International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7431
***Article Received**: 10 May 2023 **Revised:** 30 June 2023 **Accepted:** 29 July 2023*

Comparison of Classification Algorithm for Crop Decision based on Environmental Factors using Machine Learning

Guruprasad Deshpande[1], Rajani P.K.[2], Vishal Khandagle[3], Jayesh Kolhe[4]
[1]Department of Electronics and Telecommunication,
Pimpri Chinchwad College of Engineering,
Pune, India
e-mail: [1]guruprasad.deshpande19@pccoepune.org
[2]Department of Electronics and Telecommunication,
Pimpri Chinchwad College of Engineering,
Pune, India
e-mail: [2] rajani.pk@pccoepune.org
[3]Department of Electronics and Telecommunication,
Pimpri Chinchwad College of Engineering,
Pune, India
e-mail: [3] vishal.khandagale19@pccoepune.org
[4]Department of Electronics and Telecommunication,
Pimpri Chinchwad College of Engineering,
Pune, India
e-mail: [4] jayesh.kolhe19@pccoepune.org

Abstract. Crop decision is a very complex process. In Agriculture it plays a vital role. Various biotic and abiotic factors affect this decision. Some crucial Environmental factors are Nitrogen Phosphorus, Potassium, pH, Temperature, Humidity, Rainfall. Machine Learning Algorithm can perfectly predict the crop necessary for this environmental condition. Various algorithms and model are used for this process such as feature selection, data cleaning, Training, and testing split etc. Algorithms such as Logistic regression, Decision Tree, Support vector machine, K-Nearest Neighbour, Navies Bayes, Random Forest. A comparison based on the accuracy parameter is presented in this paper along with various training and testing split for optimal choice of best algorithm. This comparison is done on two tools i.e., on Google collab using python and its libraries for implementation of Machine Learning Algorithm and WEKA which is a pre-processing tool to compare various algorithm of machine learning.

Keywords: Crop Decision, Machine Learning, Algorithms, Google Collaboratory, WEKA, Accuracy, Comparison.

I. INTRODUCTION

Crop production, water management, fertilizer application, pest and disease management, harvesting, transportation of food products, packing, food preservation, food safety, food storage, and food marketing are the primary stages of agricultural products. Thanks to the Internet of Things (IoT),[1] farmers may now have a global impact on technology and make well-informed decisions that will increase production, improve resource management, and produce the desired outcomes with the least amount of waste. With IoT-based devices, technology supports our farmers throughout the entire farming process and even after. While many of them borrow money to cultivate a crop in the hopes of making a profit, the most important choice that farmers must make is usually made at random. This choice is made without much analysis and based on extremely general considerations. If the crop he chooses has the right conditions,[2] the right demand, and government assistance, his decision could boost the farmer's revenue. As a result, we have suggested a platform that can give the farmer a suitable option. Aid them in recording their profit and understanding the effectiveness of their soil. The market and farmers will benefit from this platform's ability to produce an excess amount of a certain crop.[3] The platform would assist farmers in planning their harvest under optimal weather circumstances, together with crop efficiency and demand. [11] Additionally, this platform will shield them from losses brought on by bad weather or natural calamities.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7431
***Article Received**: 10 May 2023 **Revised:** 30 June 2023 **Accepted:** 29 July 2023*

II. TOOLS AND TEACHNOLOGY

The implementation of algorithms and the program is done using following tools and technologies.

A. Google Collab

Google Collaboratory (also known as Collab) is a Jupyter-based cloud platform for sharing machine learning research and instruction. It offers free access to a potent GPU and a runtime that is fully functional for machine learning and deep learning. This work provides a thorough review of Collaboratory's hardware capabilities, performance, and restrictions. Collab supports several well-known machine learning libraries that are simple to load in your notebook.

B. Python

All ML codes are written using it. Python is a general, high-level programming. Its design philosophy prioritises readability and makes extensive use of indentation. The reason of choosing this language than other coding language is that it provides a variety of libraries and supports interfacing of various technologies easily. The libraries used during Implementation of Various learning are scikit learn which provides access to machine learning algorithm and other libraries like pandas, NumPy and matplotlib for analysing and visualizing the data and doing required operation in order to make the data ready for machine learning algorithm.

C. Weka

WEKA is an open-source piece of software (Waikato Environment for Knowledge Analysis). It has tools for preprocessing data, machine learning algorithms implemented, and visualization tools. These methods may be used with a variety of real-world data. Raw data collected on the ground serves as the foundation. Fields with null values and fields unrelated to the data collection may be present in this data. Data preparation tools can be used to clean the data. The pre-processed data may be kept and used to apply machine learning algorithms on in a local storage. Depending based on the type of machine learning model being constructed, a choice must be made amongst alternatives like Classify, Associate, and Cluster. Numerous Attributes Selection techniques are implemented by WEKA, allowing for the automated selection of traits to create a unique reduced dataset for category.

III. METHODOLOGY

Loading data loading the data set available in the model is the first we can use the cloud as well for this step and the main aim of this step is to get data and do an initial summarizing of data for further processing to get simpler to summarize the data, we would be using parameters such as mean variable variance entropy and plot data for summarizing the process.[4] The data will be segregated in the respective variable for making calculation easier and apply our algorithms to this segregated data. Denoising as it helps to produce high accuracy, denoising is used to clean the data by removing all of the null and garbage values from the dataset. Splitting Dataset to Train and Test To test the model, we will use some of the training data, and the remaining data will be used to test model. We can also use this data to check our accuracy.[5] Feature scaling method for uniformizing the independent characteristics that are present in the data within a given range. For accurate prediction, this technique is essential.

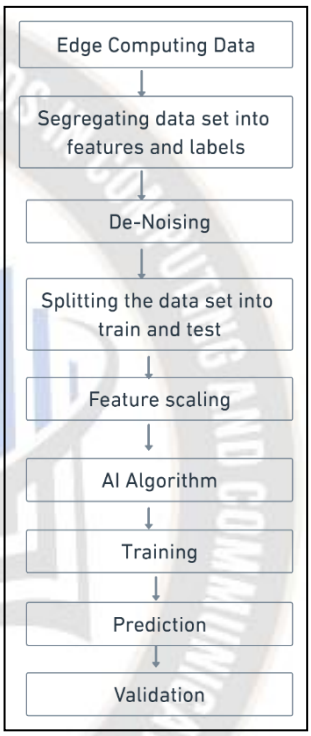

Figure 1. Flow Chart

AI Algorithm In this step, we would be applying different algorithms for different factors. Such as SVC, KNN, etc. classification algorithms in location. We would be using a Logistic regression algorithm and other algorithms and other algorithms for other factors such as Market, Rainfall, and other factors respectively. Training after applying the algorithms, we need to train our data for different datasets and prepare it for prediction and test it with the remaining data to validate as well the prediction. This is done by our pre-processed data of previous data[20]. Prediction In this step, we would be taking input from the user to predict a Demand-oriented, suitable crop from its location, and considering the factor we also need to predict the results for the same using our above model. This is a crucial step of the whole process. Validations Checking the accuracy is very much important for further improvement.[6] We would validate our data on certain parameters.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7431
***Article Received**: 10 May 2023 **Revised:** 30 June 2023 **Accepted:** 29 July 2023*

A. NAÏVE BAYES (NB)

In a variety of classification tasks, the Naive Bayes algorithm, which is based on Bayes' theorem, is commonly used. The multinomial, Gaussian and Bernoulli algorithms make up the three Naive Bayes algorithms. In Equation 1, it is depicted.

$$P(Y \mid X) = \frac{P(X \mid Y)\, P(Y)}{P(X)} \qquad (1)$$

where P (y | X) = Posterior probability

P (X | y) = Likelihood

P(X) = Evidence

P(y) = Prior probability.

A supervised machine learning method called the Naive Bayes is mostly employed for classification issues. It operates under the presumption that each feature contributes equally to the result and that the likelihood of occurrence of any feature is independent of the probability of occurrence of other features. The theorem, which determines the likelihood of events given the occurrence of another event, forms the basis of this formula. The theorem attempts to determine the likelihood of an event occurring given the truth of another event. The Nave Bayes algorithm uses a probabilistic classifier, which means it makes predictions based on likelihood. The Naive Bayes Algorithm would compute the outcome based on probability given a labelled dataset and a target variable. The complete dataset is first pre-processed, with the events and their frequencies recorded, and then categorized into a frequency table. The frequency table is then used to create a likelihood table. Finally, the posterior probability is computed using the Bayes theorem. The Naive Bayes Algorithm has several benefits, including the ability to classify data into binary and many classes. Second, in comparison to other ML techniques, it is quicker and simpler to construct. Additionally, it does not need a lot of training data. Both discrete and continuous data can be used with it. It is highly scalable and unaffected by superfluous features. When the independence assumption is valid, the Naive Bayes algorithm outperforms other algorithms. The Nave Bayes method can be used in agriculture to recommend profitable food crops. Farmers would benefit greatly from this advice on food crops, especially in the face of climate change. A better selection of food crops would increase farmer income while lowering the likelihood of crop choice failures.

B. DECISION TREES

A decision tree is a type of structure tree that resembles a flowchart and is frequently employed in supervised machine learning for classification. Every path leading from the root node to every leaf node in a DT can be converted into a set of rules, each of which is a rule. Every leaf node in a decision tree has a class assigned to it that is reachable if the attribute meets the condition of the branch that leads to it. Every internal node in a decision tree represents a test/condition or an attribute. Based on the different target attribute types, decision trees can be broadly divided into categorical and continuous variable types.

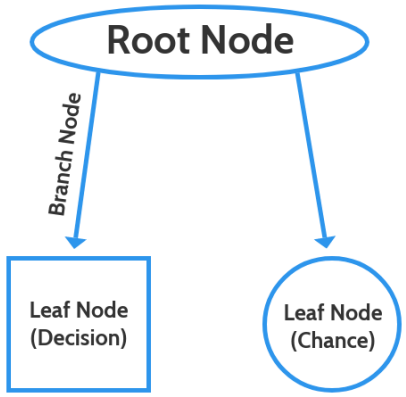

Figure 2. Decision tree

An ideal split is achieved by comparing the root node of a decision tree with other dataset qualities or features. The outputs from one class should be on a side of the tree, and those from the other class should be on the other, according to a perfect split. Every node is divided in this way until it achieves a perfect split, which results in the formation of a tree's leaf node. The choosing of attributes is the main difficulty in creating a DT. That is, it is challenging to choose which qualities to employ as root nodes or internal nodes given the abundance of available attributes. Information Gain and Gini Index are two methods that can be used to this end:

Information Gain (T, X) = Entropy (T) – Entropy (T, X) (2)

where T refers to the present state and X to the selected attribute;

$$\text{Gini index} = 1 - \sum(p)^2 = 1 - [(p+)^2 + (p-)^2] \qquad (3)$$

where p+ represents the probability of Yes and p- the probability of No.

C. SUPPORT VECTOR MACHINE (SVM)

Support vector machines (SVMs) are supervised learning models used in artificial intelligence that analyse data for order and relapse. Support Vector Machine, often known as SVM.[9] It is a supervised machine learning technology that may be used to address classification and regression issues, but it excels at classification. SVM functions well with small data sets, but it performs better and is more effective with large data sets. SVM begins by charting every point in a dataset with n features in an n-dimensional space, and each point is given a coordinate based

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7431
***Article Received**: 10 May 2023 **Revised:** 30 June 2023 **Accepted:** 29 July 2023*

on the value of its features. From this point on, the classification procedure is carried out by choosing an appropriate hyperplane that most thoroughly divides the points into two different groups. Support vectors are simply the points that define the position and orientation of the hyperplane and are positioned nearby. The margin is the distance between the hyperplane and the support vectors, and it must be maximised as much as possible to produce the most accurate hyperplane.

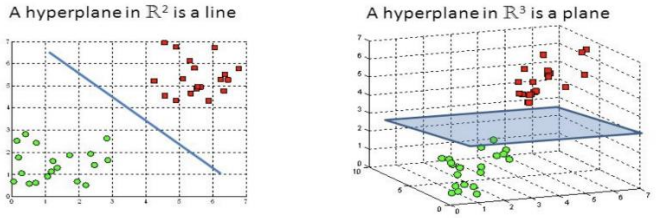

Figure 3. Support Vector Machine

First and foremost, the SVM algorithm excels at processing large, multidimensional datasets. It is quite helpful in situations where there are more samples than dimensions. Because SVM uses support vectors for training, it uses less memory.

The first drawback of the SVM approach is that because it takes more time to train the model, it is not appropriate for very large datasets. When the target classes overlap, it also produces inaccurate results. Additionally, the SVM algorithm is unable to take probability into consideration. To help with decision-making, agricultural data is categorised using SVM.

D. K-NEAREST NEIGHBOR (KNN)

The king is one of the commonly used supervised and nonparametric machine learning methods, and it is used to solve classification.

$$((x_2 - x_1)^2 + (y_2 - y_1)_2)^{1/2} \quad (5)$$

Hamming Distance is method finds distances by depending on common neighbors. |x1 − y1 | + |x2 − y2| If x1 and y1.Minkowski Distance is Similar to Euclidean ation and regression issues. The algorithms with labelled data are those under supervision. In supervised learning, labelled data is input that has already been annotated with the desired result. Given pertinent inputs, supervised learning algorithms attempt to create models from the data and predict the output data.

The technique relies on the distances between the points, which can be calculated in a few different ways. The fact that the distance must always be either zero or positive should be considered. The distance is squared, raised to a given power, or the absolute values are taken to do this. Examples of distance-finding techniques are the Following.

Manhattan Distance :

$$|x2 - x1| + |y2 - y1| \quad (4)$$

Euclidean Distance is the distance between two points, used in regular geometry. distance, an ‘‘n’’ value is needed here,

$$((x2 - x1)^p + (y2 - y1)^p)^{1/p}$$

where xi and yi are the x and y coordinates of a point on an xy plane.

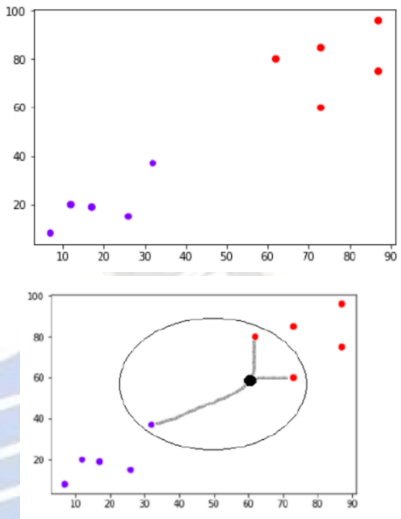

Figure 4. K-NN

The model is first loaded with data. Since kNN uses supervised learning, labelled data must be loaded before it can be used. The desired number of neighbours is then taken into consideration when declaring K. The machine learning algorithm then determines the "distance" or "relation" between each element in the dataset and the query input. After being added to an ordered collection, the distance between the element and the query input is then sorted in increasing order of the distances.[8] Finally, the first K items in the collection are chosen, and depending on whether the model is a classification problem or a regression problem, the output is delivered by choosing the mean in the former case and the mode in the later. The decision of k is crucial because it significantly affects the outcome of our ML model. The model experiences instability and produces more erroneous results if K is set too low. On the other hand, a high value of k will start supplying more faults in the model. As a result, k must have a value that is halfway between the two extremums. K should be chosen as an odd number in the case of a model where a vote is necessary to determine the output in order to guarantee a game of outcome.

The KNN algorithm's main benefits include its simplicity and ease of implementation. The algorithm can also be used to solve a variety of problems, including searching issues as well as classification and regression issues. The algorithm can also be strengthened by including extra training data. The fundamental

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7431
***Article Received**: 10 May 2023 **Revised:** 30 June 2023 **Accepted:** 29 July 2023*

drawback of KNN is that as the size of our dataset grows and the cost of computation rises, the algorithm's performance continues to decline. As a result, it is inappropriate in situations where quick results are needed. Second, to acquire the right results, we must precisely identify the value of k. Sometimes, this procedure can be challenging.

E. Logistic Regression

A approach for supervised classification is logistic regression. This is beneficial for binary data categorization issues. Data containing categorical variables having more than two values can be used with the model in general.

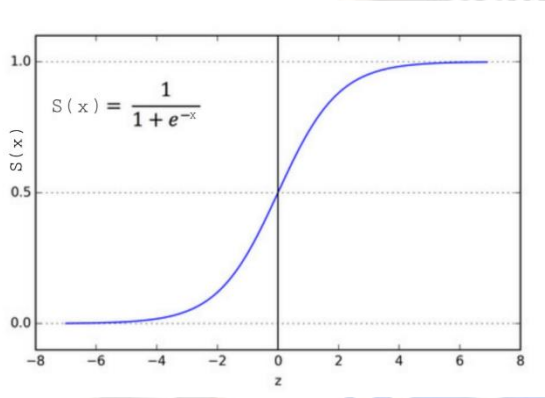

Figure 5. Logistic Reggression

This makes use of the sigmoid function seen in fig. 2. (a). The equation of the function goes as follows:

$$S(x)= \frac{1}{1+e^{-z}} \qquad (6)$$

When dealing with data having linear relationships, this approach works effectively. Additionally, it is less likely to overfit for two-dimensional datasets. Multinomial logistic regression is the name of the extended form of logistic regression.

F. RANDOM FOREST (RF)

Among the best supervised machine learning algorithms is the random forest (RF). In that it links various classifiers to solve a challenging problem and improve the performance of the model, the RF algorithm embodies the essence of ensemble learning. [19] The "forest" that is constructed using this method is made up of several decision trees. Each decision split involves the random selection of RF characteristics. By selecting features at random that encourage prediction and lead to greater efficiency, the correlation between trees is reduced.

An ML approach called Random Forest classifies data by breaking it down into subsets or decision trees, then combining the results of each tree to get the final output. The Bagging subcategory of ensemble learning techniques includes Random Forest. The Row and Feature samples from the primary dataset are chosen at random and fed into the Random Forest Technique's decision trees.

The number of decision trees in the model is decided by the analyst. Every decision tree processes the data and forecasts the outcome using its calculationsIf the result is boolean format, Random Forest picks the majority of the result or the mean/median of the result (in case the result is in numerical form). As a result, using more decision trees improves accuracy and avoids the overfitting issue.

The Random Forest method has a number of benefits. First off, Random Forest is quite well-liked since it is straightforward and reasonably simple to grasp. It can also carry out tasks including classification and regression. It also works well with huge, highly dimensional data sets, and most significantly, it greatly improves the model's accuracy and fixes the overfitting problem. Data extrapolation cannot be done using Random Forest since the findings could be unreliable. Although Random Forest may be used for both classification and regression problems, classification jobs are where it shines. Additionally, it fails to deliver accurate findings when working with sparse data. Additionally, Random Forest takes more resources, larger data, and more time for implementation. It is well known that Random Forest yields inaccurate results. Fig. explains the random forest concept.

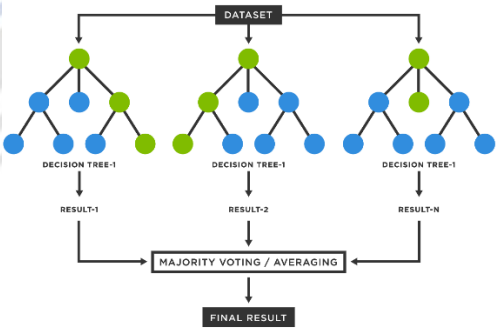

Figure 6. Random Forest

The random forest concepts. The expectation of an average of B trees is like the expectation of each given that each bagged tree is dispersed in the same way. This explains why the bias of bagged trees is the same as the bias of individual trees, hence the only way to decrease it is to reduce the variance. As opposed to progressing, where the trees are distributed differently because they are growing adaptively to eliminate bias. The variance of an average of B identically distributed random variables is 2. The variance of average is given as if the variables have a perfect symmetric distribution but a positive pairwise correlation.

$$\rho\,\sigma^2 + \frac{1+(1-\rho)}{B}\,\sigma^2 \qquad (7)$$

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7431
***Article Received**: 10 May 2023 **Revised:** 30 June 2023 **Accepted:** 29 July 2023*

The value of the second term changes very little as B increases, whereas the value of the first term stays constant. Therefore, the extent of the bagged trees' association limits the advantages of averaging. The RF concentrates on minimising bagging variance by lowering the correlation between the trees without significantly increasing variance. This approach is made possible by the tree-growing process, which randomly chooses input variables.

IV. RESULT AND DISCUSSION

In this section, it is explained the results of research and at the same time is given the comprehensive discussion. Results can be presented in figures, graphs, tables and others that make the reader understand easily [14], [15]. The discussion can be made in several sub-sections.

A. Accuracy Parameters

1) Accuracy- Its proportion of accurate predictions for the test data is its definition. It is simple to compute simply dividing the number of accurate forecasts by the total number of guesses.

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (8)$$

2) Precision- Precision is defined as the proportion of instances out of all the examples that were predicted to belong to a certain class that are really relevant (also known as true positives).

$$\text{Precision} = \frac{TP}{(TP+FP)} \quad (9)$$

3) Recall- The percentage of instances that were correctly identified as belonging to a class in relation to all of the examples in the class is known as recall.

$$\text{Recall} = \frac{TN}{(TN+FP)} \quad (10)$$

4) F1- Score- the proportion of bad incidents compared to all actual negative incidents. Thus, the denominator (TN + FP) in this equation equals the actual number of negative cases in the dataset.

$$\text{F-1 Score} = \frac{2\times TP}{2\times TP+FN+FP} \quad (11)$$

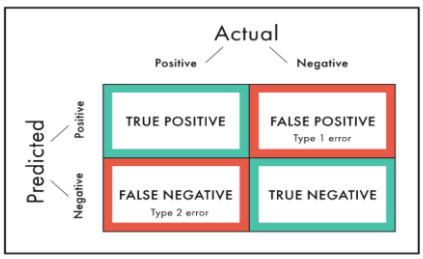

Figure 7. Confusion Matrix

B. Resuts

The Table after implementation of algorithms on the dataset available in public are shown in below tables along with the plot of their accuracy the comparison is done on the python using google collab and weka software with different training and testing spit percentage.

The below table1 shows the comparsion of algorithms as discussed above in detail. The google collab tool is used for comparison for the accuracy by using various standard python libraries.the split ratio of 80-20% is used. The resluts shown are as follows

Table 1. Collab comparison with 80-20 % split ratio

Accuracy Parameter	Accuracy	Precision	Recall	F1-Score
Logistic Regression	95.22	0.95	0.95	0.95
Random Forest	99.09	0.99	0.99	0.99
Naïve bayes	99.09	0.99	0.99	0.99
K-NN	97.5	0.98	0.97	0.98
SVM	97.727	0.98	0.97	0.98
Decision Tree	90	0.86	0.9	0.87

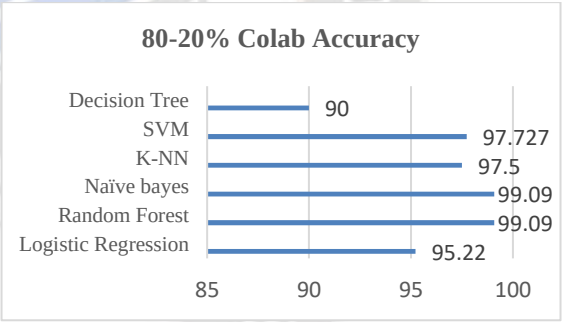

Figure 8. Visulaization of collab comparison with 80-20 % split ratio

The below table1 shows the comparsion of the machine learning algorithms as discussed above in detail. The google collab tool is used for comparison for the accuracy by using various standard python libraries.the split ratio of 70-30% is used. The resluts shown are as follows

Table 2. Collab comparison with 70-30 % split ratio

Accuracy Parameter	Accuracy	Precision	Recall	F1-Score
Logistic Regression	96.21	0.96	0.96	0.96
Random Forest	98.93	0.99	0.99	0.99
Naïve bayes	99.24	0.99	0.99	0.99
K-NN	97.5	0.98	0.97	0.98
SVM	98.03	0.98	0.98	0.98
Decision Tree	85.9	0.81	0.86	0.81

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7431
***Article Received**: 10 May 2023 **Revised:** 30 June 2023 **Accepted:** 29 July 2023*

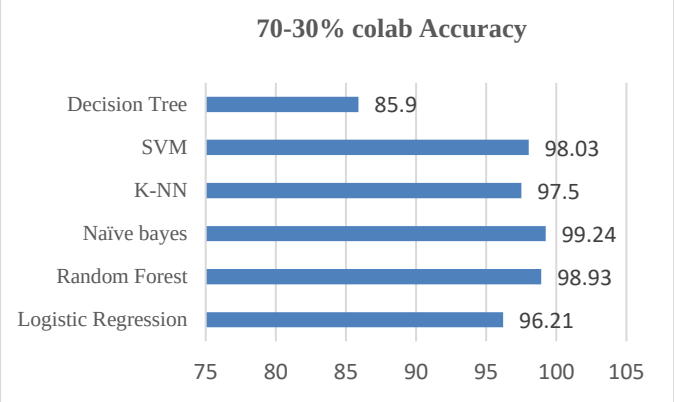

Figure 1. Visulaization of collab comparison with 70-30 % split ratio

The below table1 shows the comparsion of the machine learning algorithms as discussed above in detail. The google collab tool is used for comparison for the accuracy by using various standard python libraries.the split ratio of 60-40% is used. The resluts shown are as follows

Table 3. Collab comparison with 80-20 % split ratio

Accuracy Parameter	Accuracy	Precision	Recall	F1-Score
Logistic Regression	95.11	0.95	0.95	0.95
Random Forest	99.31	0.99	0.99	0.99
Naïve bayes	99.31	0.99	0.99	0.99
K-NN	97.5	0.98	0.97	0.98
SVM	98.06	0.98	0.98	0.98
Decision Tree	86.02	0.85	0.86	0.83

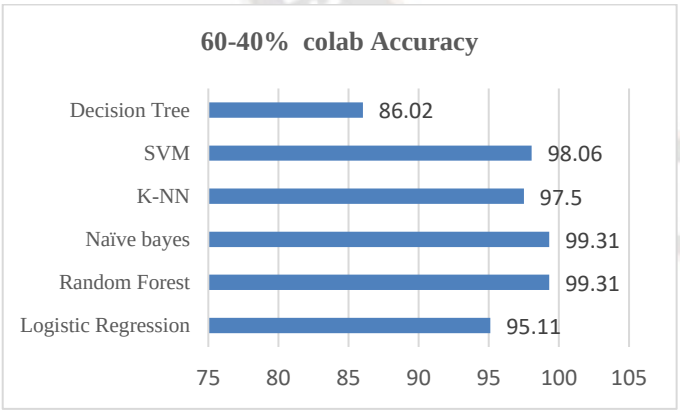

Figure 2. Visulaization of collab comparison with 60-40 % split ratio

The below table1 shows the comparsion of the machine learning algorithms as discussed above in detail. The weka tool is used for comparison for the accuracy by using various standard condition for implementation of algorithms.the split ratio of 80-20% is used. The resluts shown are as follows

Table 4. weka comparison with 80-20 % split ratio

Accuracy Parameter	Accuracy	Precision	Recall	F1-Score
Logistic Regression	98.6364	0.987	0.986	0.986
Random Forest	99.3182	0.993	0.993	0.993
Naïve bayes	99.5455	0.995	0.995	0.995
K-NN	98.4091	0.985	0.985	0.985
SVM	97.2727	0.975	0.973	0.973
Decision Tree	97.5	0.976	0.975	0.975

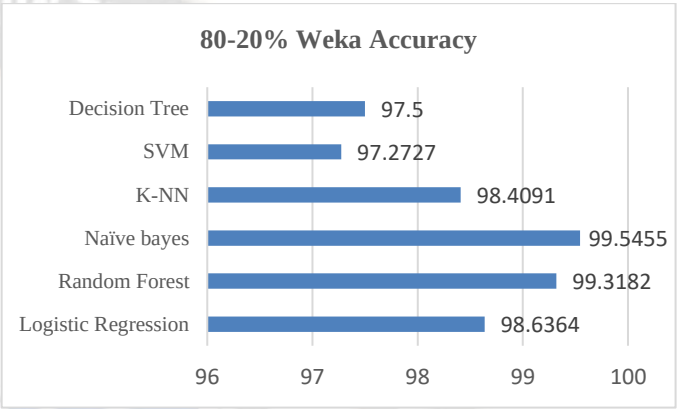

Figure 3. Visulaization of weka comparison with 80-20 % split ratio

The below table1 shows the comparsion of the machine learning algorithms as discussed above in detail. The weka tool is used for comparison for the accuracy by using various standard condition for implementation of algorithms.the split ratio of 70-30% is used. The resluts shown are as follows

Table 5.weka comparison with 70-30 % split ratio

Accuracy Parameter	Accuracy	Precision	Recall	F1-Score
Logistic Regression	97.2727	0.974	0.973	0.973
Random Forest	99.5455	0.996	0.995	0.995
Naïve bayes	99.9393	0.994	0.994	0.994
K-NN	98.18	0.983	0.982	0.982
SVM	97.121	0.974	0.971	0.971
Decision Tree	97.27	0.974	0.974	0.974

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7431
***Article Received**: 10 May 2023 **Revised:** 30 June 2023 **Accepted:** 29 July 2023*

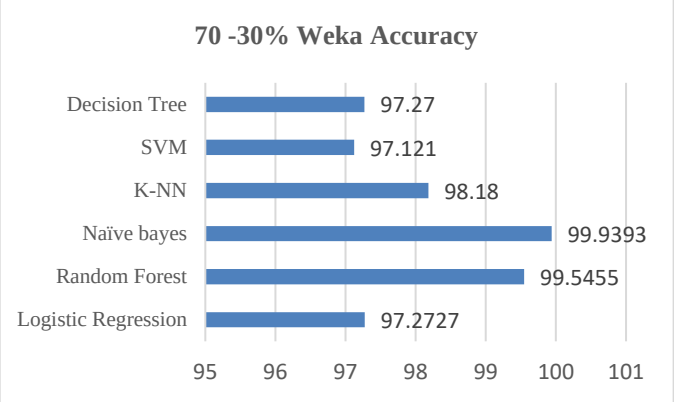

Figure 4. Visulaization of weka comparison with 70-30 % split ratio

The below table1 shows the comparsion of the machine learning algorithms as discussed above in detail. The weka tool is used for comparison for the accuracy by using various standard condition for implementation of algorithms.the split ratio of 60-40% is used. The resluts shown are as follows

Table 6. weka comparison with 60-40 % split ratio

Accuracy Parameter	Accuracy	Precision	Recall	F1-Score
Logistic Regression	97.2727	0.974	0.973	0.973
Random Forest	99.5455	0.996	0.995	0.995
Naïve bayes	99.9393	0.994	0.994	0.994
K-NN	98.18	0.983	0.982	0.982
SVM	97.121	0.974	0.971	0.971
Decision Tree	97.27	0.974	0.974	0.974

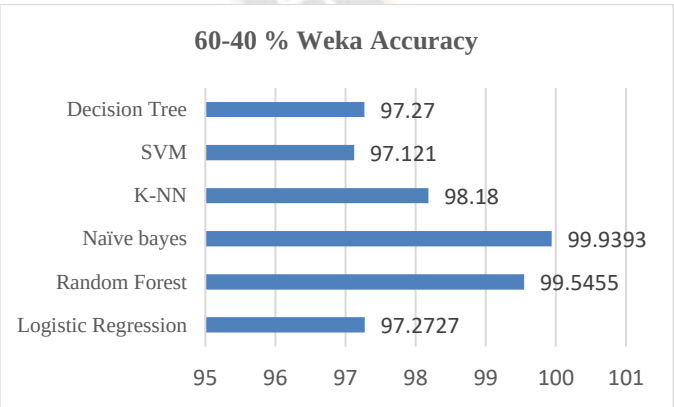

Figure 5. Visulaization of weka comparison with 60-40 % split ratio

C. Discussion

It observed that the most dominant algorithms in both the analysis are Random Forest and Naïve bayes but the accuracy varies in every analysis. The table shows that RF and NB this dataset has ben divided into training and testing in various ratios. These ratios have similar results on them the software used for this purpose are weka and the google collab checking the accuracy parameter by pre-processing as well as real time implementation. The accuracy of the Naïve bayes in the split ratio of 80-20 % is 99.24 % and of random forest is 98.93 %. The accuracy is same for theses two algorithm in split ratio of 70 % - 30 % which is 99.31 % and the best results is given in the 60 % - 40 % split i.e., of 99.31 % and 99.545 % for Naïve bayes and Random Forest Respectively. These results are calculated using the collab and using the libraries for each algorithm and tuning their respective parameters. The weka tool used shows similar results and shows different values like the accuracy of the Naïve bayes in the split ratio of 80 % - 20 % is 99.545 % and of random forest is 98.31 %. The accuracy is same for these two algorithms in split ratio of 70 % - 30 % is 99.93 % and 99.545 % and the results given in the 60 % - 40 % split are of 99. 545 % and 99.181 % for Naïve bayes and Random Forest Respectively.

V. CONCLUSION

It observed that the most dominant algorithms in both the analysis are Random Forest and Naïve bayes but the accuracy varies in every analysis. The weka and collab analysis varies but the conclusion for both of them is similar. The Naïve bayes and Random forest when splited in proper training and testing ratio can give best accuracy of the algorithm which can best used for the farming algorithm. The weka analysis provides the 70-30 split is best for the random forest algorithm and 60-40% or 80-20 % split is best for the Navies bayes algorithm. Whereas the collab analysis gives the 60-40 % as the best split ratio for both the Naïve bayes and Random Forest algorithms. Hence the best ratio can be applied on the the final model for the application purpose in the various Agri-Tech products.

REFERENCES

[1] Guruprasad Deshpande, Anjali Sakurkar and Sonali Kale "Opportunities in Crop Deciding Platform" ES Food & Agroforestry, volume 1, September 2020

[2] S. Kangasuba Raja, Rishi R, Sudarshan E, Sirijit V. "Demand-Based Crop Recommended system for farmers" IEEE International Conference on Technological Innovation in ICT for agriculture and Rural Development-2017

[3] Balaji Prabhu B.V., Dr. M Dakshayaini "Demand Prediction Model for forecasting AGRI-Needs of Society"-International conference on inventive computing and informatics-2020

[4] Soumya Sri Attaluri, Nowshatha k Batcha, Raheen Mafias "Crop plantation Recommendation using feature extraction and machine learning techniques", Journal of Applied Technology and Innovation-2020

[5] Aditya Humnabadkar, Omkar Kulkarni, Mrs. Rajani. P. K, " Automation of Vehicle Identification at Night using Light Source Recognition", 5th International Conference on Computing Communication Control and automation

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7431
***Article Received**: 10 May 2023 **Revised:** 30 June 2023 **Accepted:** 29 July 2023*

(ICCUBEA 2019), Pune, India, held on 19th to 21st September 2019

[6] Uddin, S., Khan, A., Hossain, M. et al. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak 19, 281 (2019).

[7] Kumar Pramanik, K. ., B. Rahane, S. ., Jayamani, V. N. ., Mehra, R. ., C. R., M. ., & V. Athawale, S. . (2023). Cloud Computing based Wireless Sensor Network in Data Transmission With Routing Analysis Protocol and Deep Learning Technique. International Journal of Intelligent Systems and Applications in Engineering, 11(3s), 165–169. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2556

[8] Rajani P. K. and Arti Khaparde "Video error concealment using block matching and frequency selective extrapolation algorithms", Proc. SPIE 10443, Second International Workshop on Pattern Recognition, 104431C (19 June 2017)

[9] G. M. Naidu "Applicability of Arima models in the wholesale vegetable market." Int. J. Agric. Stat. Sci., vol. 11, no. 1, pp. 69–72. (2015)

[10] Mr. Kunal Verma, Mr. Dharmesh Dhabliya. (2015). Design of Hand Motion Assist Robot for Rehabilitation Physiotherapy. International Journal of New Practices in Management and Engineering, 4(04), 07 - 11. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/40

[11] Joachims T. Making large-scale SVM learning practical. SFB 475: Komplexitätsreduktion Multivariaten Datenstrukturen, Univ. Dortmund, Dortmund, Tech. Rep. 1998. p. 28.

[12] Rajani, P. K., Neha Motagi, Komal Nair, and Rupali Narayankar. "Autonomous Smart Device for COVID-19 Detection Using Artificial Intelligence." In Handbook of Research on Applied Intelligence for Health and Clinical Informatics, pp. 128-147. IGI Global, 2022.

[13] B. Sawicka, A. H. Noaema, T. S. Hameed, and B. Krochmal-Marczak, ''Biotic and abiotic factors influencing on the environment and growth of plants,'' (in Polish), in Proc. Bioróżnorodność Środowiska Znaczenie, Problemy, Wyzwania. Materiały Konferencyjne, Puławy, May 2017.

[14] M. Marenych, O. Verevska, A. Kalinichenko, and 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., vol. 16, no. 2, pp. 183–188, 2014.

[15] S. K. Honawad, S. S. Chinchali, K. Pawar, and P. Deshpande, ''Soil classification and suitable crop prediction,'' in Proc. Nat. Conf. Comput. Biol., Commun., Data Anal. 2017, pp. 25–29.

[16] N. Rale, R. Solanki, D. Bein, J. Andro-Vasko, and W. Bein, ''Prediction of Crop Cultivation,'' in Proc. 19th Annu. Comput. Commun. Workshop Conf. (CCWC), Las Vegas, NV, USA, 2019, pp. 227–232.

[17] Ricci, A., Jankowski, M., Pedersen, A., Sánchez, F., & Oliveira, F. Predicting Engineering Student Success using Machine Learning Algorithms. Kuwait Journal of Machine Learning, 1(2). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/118

[18] Shivnath and S. Koley, ''Machine learning for soil fertility and plant nutrient management using back propagation neural networks,'' Int. J. Recent Innov. Trends Comput. Commun., vol. 2, no. 2, pp. 292–297, 2014.

[19] M. Paul, S. K. Vishwakarma, and A. Verma, ''Analysis of soil behaviour and prediction of crop yield using data mining approach,'' in Proc. Int. Conf. Comput. Intell. Commun. Netw. (CICN), Dec. pp. 766–771, 2015.

[20] S. Pudumalar, E. Ramanujam, R. R. Harine, C. Kavya, T. Kiruthika, and J. Nisha, ''Crop recommendation system for precision agriculture,'' in Proc. 8th Int. Conf. Adv. Comput. (ICoAC) , pp. 32–36, 2017.

[21] Mariammal, A. Suruliandi, S. P. Raja, and 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., vol. 8, no. 5, pp. 1132–1142, Oct. 2021.

[22] A. Liaw and M. Wiener, ''Classification and regression by random forest,'' R News, vol. 2, pp. 18–22, Apr. 2002.

[23] Datalink: Kaggle datasets download -d atharvaingle/crop-recommendation-dataset