guiping, Huber Isaiah, Archontoulis Sotirios. Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. Scientific Reports, 2021. 11. 10.1038/s41598-020-80820-1.
4. D Li, Y Miao, SK Gupta, CJ Rosen, F Yuan, C Wang, L Wang, Y Huang. Improving potato yield prediction by combining cultivar information and UAV remote sensing data using machine learning, Remote Sens. 2021; 13(16):3322. doi: 10.3390/rs13163322.
5. JR Olędzki. The report on the state of remotesensing in Poland in 2011-2014, (in Polish), Remote Sens. Environ. 2015; 53(2):113-174.
6. N Chanamarn, K Tamee, P Sittidech. Stacking technique for academic achievement prediction, in Proc. Int. Workshop Smart Info-Media Syst, 2016, 14-17.
7. R Jahan. Applying naive Bayes classification technique for classification of improved agricultural land soils,'' Int. J Res. Appl. Sci. Eng. Technol. 2018; 6(5):189-193.
8. RH Myers, DC Montgomery, GG Vining, CM Borror, SM Kowalski. Response surface methodology: A retrospective and literature survey, J Qual. Technol. 2004; 36(1):53-77.
9. M Marenych, O Verevska, A Kalinichenko, M Dacko. Assessment of the impact of weather conditions on the yield of winter wheat in Ukraine in terms of regional, Assoc. Agricult. Agribusiness Econ. Ann. Sci. 2014; 16(2):183-188.
10. W Paja, K Pancerz, P Grochowalski. Generational feature elimination and some other ranking feature selection methods, in Advances in Feature Selection for Data and Pattern Recognition, vol. 138. Cham, Switzerland: Springer, 2018, 97-112.
11. K Grabowska, A Dymerska, K Poáarska, J Grabowski. Predicting of blue lupine yields based on the selected climate change scenarios, Acta Agroph. 2016; 23(3):363-380.
12. B Sawicka, AH Noaema, TS Hameed, B Krochmal-Marczak. Biotic and abiotic factors influencing on the environment and growth of plants, (in Polish), in Proc. Bioróżnorodność Środowiska Znacze nie, Problemy, Wyzwania. Materiały Konferencyjne, Puławy, 2017.
13. Kodimalar Palanivel, Chellammal Surianarayanan. An approach for Prediction of Crop Yield Using Machine Learning and Big Data Techniques, Interanational Journal of Computer Engineering and Technology. 2019; 10(3):110-118.
14. Abbas Farhat, Afzaal Hassan, Farooque Aitazaz, Tang Skylar. Crop Yield Prediction through Proximal Sensing and Machine Learning Algorithms, 2020, Agronomy. 10. 1046. 10.3390/agronomy10071046.
15. Dilli Paudel, Hendrik Boogaard, Allard de Wit, Sander Janssen, Sjoukje Osinga, Christos Pylianidis, Ioannis N. Athanasiadis, Machine learning for large-scale crop yield forecasting, Agricultural Systems, 2021, 187. 103016, ISSN 0308- 521X.
16. BB Sawicka, B Krochmal-Marczak. Biotic components influencing the yield and quality of potato tubers, Herbalism. 2017; 1(3):125-136.
17. N Rale, R Solanki, D Bein, J Andro-Vasko, W Bein. Prediction of Crop Cultivation, in Proc. 19th Annu. Comput. Commun. Workshop Conf. (CCWC), Las

Vegas, NV, USA, 2019, 227-232.

18. DK Bolton, MA Friedl. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics, Agricult. Forest Meteorol. 2013; 173:74-84.
19. E Manjula, S Djodiltachoumy. A model for prediction of crop yield,'' Int. J Comput. Intell. Inform. 2017; 6(4):298-305.
20. G Mariammal, A Suruliandi, SP Raja, E Poongothai. Prediction of land suitability for crop cultivation based on soil and environmental characteristics using modified recursive feature elimination technique with various classifiers, IEEE Trans. Computat. Social Syst. 2021; 8(5):1132-1142.
21. Y Jeevan Nagendra Kumar, B Mani Sai, Varagiri Shailaja, Singanamalli Renuka, Bharathi Panduri. Python NLT K Sentiment Inspection using Naïve Bayes Classifier International Journal of Recent Technology and Engineering, ISSN: 2277-3878, Volume-8, Issue-2S11, Sep 2019.