Journal of Theoretical and Applied Information Technology
30th November 2022. Vol.100. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

IMPROVING SVM CLASSIFIER MODEL USING TREE STRUCTURED PARZEN ESTIMATOR OPTIMIZATION FOR CROP PREDICTION

VASANTHANAGESWARI[1], P. PRABHU[2]

[1] Ph.D., Research Scholar, Department of Computer Applications, Alagappa University, Karaikudi, Tamilnadu, India

[2] Assistant Professor in Information Technology, Directorate of Distance Education, Alagappa University, Karaikudi, Tamilnadu, India

E-mail: [1]smilingsudhamathi@gmail.com, [2]prabhup@alagappauniversity.ac.in

ABSTRACT

Economy of India profoundly relies upon farming. Conventional methods of suggestions are still utilized for farming. Farming is the cornerstone of emerging nation like India. For the income, most of their populace relies upon farming. As of now, farming is based on different approximations of fertilization amount and the sort of yield to be developed or planted. Agriculture holds a predominant position in the growth of any country's prosperity. However, there exists a major threat in the crop yield due to unpredictable and uncontrolled climatic changes, traditional farming methods, and poor irrigation services [10]. Machine learning is an arising research field in analysis of crop yield. Yield forecast is a very significant issue in agriculture. Machine learning methods are the better decision for this issue. Different Machine learning methods are utilized and assessed in agribusiness for assessing what's to come year's production of crop. This paper proposes the Improved Support Vector Machine (Tree-structured Parzen Estimator (TPEOSM)) is applied in the agricultural dataset to predict the crop yield with fertilizer amount. The performance metrics such as precision, recall, f1 score, and accuracy is evaluated. The proposed method accuracy is better than the traditional SVM classifier. Thus, we hope it will help the farmers to select the right crop for the cultivation in basis of temperature, rainfall and other weather conditions and amount of fertilizers thus to be used for the particular crop.

Keywords: *Agriculture, Machine Learning, Support Vector Machine, Tree-Structured Parzen Estimator, Supervised Classification*.

1. INTRODUCTION

India is a worldwide force to be reckoned with in the agriculture area. It is the second-biggest maker of rice, wheat, sugarcane, and so forth. Yet, presently a day, the harvest yield is diminishing because of the absence of soil supplements. Because of the accessibility of various composts, ranchers can't find the manure reasonable for the yield. Nowadays, most individuals are unaware of the importance of cultivating crops at the appropriate time and location. As a result of these cultivating methods, the climatic conditions of the seasons are also changing in the face of fundamental assets such as soil, water, and air.

Machine learning is widely used in a number of critical industries, including future healthcare, education, customer segmentation, fraud detection, financial banking, and agriculture. Machine learning (ML) techniques have evolved as an important alternative to and complement to traditional crop production modelling. When compared to other study fields in agriculture, such as soil and water management and livestock production,

ML techniques are increasingly being used in crop production research. Machine learning is an artificial intelligence application that allows a system to learn from examples and experience without having to program it explicitly. Machine learning is a class of techniques that helps software program to improve their accuracy in predicting outcomes from research systems. To obtain a greater understanding of crop yield, we must analyze large amounts of data using a machine learning algorithm, which will provide

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

accurate results, and recommend a better crop to the farmer and fertilizer for the suitable crop.

The remaining paragraphs are arranged as follows; Chapter II discusses related works. The detailed description of the suggested methodology's various phases is covered in Chapter III. The debate and results are covered in Chapter IV.

2. RELATED WORKS

This section describes a review of the literature on crop yield prediction using various algorithms, as well as their accuracy and limitations.

E. Manjula, S. Djodiltachoumy (2017) [1], has chosen Nitrogen, Phosphorus, Potassium, Calcium, Magnesium, Sulphur, Iron, Zinc, and other nutrients are chosen for investigation of soil supplements using Nave Bayes, Decision Tree, and a hybrid approach of Nave Bayes and Decision Tree. The classification algorithms' performance is compared in terms of accuracy and execution time.

Vaneesbeer Singh, Sarwar, Abid (2017) [2], described a method for predicting the yield category based on the status of macronutrients and micronutrients in a dataset using various Machine Learning approaches. The dataset used in the crop yield prediction came from Jammu's Krishi Bhawan (Talab-Tello). Macronutrients (ph, Oc, Ec, N, P, K, S) and Micronutrients (Zn, Fe, Mn, Cu) found in samples taken from different parts of Jammu District are the parameters present in the data. Following the investigation, machine learning methods are used to estimate the yield category. The crop yield will be specified by the category, as projected. The problem of predicting agricultural yield is described as Classification, and many classifier methods are utilized to solve it.

Chougule, Vijay Kumar Jha and Debajyoti Mukhopadhyay (2019) [3], suggests appropriate fertilizer requirements to farmers. Fertilizer recommendations are made based on the soil's nitrogen, phosphorus, and potassium (NPK) level, as well as a fertilizer recommendation system Provides recommendations for crop suitability in a certain region. The framework involves irregular woods calculation for crop proposal as it works effectively on tremendous dataset and can deal with missing values. Paper depicts how k-implies grouping can utilized for foresee best reasonable manure for the yield in light of given accessible NPK content in the sand.

Deepa, K. Ganesan (2018) [4], proposed the results of the rough set technique, VIKOR method, and Shannon's Entropy approach were combined to create a soft decision system. Based on the soft decision system and the bijective soft set technique, classification rules were constructed for five agriculture crops: paddy, groundnut, sugarcane, cumbu, and ragi. The proposed method is also effective for improving soil fertility by providing fertilizer recommendations for optimal crop development circumstances. Experiments show that the suggested model may effectively recommend pattern optimization and increase annual yield. In a climate-changing environment, the method can help identify the location to assess crop adaptability under given nutrient levels and provide insight into nutrient appropriateness assessments for certain crops.

Selva Anushiya, A. M. Rajeswari (2021) [5], developed a classification model using a combined method of rough set and fuzzy to assist farmers in making crop production decisions. In this work, A classification model is created using a combined method of rough set and fuzzy to assist farmers in making crop production decisions. A classification model is created using a combined method of rough set and fuzzy to assist farmers in making crop production decisions. For feature selection, the fuzzy model is employed, and for rule induction, the rough set-based approach is used. Support vector and random forest regression have been shown to be more accurate. when compared to the rough set-based technique.

Ahmed, U., Lin, J. C. W., Srivastava, G., & Djenouri, Y (2021) [6], presents nutrition recommendations based on a new genetic algorithm (IGA) that analyses time-series sensor data to suggest various crop settings. The proposed method is also effective for improving soil fertility by providing fertilizer recommendations for optimal crop development circumstances. Experiments show that the suggested model may effectively recommend pattern optimization and increase annual yield. In a climate-changing environment, the method can help identify the location to assess crop adaptability under given nutrient levels and provide insight into nutrient appropriateness assessments for certain crops.

M. S. Suchithra, M. L Pai (2020) [7], analyzed nutrition recommendations based on a new genetic algorithm (IGA) that analyses time-series sensor data to

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

suggest various crop settings. The proposed method is also effective for improving soil fertility by providing fertilizer recommendations for optimal crop development circumstances. Experiments show that the suggested model may effectively recommend pattern optimization and increase annual yield. In a climate-changing environment, the method can help identify the location to assess crop adaptability under given nutrient levels and provide insight into nutrient appropriateness assessments for certain crops.

M. Kalimuthu, P. Vaishnavi and M. Kishore (2020) [8], examined the technologies by utilizing machine learning, one of the most advanced technologies in crop prediction, his study work assists the rookie farmer in guiding them in sowing the appropriate crops. Naive Bayes, a supervised learning algorithm, proposes a method for doing so. The seed data of the crops is collected here, together with the necessary characteristics such as temperature, humidity, and moisture content, to aid the crops' growth. A mobile application for Android is also being created in addition to the software. In order to begin the prediction process, users are asked to enter parameters such as temperature and location, which will be taken automatically by this application.

J. J. I. Haban, J. C. V. Puno, A. A. Bandala, et.al [9], has created a fuzzy logic-based programme that will supply the right amount of fertilizer to the soil. Season, nitrogen, phosphorus, and potassium levels are examples of inputs for the fuzzy logic system. In this publication, the researchers advised four types of fertilizer to use: Complete, Urea, Solophos, and Muriate of Potash. The input parameters and fuzzy rules will determine the fertilizer combination and amount.

A. k. Gajula, J. Singamsetty, V. C. Dodda and L. Kuruguntla [10], detects soil quality and predicts the best crop for cultivation using the K-Nearest Neighbors (KNN) algorithm. Our programme takes temperature and soil quality as inputs. Furthermore, our technique recommends fertilizer based on the crop expected. The results of the tests reveal that our system effectively forecasts crop selection and production, which is extremely beneficial to farmers.

T. Siddique, D. Barua, Z. Ferdous and A. Chakrabarty [11], employed Multiple Linear Regression (MLR) and K-Nearest Neighbor Regression (KNNR). During the analysis, Multiple Linear Regression (MLR) produced the most accurate results, and it was included into an android application. The android application system may also create a schedule for the entire agricultural process for a certain crop, such as when to apply fertilizers, when to water, and so on.

K. Archana, K.G. Saranya [12], focuses on macronutrients (NPK), pH and electrical conductivity in the soil, as well as temperature, to provide the best crop recommendations. Crop rotation, crop output prediction and forecasting, and fertilizer advice are all built into the proposed system. In this project, a system is created that includes an agricultural dataset and uses a voting-based ensemble classifier algorithm to recommend optimal crops. Crop yield forecasting and prediction will boost agricultural output. Crop rotation on a regular basis will boost soil fertility. This technology helps farmers make informed fertilization decisions. This system had a 92 percent accuracy rating.

Ghadge, Kulkarni, et.al [13], has proposed the objective of the study is to assist farmers in determining soil quality using a data mining approach. As a result, the system focuses on assessing soil quality in order to anticipate which crops are viable for cultivation based on soil type and to maximize crop production by proposing appropriate fertilizer.

S. Jaiswal, T. Kharade, et.al [14], focuses on assisting farmers by answering their agricultural questions and providing a profile of fundamental requirements via a web application, as well as recommending potential government plans to assist farmers. The suggestion system also keeps farmers up to date on recent agricultural trends, as well as new government plans and programmes.

S.M. Pande, Ramesh, et.al [15], implemented a mobile application is used to connect farmers with the proposed system. GPS assists in determining the user's location. As input, the user enters the area and soil type. Machine learning algorithms can be used to identify the most lucrative crop list or to forecast crop yields for a user-selected crop. Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest (RF), Multivariate Linear Regression (MLR), and K-Nearest Neighbour (KNN) are examples of machine learning algorithms. some of the Machine Learning techniques used to forecast agricultural productivity. The Random Forest produced the best results, with a 95 percent accuracy rate. In addition, the system recommends the optimal time to apply fertilizers to increase production.
Y. Ozaki, Y. Tanigaki, et.al [16], has revealed a useful multiobjective Bayesian optimization algorithm

is what we suggest. Multiobjective Tree-structured Parzen Estimator is an extension of the widely utilized Tree-structured Parzen Estimator (TPE) technique (MOTPE) to solve complex problems and for time consuming they compared with other conventional algorithms.

P. Prabhu, N. Anbazhagan(2013)[18],they have suggested the FI-FCM technique for business intelligence in this research based on frequent item sets and Fuzzy C Means clustering to extract the intelligence from the dataset in order to increase business intelligence and make decision-making more effective The suggested method, FI-FCM, aims to cluster the objects into frequent item sets based on stated rules, and then analyze the clustered objects to gain knowledge.

P. Paulraj and Neelamegam (2014) [17] The information from the dataset is extracted using the suggested algorithm's frequent itemset mining and clustering technique, which aims to increase the effectiveness of decision-making and boost business intelligence.

3. PROPOSED TREE-STRUCTURED PARZEN ESTIMATOR OPTIMIZED SUPPORT VECTOR MACHINE (TPEOSM)

Many Farmers are facing difficulty for finding suitable crop and estimating the amount of fertilizer. The proposed methodology helps them to overcome this difficulty. This research work aims to predict the crop yield and the amount of fertilizer for the suitable crop.

The figure 1 shows Proposed Tree-structured Parzen Estimator Optimized Support Vector Machine (TPEOSM) Methodology for prediction of suitable crop and fertilizer.

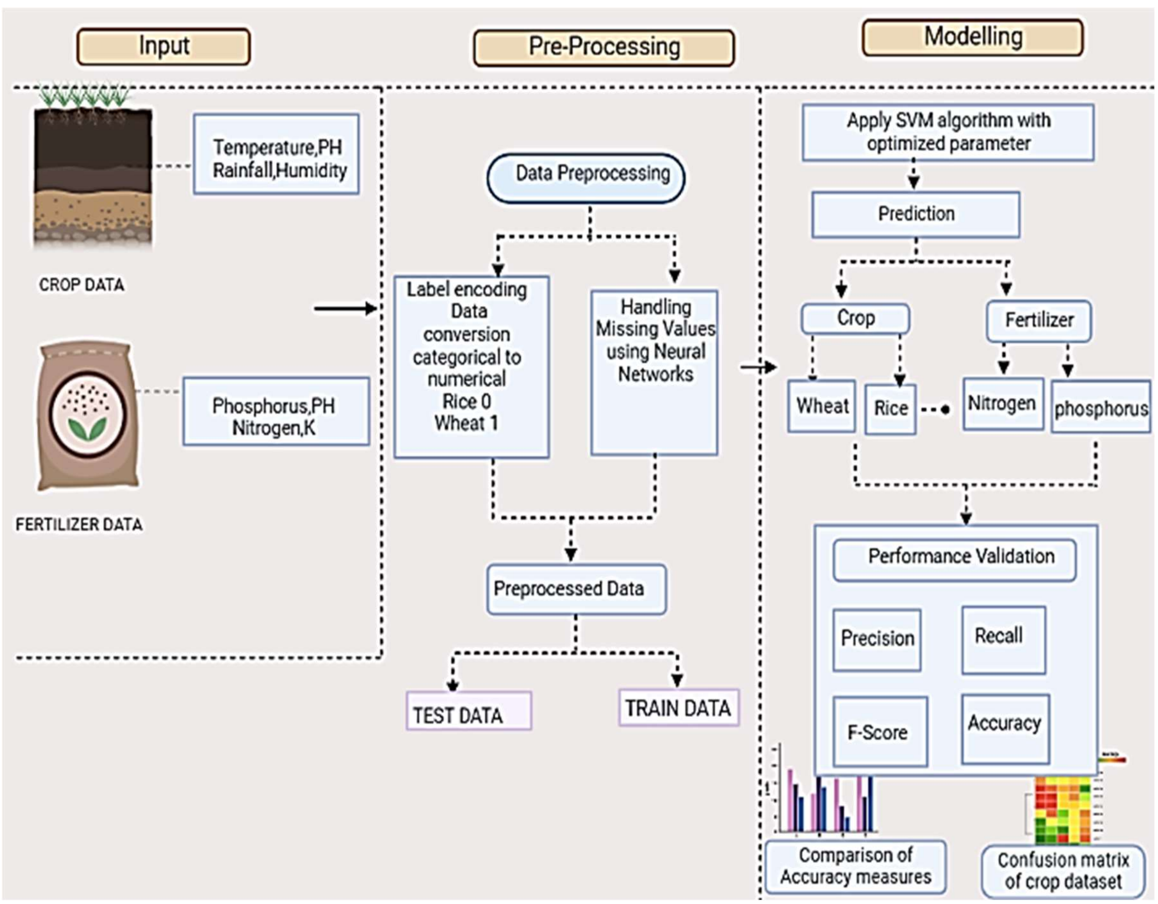

Figure 1 Design of Proposed Tree-structured Parzen Estimator Optimized Support Vector Machine (TPEOSM)Methodology

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.1 Loading Dataset

The task relevant crop and fertilizer data with its features are loaded for preprocessing.

3.2 Preprocessing Phase

Data preprocessing is the process of transforming raw data into processed data that can be used by a
machine learning model. It is the initial and most important step in developing a machine learning model.

3.2.1 Handling Missing Values(if any)

Missing values can be replaced using these five methods: deleting rows, replacing with mean, median, mode, assigning a unique category, and predicting missing values.

3.2.2 Label Encoding

Label encoding is the process of translating labels (data types) into numeric format so that they may be read by machines. For handling categorical variables, label encoding is a common encoding method.

Machine learning algorithms can then better decide how those labels should be used. In supervised learning, it is a crucial pre-processing step for the structured dataset. In this work the label column that is the crop names are in categorical value they are encoded to the numerical values such as rice,wheat are indexed as 0,1 using the label conder.

Here the sample encoded values are displayed.It can further used for the optimization process of support vector machine Algorithm. Table 1 shows the sample label encoding of various crops.

Table 1 Data Type Conversion

Label	Name of the crop
0	Adzuki Beans
1	Black gram
2	Chickpea
3	Coconut
4	Coffee
5	Cotton
6	Ground Nut
7	Jute
8	Kidney Beans
9	Lentil
10	Moth Beans

3.3 Modeling Phase

The Classification algorithm is a Supervised Learning technique that uses training data to identify the category of new observations. A programme in Classification learns from a given dataset or observations and then classifies new observations into one of several classes or groups. In this section a new classification model,Tree-structured parzen estimator optimized Support Vector Machine has been proposed.

SVM provides better result when compared to other conventional Algorithms,Hence SVM parameters are tuned using the optimization algorithm. Supervised Learning algorithms are used for both classification and regression problems. However, it is primarily used in Machine Learning for Classification problems. The SVM algorithm's goal is to find the best line or decision boundary for classifying n-dimensional space.SVM parameters are C and Kernel .These parameter C and Kernel are assigned with default value which results in low accuracy in classification. Hence Parameter tuning required. TPEOSM provides solution for optimziaton problems. So in this work Opimization method is used to find optimal parameters for SVM.

3.3.1 Parameter Optimization Using Tree-structured Parzen Estimator

In parameter optimization, parameter values that is suitable for better result are selected. For instance, Minimum objective function is desired for some appropriate dataset. Optimization can be done in several ways. In an extremely broad manner, they can be separated into two general classifications, in particular, parameter optimization and function enhancement strategies. In this proposed work Tree-Structured Parzen Estimator is used for the parameter optimization in the support vector machine. The Tree-structured Parzen Estimator (TPE) is an SMBO (sequential model-based optimization) method. SMBO methods build a model to estimate the performance of hyperparameters based on previous measurements in a sequential manner, and then choose new hyperparameters to test based on this model.

The figure 2 shows the steps involved in Tree-structured Parzen Estimator.

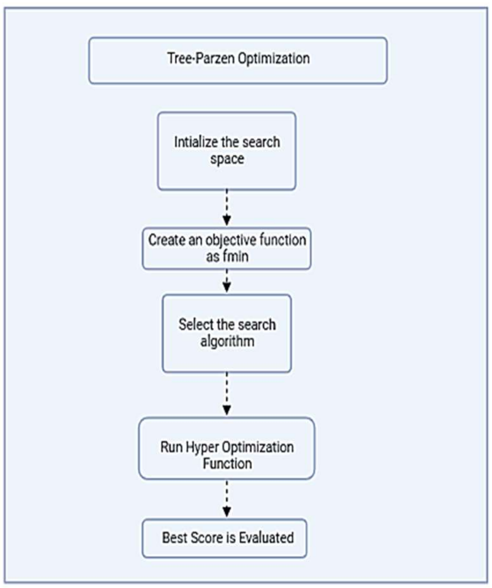

Figure 2 Tree-structured Parzen Estimator Optimization Method

The TPE method models P(x|y) and P(y), where x denotes hyperparameters and y denotes the corresponding quality score.An surrogate model is built p(score/hyperparameters) for p(y/x) which is the objective function.By creating an surrogate model for an objective function makes the optimization easier. Bayesian methods work by selecting the hyperparameters that perform the best on the surrogate function to determine the next set of hyperparameters to evaluate on the actual objective function. The next set of hyperparameters are selected using the surrogate function.

$$EIy*(x) = \int_{-\infty}^{y*}(y*-y)p\left(\frac{y}{x}\right)dy \quad (1)$$

Where,

y* is the threshold value,
x is the proposed set of hyperparameters,
y is the actual value of objective function.
If p(y/x) is positive it yields better.

$$p\left(\frac{y}{x}\right) = p\left(\frac{x}{y}\right) * p(y)/p(x) \quad (2)$$

$$p\left(\frac{x}{y}\right) = \{l(x), y < y*, g(x), y \geq y^* \quad (3)$$

For the expected improvement l(x) is taken under threshold values gives the better values.

$$EIy*(x) = \gamma y * l(x) - l(x)\int_{-\infty}^{y*}\frac{p(y)dy}{\gamma l(x)+(1-\gamma)g(x)}\alpha\left(\gamma+\frac{g(x)}{l(x)}\right)1$$

(4)

It reduces the running time of hyperparameter tuning best scores on the testing set.

3.3.2 Apply optimized parameters in SVM for prediction

C and Kernel are the parameter of the Support vector machine. These optimized parameter values of SVM classifiers are applied in the test data for better classification of crops and to provide better accuracy compared to the conventional algorithms.

3.3.3 Evaluation

In this work the which crop is suitable for farming is predicted based on the attributes temperature,humidity,rainfall and pH values then the amount of fertilizer for the particular crop is predicted by means of using the optimization methods in the test data.The output of the optimization algorithm is fitted to the test data to obtain the desired result. The resuls are tested with various performace mesures using precision,recall,f measure and accuracy.

3.4 Algorithm (TPEOSM)

Input:

//Crop and fertilizer dataset
CROP – (Ph, Temperature, Humidity, Class)
FERTILIZER – (Crop, Nitrogen, Phosphorus, Potassium, PH)

Output:
Crop types – (Rice, Wheat, Barley…….)
Fertilizer Amount (Crop, Nitrogen-80, Phosporus-40)
Begin
Phase 1: Load Crop and Fertilizer dataset

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Phase 2: // Preprocessing
2.1 Label encoding
Converting Datatype (Categorical to Integer)
2.2 Handling Missing values using Neural Network

Phase 3: Modeling
3.1 Perform Tree structured Parzen Estimator Optimization for SVM classifier
1.Initialize the Search Space, Input parameter is specified.
2.Objective function is defined, function to minimized is defined, parameter value is taken from the search space and return the output as loss.
3. By using the selected hyperparameters score is calculated (hyperparameter, score) for the training dataset.
4. The collected observations are sorted by the score values and split into two groups as l(x1) and g(x2) based on threshold value.
5. The primary gathering (x1) contains perceptions that gave the best scores and the subsequent one (x2) - any remaining perceptions.
3.Suggested select algorithm is used it provides a sequential search.
4. Parzen estimators are used to evaluate the data.
5. Hyperparameters are drawn from x1.
6. Maximum no of evaluations are specified.
7. Then, the best score is calculated with the parameters in the l(x1).
3.2 Apply SVM Classifier with optimized parameter for prediction
3.3 Prediction Evaluation
Calculate Precision, recall, f measure and accuracy
End

4 EXPERIMENTAL SETUP

A.Software Tool

In this research work , Python 3.9 software and Jupyter Notebook is previously known as IPython Notebooks is used for implementation. It is an extremely capable general-purpose programming language. Python has simple easy-to-use syntax. It uses libraries such as pandas, NumPy, keras and so on.

B.Dataset Description

Agricultural and fertilizer dataset collected from the public kaggle repository.The crop dataset is comprises of temperature,humidity,rainfall, ph value and fertilizer dataset comprimises of Nitrogen,phosphorus,potassium and pH. Table 2 shows the sample crop dataset and Table 3 shows the sample fertilizer dataset.

Table.2 Crop Dataset

Crop	N	P	K	pH
Rice	80	40	40	5.5
Barley (JAV)	70	40	45	5.5
Maize	80	40	20	5.5

Table 3. Fertilizer datasset

Temperature	humidity	pH	rainfall	label
20.87	82.00	6.502	202.93	Rice
26.08	56.06	6.00	152.09	wheat
28.95	89.07	6.42	57.65	Mung Bean

C.EvaluationMeasure

Evaluation Measures is that the extent to which changes in outcomes may be attributed to the programme is measured by evaluating how successfully the programme activities meet expected objectives. Statistical Measures are a descriptive analytical technique that provides an overview of a data set's properties. The classification errors are calculated using statistical measurements as well as ROC and PRC curves. The Precision,Recall and F Measure are determined.

The number of positive class predictions that actually belong to the positive class is measured by precision.

$$Precision = TP/TP + FP \quad (5)$$

The number of positive class predictions made out of all positive examples in the dataset is measured by recall.

$$Recall = TP/TP + FN \quad (6)$$

F-Measure generates a single score that accounts for both precision and recall concerns in a single number.

$$F - Measure = TP/TP + {}^{1}/_{2}(FP + FN) \quad (7)$$

5.RESULT AND DISCUSSION

Agricultural dataset is used for crop and fertiizer prediction. The accuracy and measured values are calculated. The results of the confusion matrix and statistical measurements are explained.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5.1 Performance metrics

The evaluation metrics such as accuracy, the confusion matrix, precision, recall, and F1 score provide a deeper understanding of prediction results. Table 4 shows Precision,Recall,F-measures for crop dataset.

Table 4 Precision,Recall,F-measures for crop dataset

Classifier	Precision	Recall	F-Measure
SVM	0.66	0.65	0.61
Naïve Bayes	0.78	0.69	0.68
Decision Table	0.66	0.73	0.67
AdaBoost	0.37	0.59	0.45
Random Forest	0.81	0.82	0.81
J48	0.68	0.70	0.67
TPEOSM	0.86	0.84	0.84

In the above Table 4 Precision ,recall and F-measures are calculated for the crop dataset are shown TPEOSM proposed classifier gives higher accuracy when compared to other conventional algorithms as TPEOSM.

Table 5 shows the comparison of Precision,Recall,F-measures obtained using TPEOSM model using crop dataset.

Table 5 Comparison of Crop-wise Precision, Recall-F-measures and support using TPEOSM model

Crop	precision	recall	f1-score	support
0	1.00	1.00	1.00	24
1	1.00	0.97	0.98	30
2	1.00	1.00	1.00	27
3	0.97	1.00	0.98	30
4	0.86	0.91	0.88	33
5	0.96	0.93	0.95	28

In the above Table 5 shows Precision,Recall,F-measures individual crop is shown.

Table 6 shows the comparison of Precision,Recall,F-measures obtained using TPEOSM model using fertilizer dataset.

Table 6 Comparison of Precision, Recall-F-measures and support using TPEOSM model

Classifier	Precision	Recall	F-Measure
SVM	0.82	0.78	0.78
Naïve Bayes	0.78	0.69	0.68
Decision Table	0.66	0.73	0.67
AdaBoost	0.37	0.59	0.45
Random Forest	0.81	0.82	0.81
J48	0.68	0.70	0.67
TPEOSM	0.89	0.89	0.88

Table 4 and Table 6 shows that the proposed methodology measures are better than the conventional algorithms.Adaboost has the least measure value of 0.37 and the TPEOSM algorithm measures 0.89.

5.2 Confusion Matrix

An N x N matrix is used to evaluate the performance of a classification model, where N is the number of target classes. The matrix compares the actual goal values to the machine learning model's predictions. Confusion matrix is a table that is used to define a classification algorithm's performance. The performance of a classification algorithm is shown in Figure 3 and figure 4 and summarised using a confusion matrix.

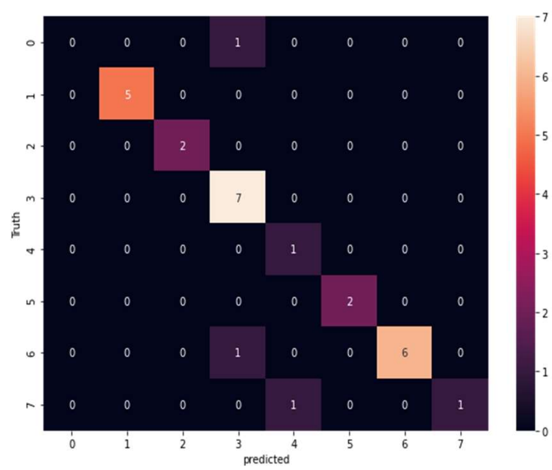

Fig 3 Confusion Matrix of Conventional SVM Algorithm of Dataset

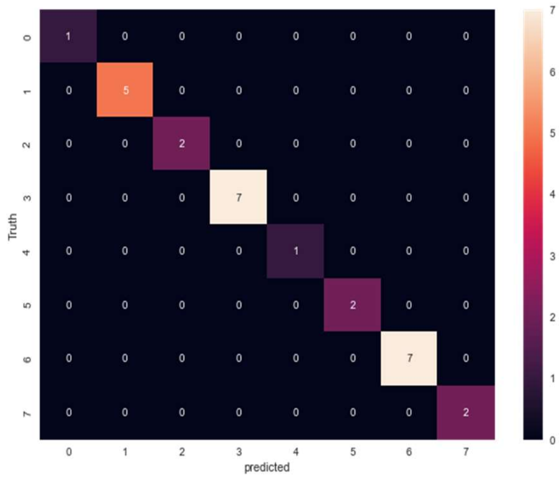

Fig 4 Confusion Matrix of Proposed Algorithm of Dataset

In this Table Eight crops are sampled for the confusion matrix it indicates the misclassified crop

in different shades. when compared to the conventional algorithm proposed algorithm classifies the crop accurately. It increases the accuracy of the algorithm by reducing the misclassification of crops.

Confusionmatrix for sampled crop dataset is shown for clear view of classification it is compared with conventional algorithm.

5.3 Classification Errors

The R^2, MAE, and RAE coefficients are used to calculate classification errors. The average Error of the dataset is given by MAE. RAE gauges the model's performance by taking the root relative square error of the two numeric vectors. The proportion of the dependent variable's variance is shown by the coefficient of determination, also known as R-squared.

$$MAE=\frac{1}{N}\Sigma_{i=1}^{N}|y_i-\hat{y}| \quad (5)$$

Where,

ŷ-Predicted value of y

y-Mean value of y

$$R^2=1-\Sigma(y_i-\hat{y})/\Sigma(y_i-\hat{y})^2$$

$$RAE=\Sigma_{i=1}^{n}|y_i-\hat{y}_i|/\Sigma_{i=1}^{n}|y_i-\hat{y}_i|$$

Where,

$$\hat{y}=1/n\Sigma_{i=1}^{n}y_i$$

Algorithm	R^2	MAE	RAE
SVM	0.36	0.27	0.90
Naïve Bayes	0.30	0.17	0.56
Decision Table	0.27	0.15	0.49
Random Forest	0.13	0.15	0.21
J.48	0.29	0.14	0.33
Adaboost	0.38	0.30	0.99
TPEOSM	0.11	0.12	0.10

Table 8 Error values of the various classifier

Table.8 Error values of various classifier are discussed when compared to the conventional classifiers the proposed optimized TPEOSM error values measure low error value.

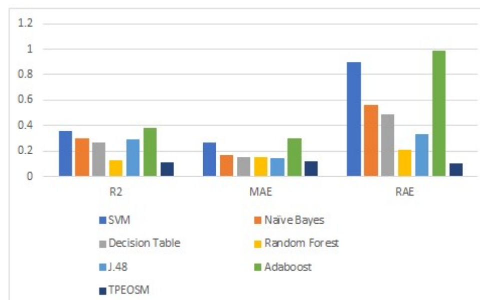

Fig.5 Comparison of Error values of various classifiers.

The figure 5 represents the R^2 MAE, RAE error values graph for the various classifiers with proposed TPEOSM classifier.

Here SVM is optimized by using Tree-structured Parzen Estimator as optimized SVM provides better accuracy than the conventional algorithm. When

SVM is optimized using Partical swarm optimization, Tree-structured Parzen Estimator and Bayesian Optimizer.

Table 9 Comparison of Optimizers for the crop dataset

Optimizer	values
PSO	0.66
TPE	0.86
Bayesian	0.84

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 10 Comparison of Optimizers for the fertilizer dataset

Optimizer	values
PSO	0.85
TPE	0.88
Bayesian	0.86

In the above table various optimization methods are implemented with SVM algorithm are compared for both crop and fertilizer data by comparing with various algorithm we can conclude that TPE optimizer provides best optimizer with better accuracy as 0.89 for crop dataset and 0.88 for fertilizer dataset.

6. CONCLUSION

An important issue in agriculture is Prediction of crop yield and to get the high yield of the crop by using appropriate fertilizer for the crop. A new method TPEOSM have been implemented for the better performance. Crop data with amount of fertilizer data is taken as sample. Performance metrics such as confusion matrix, Precision, Recall, F1 score, ROC curve have been calculated and concluded that TPEOSM gives better accuracy. When compared to the conventional Algorithm SVM Proposed TPEOSM scores 0.27% accuracy higher than the Conventional SVM.

REFERENCES

[1] Manjula E, Djodiltachoumy S. Data mining technique to analyze soil nutrients based on hybrid classification. International Journal of Advanced Research in Computer Science. 2017 Sep 1;8(8).

[2] Singh V, Sarwar A, Sharma V. Analysis of soil and prediction of crop yield (Rice) using Machine Learning approach. International Journal of Advanced Research in Computer Science. 2017 May 15;8(5).

[3] Chougule A, Jha VK, Mukhopadhyay D. Crop suitability and fertilizers recommendation using data mining techniques. InProgress in Advanced Computing and Intelligent Engineering 2019 (pp. 205-213). Springer, Singapore.

[4] Deepa N, Ganesan K. Multi-class classification using hybrid soft decision model for agriculture crop selection. Neural Computing and Applications. 2018 Aug;30(4):1025-38.

[5] Selva Anushiya, A.M. Rajeswari, (2021). Crop Suitability Prediction and Fertilizer Recommendation Using Classification Techniques. In Advances in Automation, Signal Processing, Instrumentation, and Control (pp. 1917-1923). Springer, Singapore.

[6] Ahmed, U., Lin, J. C. W., Srivastava, G., & Djenouri, Y. (2021). A nutrient recommendation system for soil fertilization based on evolutionary computation. Computers and Electronics in Agriculture, 189, 106407.

[7] Suchithra, M. S., & Pai, M. L. (2020, February). Data mining based geospatial clustering for suitable recommendation system. In 2020 International Conference on Inventive Computation Technologies (ICICT) (pp. 132-139). IEEE.

[8] M.Kalimuthu, Vaishnavi and Kishore, (2020). "Crop Prediction using Machine Learning," Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 2020, pp.926-932, Doi:10.1109/ICSSIT48917.2020.9214190.

[9] Haban JJ, Puno JC, Bandala AA, Billones RK, Dadios EP, Sybingco E. Soil Fertilizer Recommendation System using Fuzzy Logic. In 2020 IEEE REGION 10 CONFERENCE (TENCON) 2020 Nov 16 (pp. 1171-1175). IEEE.

[10] kumar Gajula A, Singamsetty J, Dodda VC, Kuruguntla L. Prediction of crop and yield in agriculture using machine learning technique. In2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT) 2021 Jul 6 (pp. 1-5). IEEE.

[11] Siddique T, Barua D, Ferdous Z, Chakrabarty A. Automated farming prediction. In2017 Intelligent systems conference (IntelliSys) 2017 Sep 7 (pp. 757-763). IEEE.

[12] Archana K, Saranya KG. Crop Yield Prediction, Forecasting and Fertilizer Recommendation using Voting Based Ensemble Classifier. SSRG Int. J. Comput. Sci. Eng. 2020 May; 7:1-4.

[13] Ghadge R, Kulkarni J, More P, Nene S, Priya RL. Prediction of crop yield using machine learning. Int. Res. J. Eng. Technol. (IRJET). 2018 Feb;5.

[14] Jaiswal S, Kharade T, Kotambe N, Shinde S. Collaborative recommendation system for agriculture sector. In ITM web of conferences 2020 (Vol. 32, p. 03034). EDP Sciences.

[15] Pande Sm, Ramesh Pk, Anmol A, Aishwarya Br, Rohilla K, Shaurya K. Crop recommender system using machine learning approach. In2021 5th International Conference on Computing Methodologies and Communication (ICCMC) 2021 Apr 8 (pp. 1066-1071). IEEE.
[16] Ozaki Y, Tanigaki Y, Watanabe S, Nomura M, Onishi M. Multiobjective Tree-Structured Parzen Estimator. Journal of Artificial Intelligence Research. 2022 Apr 8; 73:1209-50.
[17] Paulraj P, Neelamegam A. Improving business intelligence based on frequent item sets using k-means clustering algorithm. In Networks and Communications (NetCom2013) 2014 (pp. 243-254). Springer, Cham.
[18] Prabhu P, Anbazhagan N. FI-FCM algorithm for business intelligence. In Mining Intelligence and Knowledge Exploration 2013 (pp. 518-528). Springer, Cham.
[19] KaggleRepository https://www.kaggle.com/c/fp.