

Classification of Fertiliser Type Based on Soil Minerals Using Voting Classification Over Decision Tree

K. Bandaiah^a and L.Rama Parvathy^{b,1}

^aResearch Scholar, Dept. of CSE, Saveetha School of Engineering,

^bProfessor, Dept. of AI&DS, Saveetha School of Engineering,
SIMATS, Chennai, Tamilnadu, India

Abstract: To estimate the type of fertilizer based on the soil minerals using the voting classifier. For forecasting fertilizer type accuracy %, a Voting Classifier with a sample size of 10 and a Decision Tree with a sample size of 10 was iterated at various times. A supervised learning algorithm is a Decision Tree. It constructs a "forest" using an array of decision trees, typically trained to use the "bagging" method. A Novel Voting Classification is a predictive model that learns from several models and predicts an output (class) based on the result representing the greatest likelihood of being the chosen class. The Novel Voting Classifier produced substantial results with 96 percent accuracy, compared to 94% accuracy for the Decision Tree. The Novel voting classifier and the Decision Tree showed statistical evidence of $p=0.001$ ($p<0.05$). Voting Classifier is the most effective algorithm that classifies the type of fertilizer based on soil minerals with more accuracy than the Decision Tree.

Keywords: Novel Voting classifier, Decision Tree Algorithm, Classification, Fertiliser type, Machine learning, Accuracy.

1. Introduction

The main job of the machine learning algorithms is to identify objects and separate them into different categories, which is called classification. It also changes the massive amount of the data into the typical discrete values [1]. A Novel Voting Classification is a predictive model that learns from several models and predicts an output (class) based on the result representing the greatest likelihood of being the chosen class [2]. A complex voting ensemble in classification includes accumulating the votes for crisp class labels from different models and predicting the class with the most votes [3]. In a soft voting ensemble, the projected probabilities for class labels are added, and the class label with the highest sum probability is predicted [4].

For the object recognition of dangerous photos, around 15 publications were published in IEEE, and ten articles were published in Google Scholar [5]. The essential ideas of the decision tree classifier, a multistage classification approach, are presented

¹L. Rama Parvathy, Department of AI&DS, Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences, Chennai, India. Email: ramaparvathy1.sse@saveetha.com

in this study. Two strategies for creating decision trees are mentioned, as well as experimental findings. Each design method's respective merits and disadvantages are examined. The following is a list of typical remote sensing applications [6]. We employed a decision tree to categorize coastal zones in this paper. Still, we used a purity picture index (PPI) to select end members as training images while picking data into meaningful information. This may decrease the effect of pixel intensities on feature learning before finally classifying with decision tree techniques.

The research gap found in the study is that the classification and prediction accuracy of the particular fertilizer type is more minor, making the prediction less accurate. Also, the data used for the forecast is invalid and does not contain all the features required to predict the fertilizer correctly and accurately. This research focussed on the prediction of the fertilizer using more data with the help of a Novel voting classifier over a Decision Tree.

2. Materials and Methods

The suggested research is being carried out at the DBMS Lab of the CSE Department, SSE, SIMATS, Chennai. For the study, two separate groups were employed. The voting classification method is in group one, while the Decision Tree algorithm is in group two. Each group receives ten samples, for a total of twenty samples, with an alpha of 0.05 and a beta of 0.2, a 95 percent confidence interval, and an 80% pretest power [7].

In this research study, the performance of two algorithms, Novel voting classifiers and Decision Tree, are compared. The fertilizer prediction was made using the dataset collected from the Kaggle website. The dataset contains these columns of Temperature, Humidity, Moisture, Soil Type, Croup Type, Nitrogen, Potassium, Phosphorus, and Fertiliser Name. All the columns of the data have been preprocessed; irrelevant data has been removed. The dataset has been taken from the given source. This dataset has been collected by researching various websites and journals. The dataset contains all the data on the minerals that the land includes, which will be helpful while detecting the correct fertilizer type. Also, the textual data that is present in the dataset has been corrected to the numerical data.

2.1. Novel Voting Classifier

Voting is a machine learning ensemble algorithm. A voting ensemble in regression entails making a prediction that is the average of several other regression models. In classification, a complex voting ensemble is used to forecast which class will receive the most votes by adding the votes for crisp class labels from other models. In a soft voting ensemble, the projected probabilities are added together, predicting the class label with the highest sum probability. A democratic troupe (also known as a "dominant party voting outfit") is a collective AI model that integrates forecasts from other models. In an ideal environment, it is a technique for improving model execution to obtain preferred performance over any one model in the collection. In a democratic gathering, projections from many models are blended. It has the potential to be employed for relapse or characterization. Because of the order, forecasting each mark is added together, and the name with the most votes is predicted. Relapse Ensemble Voting: Predictions are the norm

2.2. Decision Tree Model

The supervised learning algorithms family includes the Decision Tree algorithm. The decision tree technique, unlike many other supervised learning methods, can be used to handle regression tree challenges. The goal of using a Decision Tree is to train an algorithm that can be used to forecast the category or quantity of the target variable by inferring fundamental decision rules from the data. We begin at the top of the tree when using Decision Trees to anticipate a record's class label. The values of the record's attribute and the root attribute are compared. Based on the evaluation, we go to the next node by following the branch corresponding to that value. The Decision Tree is a supervised learning approach that may be used to solve issues in classification and regression. Anaconda Spyder Software was used to test the proposed model (Jupyter Notebook). The system utilized a Windows 10 operating system, an i5 CPU, 256GB SSD, and 12GB RAM.

The statistical analysis is done using IBM's SPSS statistical tool with version 26. The independent variables are image type, and the dependent variables are image size. In SPSS, the dataset is prepared using the ten samples from each algorithm, and the total number of samples is 20. Group id is given 1 for the voting classifier and 2 for the Decision Tree.

3. Results

Group Statistics, the mean accuracy and standard deviation for the Novel Voting Classifier are 95.10 and 1.83787. The Decision Tree means accuracy is 88.65, and the standard deviation is 1.91558. The Novel Voting Classifier also obtained a standard error suggest a rate of 0.58119, whereas the Decision Tree algorithm got an error mean rate of 0.60576, as given in Table 1.

Table 1. Descriptive Statistics of the mean and standard deviation of two groups with each sample size of 10 using Paired T-Test.

	Groups	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	Voting	10	95.10	1.83787	0.58119
	Decision Tree	10	88.65	1.91558	0.60576

An independent sample test of 10 samples was performed using a statistical analysis tool, and the Novel Voting Classifier obtained a mean difference of 18 and a standard error difference of 1.00222. When compared to other algorithm's performance, the Novel Voting Classifier performed better than the Decision Tree Algorithm, as given in Table2.

Table 2. Independent sample T-test

		F	Sig	t	df	Sig (2_tailed)	Mean Diff.	Std. Error Diff.	Lower	Upper
Predicted	Equal Variances assumed	0.045	0.835	7.683	18	0.000	6.45	0.839	4.686	8.213
Actual	Equal Variances not assumed			7.683	17.696	0.000	6.45	0.83948	4.686	8.213

Figure 1. represents a bar chart of the simple bar mean accuracy, with Novel Voting Classifier and Decision Tree Algorithm, the bar chart representing the comparison of mean accuracy of Novel Voting Classifier is 95.100 and Decision Tree Algorithm is 88.65.

Management [8] variables to predict the correct N2O emission and also compared the working ability of Random Forest with several other regression models [9].

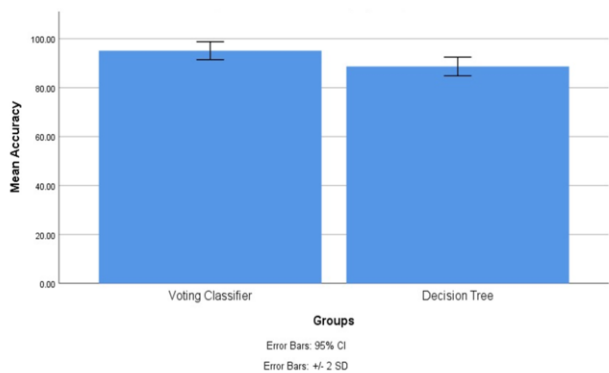


Figure 1. Comparison of Voting Classifier and Decision Tree Analysis in terms of mean accuracy.

According to their research, the Random Forest surpassed regression models in prediction. They also revealed that two crucial input factors, fertilization, and crop type, had an impact on their findings. This research predicts the quantity of fertilizer required for a good yield and its component crops, such as nitrogen and potassium—weather, soils, and land management all impact the type and quantity of fertilizer used. The researchers analyzed data from 273 field experiments in Quebec (Canada) between 1979 and 2017. Predictions were developed, evaluated, and contrasted using k-nearest neighbors, random forest, and Gaussian processes.

4. Discussion

In this study, we observed that the voting classifier has better accuracy than the Decision Tree algorithm by using the independent sample t-test. The mean accuracy of prediction using the Voting classifier is 98.0% and that of using the Decision Tree is 88%. This paper used the historical data of records [10] from Uniform Soybean Tests (UST) in North America [9] and built a Long Short Term Memory - Recurrent Neural Network based model and predicted the genotype of the crop. Also, gathered data according to the weather reports. For yield prediction, the combination of the two models outperformed Random Forest (RF), LASSO regression, and the data-driven USDA model[11]. The proposed work aimed to predict nitrous oxide emission based on local information and used the environment and crop.

5. Conclusion

In this paper, we have used both the algorithms Voting classifier and Decision Tree to predict the correct fertilizer type. When it comes to predicting the proper fertilizer type, it is evident that the voting classifier outperforms the Decision Tree.

Reference

- [1] Shalev-Shwartz S, Ben-David S. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014.
- [2] Casella A, Sanchez L. Storable Votes and Quadratic Voting. An Experiment on Four California Propositions. Epub ahead of print 2019. DOI: 10.3386/w25510.
- [3] M D, Dhanush M, Hasifa AS, et al. Crime Prediction and Forecasting using Voting Classifier. *2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*. Epub ahead of print 2021. DOI: 10.1109/icecct52121.2021.9616911.
- [4] Sherazi SWA, Bae J-W, Lee JY. A soft voting ensemble classifier for early prediction and diagnosis of occurrences of major adverse cardiovascular events for STEMI and NSTEMI during 2-year follow-up in patients with acute coronary syndrome. *PLOS ONE* 2021; 16: e0249338.
- [5] Swain PH, Hauska H. The decision tree classifier: Design and potential. *IEEE Transactions on Geoscience Electronics* 1977; 15: 142–147.
- [6] Su X, Khoshgoftaar TM, Greiner R. Making an accurate classifier ensemble by voting on classifications from imputed learning sets. *International Journal of Information and Decision Sciences* 2009; 1: 301.
- [7] Malhotra R, Khanna M. Software change prediction using voting particle swarm optimization based ensemble classifier. *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. Epub ahead of print 2017. DOI: 10.1145/3067695.3076007.
- [8] Alsayegh AY. *The Effect of Nitrogen Fertilizer Placement on Quality and Quantity of Spring Wheat on Dryland*. 1959.
- [9] Boguta P, Skic K, Sokolowska Z, et al. Chemical Transformation of Humic Acid Molecules under the Influence of Mineral, Fungal and Bacterial Fertilization in the Context of the Agricultural Use of Degraded Soils. *Molecules*; 26. Epub ahead of print 13 August 2021. DOI: 10.3390/molecules26164921.
- [10] Shook J, Gangopadhyay T, Wu L, et al. Crop yield prediction integrating genotype and weather variables using deep learning. *PLoS One* 2021; 16: e0252402.
- [11] Phillibert A, Loyce C, Makowski D. Prediction of N₂O emission from local information with Random Forest. *Environ Pollut* 2013; 177: 156–163.