3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE-2020), 07-08 February 2020, (IEEE Conference Record # 48199)

Impact of Machine Learning Techniques in Precision Agriculture

Rahul Katarya, Ashutosh Raturi, Abhinav Mehndiratta, Abhinav Thapper
Department of Computer Science & Engineering,
Delhi Technological University, New Delhi, India
rahuldtu@gmail.com, ashuhhh07@gmail.com, abhinavmehndiratta@gmail.com, abhinav.thapper3@gmail.com

***Abstract*— Agriculture and its accessories contribute to approximately 17% of India's GDP and it is still the most popular occupation amongst 70% of India's population. The agriculture sector provides different outputs used by diverse segments which include, but not limited to, use as raw materials by different industries, sources of nutrition and businesses, etc. However, the different methods used for growing crops are still mostly traditional and even borderline outdated. Indian farmer still struggles when it comes to picking up the right crop for right biological and non-biological factors. Thus to accelerate the yield of crops, different AI techniques been proposed worldwide and in this paper, we present a summarization of these different approaches. These techniques are a part of the paradigm, Precision Agriculture, more specifically 'crop recommender systems'. The diverse procedures presented in this paper include KNN, Similarity-based Models, Ensemble-based Models, Neural Networks, etc. These algorithms take into account various different factors that are external in nature like meteorological data, temperature and others like soil profile and texture to give best recommendations which not only lead to better yields but also minimum use of resources and capital.**

Index Terms— precision agriculture, crop recommendation systems, AI, machine learning, classification.

I. INTRODUCTION

Agriculture has been the cornerstone of almost all the ancient civilizations for the very primary reason for sustenance. Today, agriculture is a $2.4 trillion industry worldwide is one of the most prominent contributors to the development of third world countries. However, agriculture inherently faces many problems, most of which are highly unpredictable in nature, lack of rain, floods, blight drastic weather changes to name a few. These factors along with the unmeasured use of insecticides and chemical fertilizers, institutional factors like lack of credit, not enough subsidies by the governing body and corruption lead to alienation of the farming body and an increase in the debt which ultimately leads to suicides and families laden with more debt. The above reasons necessitate the ushering in of Artificial Intelligence and the Internet of Things [1] in the field of agriculture to use statistical prowess to provide better yields at relatively lower costs. Along these lines, we present various frameworks based on precision agriculture.

Precision Agriculture [2] includes the use of technology in proposing fertilizers, farming techniques, crops among other things to farmers. The frameworks we discuss are focused on a subsection of precision agriculture called crop recommendation systems which use different machine learning algorithms to recommend crops depending on certain rules and data. The correctness of the recommendation depends on the type and the amount of data fed. The statistical nature of these algorithms can lead to a significant increase in yield. A high accuracy measure is desired as the consequences of the otherwise will be immense which include waste of seeds, time, drastic loss in productivity, etc. Many different predictors can be used for recommendations like temperature, soil properties, humidity, etc. In India, the concept of precision agriculture is not fully embraced and farmers still use traditional methods of growing crops which do not lead to great yield and runs a high risk of failure whilst letting a better alternative of technology-based agriculture go unused.

II. LITERATURE SURVEY

Ensemble methods [3] belong to that branch of machine learning blend several learning methods into one model. The prominent feature of these algorithms is that they give better performance than any of their sub-models. Some examples of these are boosting, bagging and stacking. Majority Voting is a famous ensembling technique in which at least two base learners are required. Learners are selected in such a way that they are aggressive yet supportive of each other. The basic motive is to achieve a better measure of performance than any of the individual models. This method has been used in [4], [5] for crop prediction. Recommendations are done for a localized area of Madurai district in Tamil Nadu, in the form of a UI based system in [4]. The learners used in this system include Random Tree [6], Naive Bayes [7], KNN [8] and CHAID [9]. The most important features are extracted and are introduced to the ensemble model which generates the induction rules for the recommendation. A very similar approach is adopted in [5] which uses, learners Naive Bayes, Linear SVM [10] and Random Forest.

Data mining is a technique of, acquiring, preparing and finding patterns in data to gain useful insights from historical data. It is based on the commonness of machine learning, statistics and database management. There is a growing need to extrapolate the existing knowledge of the

978-1-7281-1683-9/20/$31.00 ©2020 IEEE

farmers by harnessing the vast amounts of past information available to us. Such an approach is utilized in [11] which propose that different types of classification algorithms may behave differently when exposed to data of different crops. It uses a large number of classification-based algorithms like ANN [10] and KNN (for wheat and potato), Harmonic analysis of NDVI Time-series algorithm (for sugarcane), J48, LAD Tree (for rice) and many others and compares their accuracies. An android application is presented in [12] which leverages the power of ANN and proposes a crop calculator that recommends best crops on the basis of different values provided like soil Ph., water availability, temperature, month of cultivation, in the backend on the web and data server. Also provides a subscription to the farmers and gives personalized information to them and also takes feedback so the system can be made more user-friendly. A simple comparative study is presented in [13] which compares classification algorithms like J48 [14], LAD Tree [15], LWL and IBK using the WEKA tool [16] to classify into different seasons and the preprocessing was done using DSS. The performance is compared on the basis of different metrics like RMSE, MAE, and RAE [17].

The RSF Recommender [18] presents a modular approach where different databases are used for storing information related to temperature, seasonal crops and crop growth rate. The country is divided into regions called Upazilas. It has a UI based system and works by detecting the location and identifying the Upazilas which provides the information regarding the different thermal, demographic and physiographic properties of that specific Upazila. Using this information and the information maintained in the database regarding the seasonal crops most similar Upazilas are chosen using Pearson Correlation Similarity and top N most similar Upazilas are selected. Finally, these along with the crop production rates are used to recommend crops keeping in mind the season by the specialized algorithm.

Big Data analytics [19] is the study of examination of very large datasets with an attempt to find patterns, inherent collations and hidden insights. An expert system [20] has a knowledge base that is acquired from domain experts, an inference engine, which uses the knowledge and statistical methods to draw interpretations from the database and an interface, which allows input/extraction of knowledge or rules into/from the knowledge base. An amalgamation of big data analytics and expert systems is presented in [21], it gives insight into a decision system aided by big data analysis, for guiding agricultural production. An expert system has a knowledge base that is acquired from domain experts, an inference engine, which uses the knowledge and statistical methods to draw interpretations from the database and an interface, which allows input/extraction of knowledge or rules into/from the knowledge base. An intelligent agriculture decision system is presented with different components like knowledge acquisition, knowledge representation and reasoning, knowledge base and establishing a reasoning program. It is integrated with a Big Data framework which improves the speed and performance of the decision system as a whole. As an instance of the combination of decision system and big data analytics, a wheat agricultural intelligent decision system was presented in the end which concludes decision making for a variety of tasks like prevention and control of diseases and pests, green stage decision, selection decision of variety, sowing time decision, among others.

A Crop Selection Method (CSM) [22] categorizes the crops as annual (crops that can be grown anytime in the whole year), seasonal (crops that can be grown in a specific season), long-term crops (take a long time to grow) and short-term (take a short time to grow) crops. In many tropical and subtropical countries including India, the amount of rain received is indicative of the crop yield that can be expected that year. According to the approach discussed in [22], multiple sequences of crops are possible from the four categories but only the sequence which gives the best mean yield is selected. For the selection of the best sequence of crops, prediction of the yield rate of the crop is imperative and is predicted based on features like water density, weather, soil type, crop type. Using the predicted yields of the crop multiple sequences are possible, they are filtered and the best sequence is selected.

III. DISTINGUISHING STUDY

This section presents the tabular comparison of the techniques discussed above.

Table 1. A comparative study of different frameworks

S.No.	Year, Reference	Methodology	Inferences
1	2016 [4]	1.Random Tree 2.NaiveBayes 3.KNN 4.CHAID 5.WEKA tool	1. Ensemble model-based. 2. Majority voting 3. Induction rules generated.
2	2017 [11]	1.Multiple Linear regression 2.Harmonic analysis of NDVI time series 3.KNN 4.ANN 5.LADTree 6.C4.5 7.Naive Bayes 8.Decision Tree using ID3	1. Different crops give better yield predictions for different models. 2. Past data required. 3. Improvement required
3	2017 [18]	1. Location	1. Similar crops are

		Detection 2. Pearson Correlation Similarity 3. UI based system with different modules	grown in similar areas. 2. Crop recommendation done by crop production rates
4	2017 [12]	1. ANN 2. Android application development for subscription-based system 3.Personalized support to farmers	1. System has information about the previous crops planted 2. System takes feedback
5	2018 [21]	1. Big Data Analysis 2. Expert System 3. Inference engine 4. Interface for the system	1. Expert system can be used for the KB 2. Big data technology improves processing speed and accuracy
6	2018 [13]	1. J48 2. LAD Tree 3. LWL 4. IBK 5. Classifiers compared with Specificity, Sensitivity, Accuracy, RMSE, MAE, and RAE	1. IBK obtained the highest accuracy 2. LAD Tree showed the least accuracy 3. Pruning of data can help in minimization of errors 4. Weather and soil data if incorporated can improve the prediction
7	2015 [22]	1. Crop Sequencing Technique 2. Past data of crop yield 3. Prediction method with good accuracy.	1. The sequence with the highest net yield rate can be selected for growing.
8	2018 [5]	1. Linear SVM 2. Naive Bayes 3. Random Forest	1. Ensemble model-based. 2.Majority voting

IV. BASIC ARCHITECTURE OF A CROP RECOMMENDATION SYSTEM

After reading through multiple research papers pertaining to different techniques like data mining, prediction, recommendation, in this section we present a fundamental approach in the form of a flowchart, Fig.1, for applying these procedures. The first step is data collection, this includes but not limited to data mining, sensor data, etc. Data is preprocessed by different methods like scaling, removing or substituting missing data and dealing with outliers. After this, the data is split into train and test data. By examining the data, we extract the most important attributes using techniques like principal component analysis (PCA), linear discriminant analysis (LDA). These extracted features are used to train models like Naive Bayes, Random Forest, KNN, etc. using training data and performance is evaluated on test data using techniques such as cross-validation, accuracy, RMSE, precision, recall. The best performing model is selected and subsequently used to classify and recommend crops.

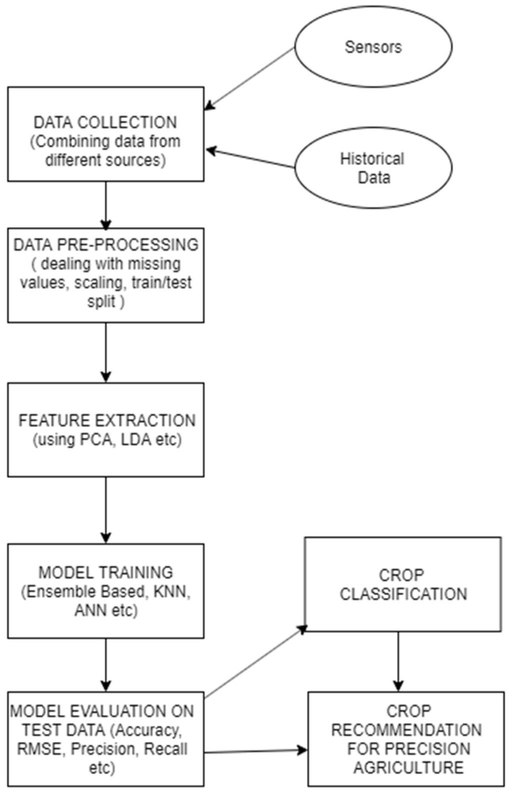

Fig.1 Workflow of a basic crop recommender

V. EXPERIMENT & RESULTS

We have summarized numerous AI techniques which are used to recommend crops. Different models can be better for different crops in terms of accuracy and precision for recommendation and classification. So along the same lines, we compare the performance of four very commonly used models namely Random Forest, Logistic Regression, Naive Bayes and Decision Table to classify the crops into four seasons i.e., Kharif, Rabi, Summer and Whole Year using dataset [23]. We have calculated the performance of these models on the dataset for Uttar Pradesh and Karnataka, states of India in terms of relative absolute error (RAE), root mean squared error (RMSE), mean absolute error (MAE), accuracy and root relative squared error (RRSE) as performance matrices in Table 2 and Table 3, respectively. For both states, Random Forest performs the best and the confusion matrices for Uttar Pradesh and Karnataka for Random Forest are given in Fig.2 and Fig.4. respectively. A comparison of the value of F1 score for algorithms used for both the states is also given in Fig.3 and Fig.5.

$$\text{Accuracy} = \frac{\textit{Number of correctly classified instances}}{\textit{Total number of instances}} \quad \dots (1)$$

$$\text{RMSE}=\sqrt{\frac{\sum_{i=1}^{n}(Predicted_i - Original_i)^2}{\text{Total number of instances}}} \quad \dots (2)$$

$$\text{RAE}=\frac{\sum_{i=1}^{n}|Predicted_i - Original_i|}{\sum_{i=1}^{n}|Original_i - \overline{Original}|} \quad \dots (3)$$

$$\text{RRSE}=\frac{\sum_{i=1}^{n}(Predicted_i - Original_i)^2}{\sum_{i=1}^{n}(Original_i - \overline{Original})^2} \quad \dots (4)$$

$$\text{MAE}=\frac{\sum_{i=1}^{n}|Error_i - \overline{Error}|}{\textit{Total number of errors}} \quad \dots (5)$$

$$\text{F1 Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad \dots (6)$$

Predicted and *Original* in equations (2), (3) and (4) denote the predicted value by the model and the actual value that is observed respectively.

Subscript *i* denotes the index in equations (2), (3), (4) & (5).

In equation (3) and (4), $\overline{Original}$ denotes the mean of all the observed values.

In equation (5), $\overline{Error}$ denotes the mean of all the errors occurring in the values.

The F1 score or the F1 measure is a measure of the accuracy of test data and it strikes a balance between recall and precision. It considers precision and recall values for calculating the score. The best-case scenario is when the F1 score reaches a value of 1 while the worst being 0. It is mostly used when the dataset given has highly skewed output classes.

Precision in equation (6) is the ratio of the number of true positives and total positive class instances the model predicted. The positive class instances include both true positives and false positives. High precision can be seen as a lesser number of false positives predicted by the model. It is also called positive predictive value (PPV).

Recall used in equation (6) is the ratio of the number of true positives and the number of true positives added with the number of false negatives. High recall can be seen as lesser number of false negatives. It is also called true positive rate or sensitivity.

True positive occurs when the model correctly predicts the class as positive. Similarly, when the model correctly predicts the class as negative, it is called true negative.

On the other hand, when the model incorrectly predicts a positive class as a negative class, it is called a false negative. Similarly, when a model incorrectly predicts a negative class as a positive class, it is called a false positive.

Table 2. Performance metrics for different algorithms for the state of Uttar Pradesh

Model	**Accuracy**	**RMSE**	**RAE**	**RRSE**	**MAE**
Random Forest	84.174%	0.2354	0.282	0.5802	0.092
Logistic	78.481%	0.2546	0.393	0.6276	0.129
Naive Bayes	40.59%	0.4854	0.832	1.1964	0.274
Decision Table	80.739%	0.2515	0.420	0.6200	0.138

For Uttar Pradesh, Random Forest correctly classifies the crops with the best accuracy, followed by Decision Table and Logistic Regression whilst Naive Bayes shows the minimum accuracy.

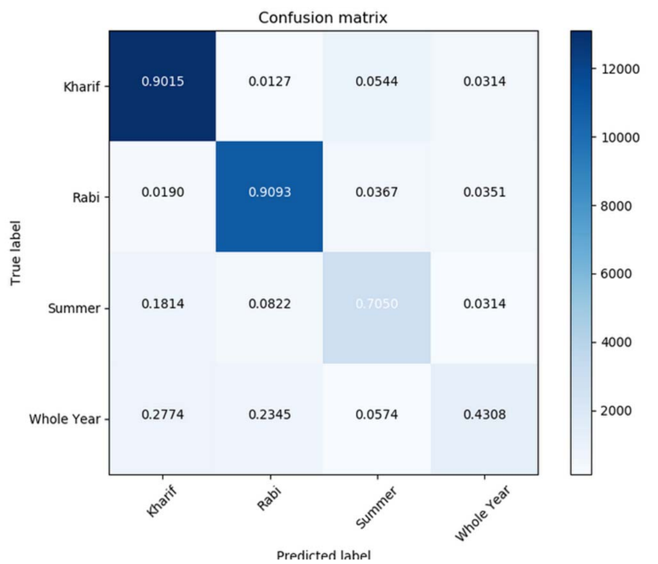

Fig.2 Confusion matrix for Random Forest (Uttar Pradesh)

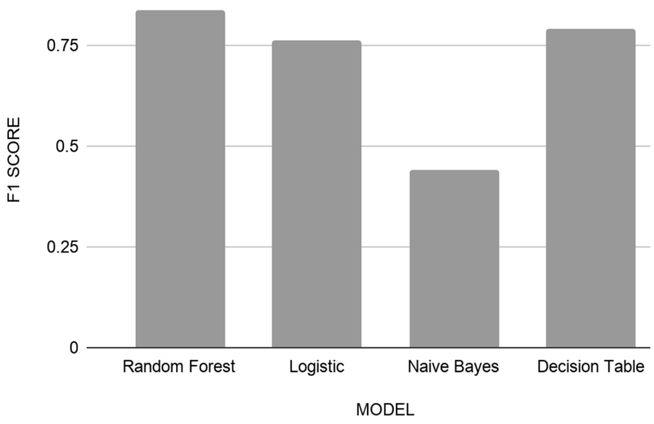

Fig.3 F1 Score for different models (Uttar Pradesh)

Table 3. Performance metrics for different algorithms for the state of Karnataka

Model	Accuracy	RMSE	RAE	RRSE	MAE
Random Forest	75.598%	0.2848	0.376	0.6737	0.134
Logistic	69.354%	0.2978	0.494	0.7044	0.176
Naive Bayes	63.635%	0.3419	0.577	0.8088	0.206
Decision Table	72.327%	0.2910	0.512	0.6884	0.182

For Karnataka, Random Forest correctly classifies the crops with the maximum accuracy followed by Decision Table and Logistic Regression while Naive Bayes shows the least accuracy.

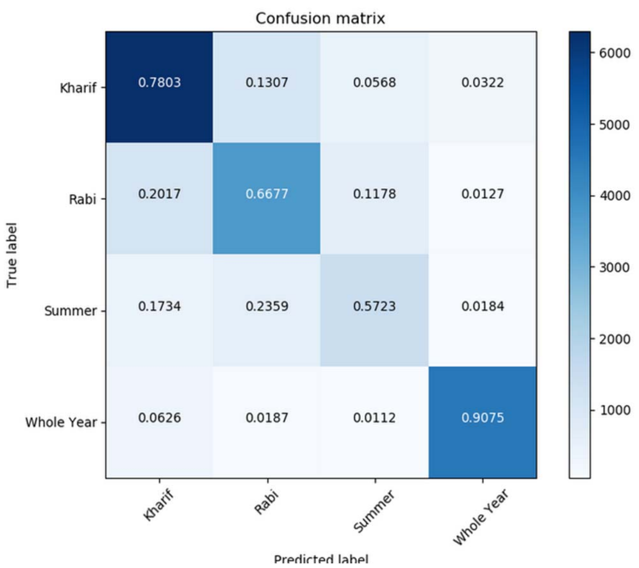

Fig.4 Confusion matrix for Random Forest (Karnataka)

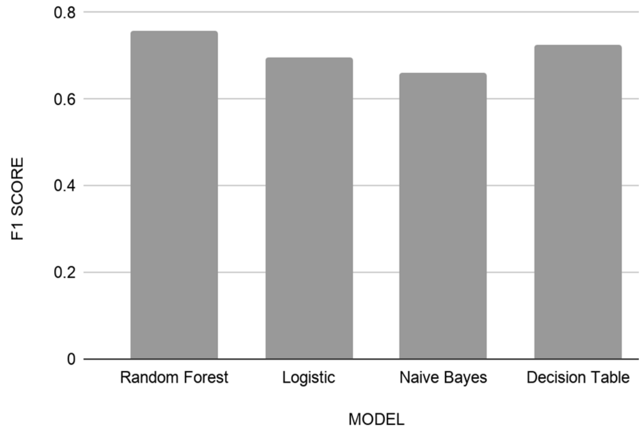

Fig.5 F1 Score for different models (Karnataka)

VI. CONCLUSION& FUTURE WORK

We concluded that there is still a need for an efficient crop recommendation system. Most of the ensemble methods ensure good accuracy and incorporating big data analysis and data mining can be beneficial for a large scale system. What we require is a method to collect data efficiently and process it through a network of sensors. We also need to include other factors while recommending crops such as the economic state of the farmer, availability of fertilizers and topography. Since temperature and rainfall have a huge impact on yield, we also require a robust system that can predict the weather accurately. For future work, we need a special focus on soil properties and nutrients as they are the major factors for crop recommendation. A web/mobile application with an easy-to-use interface is desired, which will make this technology more accessible to the farmers and subsequently lead to more profitable yields.

REFERENCES

[1] S. Singh and N. Singh, "Internet of Things (IoT): Security challenges, business opportunities & reference architecture for E-commerce," *Proc. 2015 Int. Conf. Green Comput. Internet Things, ICGCIoT 2015*, pp. 1577–1581, 2016.

[2] S. Babu, "A software model for precision agriculture for small and marginal farmers," *c2013 IEEE Glob. Humanit. Technol. Conf. South Asia Satell. GHTC-SAS 2013*, pp. 352–355, 2013.

[3] Y. Ren, L. Zhang, and P. N. Suganthan, "Ensemble Classification and Regression-Recent Developments, Applications and Future Directions [Review Article]," *IEEE Comput. Intell. Mag.*, vol. 11, no. 1, pp. 41–53, 2016.

[4] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika, and J. Nisha, "Crop recommendation system for precision agriculture," *2016 8th Int. Conf. Adv. Comput. ICoAC 2016*, vol. 6, no. V, pp. 32–36, 2017.

[5] N. H. Kulkarni, G. N. Srinivasan, B. M. Sagar, and N. K. Cauvery, "Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique," *2018 3rd Int. Conf. Comput. Syst. Inf. Technol. Sustain. Solut.*, pp. 114–119, 2019.

[6] A. Gupta and P. Jain, "A Map Reduce Hadoop Implementation of Random Tree Algorithm based on Correlation Feature Selection," *Int. J. Comput. Appl.*, vol. 160, no. 5, pp. 41–44, 2017.

[7] G. Singh, B. Kumar, L. Gaur, and A. Tyagi, "Comparison between Multinomial and Bernoulli Naïve Bayes for Text Classification," *2019 Int. Conf. Autom. Comput. Technol. Manag.*, pp. 593–596, 2019.

[8] S. Taneja, C. Gupta, S. Aggarwal, and V. Jindal, "MFZ-KNN-A modified fuzzy based K nearest neighbor algorithm," *Proc. - 2015 Int. Conf. Cogn. Comput. Inf. Process. CCIP 2015*, pp. 1–5, 2015.

[9] Z. Mingqi and S. Zhijun, “Research on CHAID decision tree model based on rating of China’s small enterprises,” *2008 Int. Conf. Wirel. Commun. Netw. Mob. Comput. WiCOM 2008*, pp. 1–5, 2008.
[10] S. Umadevi and K. S. J. Marseline, “A survey on data mining classification algorithms,” *Proc. IEEE Int. Conf. Signal Process. Commun. ICSPC 2017*, vol. 2018-Janua, no. July, pp. 264–268, 2018.
[11] Y. Gandge and Sandhya, “A study on various data mining techniques for crop yield prediction,” *Int. Conf. Electr. Electron. Commun. Comput. Technol. Optim. Tech. ICEECCOT 2017*, vol. 2018-Janua, pp. 420–423, 2018.
[12] R. Shirsath, N. Khadke, D. More, P. Patil, and H. Patil, “Agriculture decision support system using data mining,” *Proc. 2017 Int. Conf. Intell. Comput. Control. I2C2 2017*, vol. 2018-Janua, pp. 1–5, 2018.
[13] S. Mishra, P. Paygude, S. Chaudhary, and S. Idate, “Use of data mining in crop yield prediction,” *Proc. 2nd Int. Conf. Inven. Syst. Control. ICISC 2018*, no. Icisc, pp. 796–802, 2018.
[14] R. Patil and V. M. Barkade, “Class-Specific Features Using J48 Classifier for Text Classification,” *Proc. - 2018 4th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2018*, pp. 1–5, 2018.
[15] S. R. Kalmegh, “Comparative Analysis of WEKA Data Mining Algorithm RandomForest, RandomTree and LADTree for Classification of Indigenous News Data,” *Int. J. Emerg. Technol. Adv. Eng.*, vol. 9001, no. 1, pp. 507–517, 2008.
[16] The University of Waikato, “WEKA: Waikato Environment for Knowledge Analysis,” *Proc. New Zeal. Comput. Sci. Res. Students Conf.*, pp. 57–64, 2014.
[17] A. Botchkarev, “Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology,” pp. 1–37, 2018.
[18] M. J. Mokarrama and M. S. Arefin, “RSF: A recommendation system for farmers,” *5th IEEE Reg. 10 Humanit. Technol. Conf. 2017, R10-HTC 2017*, vol. 2018-Janua, pp. 843–850, 2018.
[19] A. Londhe and P. P. Rao, “Platforms for Big Data Analytics : Trend towards Hybrid Era,” *2017 Int. Conf. Energy, Commun. Data Anal. Soft Comput.*, pp. 3235–3238, 2017.
[20] C. Yau and A. Sattar, “Developing expert system with soft systems concept,” *Proc. IEEE Int. Conf. Expert Syst. Dev.*, pp. 79–84, 1994..
[21] J. C. Zhao and J. X. Guo, “Big data analysis technology application in agricultural intelligence decision system,” *2018 3rd IEEE Int. Conf. Cloud Comput. Big Data Anal. ICCCBDA 2018*, pp. 209–212, 2018.
[22] R. Kumar, M. P. Singh, P. Kumar, and J. P. Singh, “Crop Selection Method to maximize crop yield rate using machine learning technique,” *2015 Int. Conf. Smart Technol. Manag. Comput. Commun. Control. Energy Mater. ICSTM 2015 - Proc.*, no. May, pp. 138–145, 2015.
[23] Onkar Kadam, Last updated Sept’ 2017, ‘Crop Data Analysis’ https://www.kaggle.com/onkarkadam/crop-data-analysis