IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

CROP PRODUCTION-ENSEMBLE MACHINE LEARNING MODEL FOR PREDICTION

Sundarraj M[1], Susanth S[1], Vinusuriya K[1], Tharun R[1], Praveen K[1], Karthikeeyan K[2], Rajagopal T K P[2]

[1] Students, Department of Computer Science and Engineering,

[2]Assistant Professor, Department of Computer Science and Engineering,

[1,2]Hindusthan College of Engineering and Technology, Coimbatore, Tamil Nadu, INDI**A**

Abstract- *Data Mining is the maximum plausible method of the prevailing virtual international for reading mass of records units to gain left out relationship. The approach used for the evaluation of statistical records over a time frame is the time collection evaluation. This method is clinical and dependable in forecasting occasions to observe over a period. Probability of manufacturing should continually be expected to the close to perfection with the aid of using time collection evaluation. In this work, meals manufacturing is the thrust for prediction. The distinguished type strategies on this take a look at are the ANFIS (Adaptive Neuro Fuzzy Inference System) and Naive Bayes. In this project, ANFIS and Naïve are the proposed ensemble version used to task the crop manufacturing over a time frame. The parameters used one at a time for prediction of output are the accuracy and the type error. The agricultural sector is the prime occupation of India. Researchers are developed a various scientific technology in the agriculture field for better yield. In this paper, we try to form an model using various machine learning algorithms for better rice production. Crop production and prediction utilizing AI Strategies aims to deliver improved outcomes, but the ensemble model provides better predictive results compared to the individual algorithm. We tried to use combination of symmetric machine learning algorithms to form an ensemb le model for better prediction.*

Keywords- *Climate, RGF (Regularized Greedy Forest), Soil composition, CSM (Crop Selection Method), GBDT (Gradient Boosted Decision Tree), regularization, regression problem.*

I. INTRODUCTION

Before the arrival of pc age, extraction of sample from statistics units become executed best via way of means of guide methods. But now with the incredible improvement of pc technology, series of statistics set, type and garage as splendidly increased. This has made massive alternate in Pattern recognition. In order to discover

specific sample from the massive statistics units, and utility is advanced via way of means of the usage of unique automatic set of rules withinside the area of Data mining. Machine getting to know has been advanced in Data Mining as a version in getting to know idea via way of means of the usage of the pc. Given massive statistics units, prediction of recent units of statistics are advanced the usage of getting to know idea via way of means of this version via way of means of schooling and testing. With the intention of predicting an outcome, growing a version with the item of producing type is popularly known as modeling. The type in statistics mining method is predicting the price of a goal variable via way of means of producing a version primarily based totally on a few attributes express variable. By this method, type of a given statistics is primarily based totally on elegance labels and schooling.

II. CLASSIFICATION ALGORITHM ANFIS

It also assists in identifying the relevance of attributes which significantly affect the crop yield. Its accurate prediction can solve various problems related to crop. We give the the input dataset to the prediction of crop ANFIS. ANFIS are a class of adaptive networks that are functionally equivalent to fuzzy inference systems.. ANFIS uses a hybrid learning algorithm. The membership degrees of each function are computed by using the premise parameter set, namely {a,b,c}. The second layer is responsible of generating the firing strengths for the rules. Due to its task, the second layer is denoted as "rule layer". The role of the third layer is to normalize the computed firing strengths, by dividing each value for the total firing strength. The fourth layer takes as input the normalized values and the consequence parameter set {p,q,r}. The values returned by this layer are the defuzzificated ones and those values are passed to the last layer to return the final output. **Adaptive Neuro-Fuzzy Inference System** (ANFIS) blends advantages of both Artificial Neural Networks (ANNs) and Fuzzy Logic (FL) in a single framework. It provides accelerated learning capacity and adaptive interpretation capabilities to model complex patterns and apprehends nonlinear relationships. In execution time comparison the ANFIS is the highest when compared with naive bayes.

NAIVE BAYES

Based on Bayesian theorem the Naive Bayes classification technique is developed. When the value of inputs is very high, this technique is most suitable. Simple Bayes or Idiot Bayes are the other names of Bayes classifiers. The probability of b instance in class in (P(c), P(x), and P (x|c)) is represented by the Bayes theorem. For the chosen class, denotes the probability of generating b instance. Denotes the probability of the happening of is probability of the happening of d. When the attribute is single, this theorem of classification is done easily. The Naive Bayes classification is done by elaborating this hypothesis for various attribute by the use of the formula where probability of class is denoted by (which generates the value for attribute 1, the value for 2nd attribute is generated by the probability of class indicated.

III. CROP YIELD PREDICTION ANALYSIS

Agricultural data is being produced constantly and enormously. As a result, agricultural data has come in the era of big data. Smart technologies contribute in data collection using electronic devices. In our project we are going to analyses and mine this agricultural data to get useful results using technologies like data analytics and machine learning and this result will be given to farmers for better crop yield in terms of efficiency and productivity. The work will help farmers to increase the yield of their crops. Also implemented Naive Bayes algorithm for finding out the exact crop. Thus, we implemented a system which will predict the crop name and approximate yield in a particular farm. Naive Bayes uses probability of crop being grown in those circumstances. Hence in the following result probabilities are calculated and most probable crops are chosen for further final accumulation.

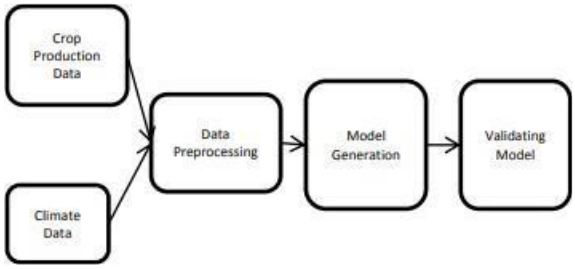

Figure 1 Overview of the Proposed Work

IV. PROBLEM IDENTIFICATION

Entropy primarily based totally combining prediction version for the unit crop yield prediction. The gray forecasting version and radial foundation feature neural community forecasting fashions are blended to boom the accuracy of prediction. This conjunction forecasting version is much less unstable in exercise and extra intuitive and feasible. A singular system gaining knowledge of version that's used to remedy the crop choice problem. In their paintings, a way named Crop Selection Method (CSM) is proposed to perceive the crop choice of a region. The end of the paintings is that the internet crop yield will increase through right choice of vegetation the use of CSM. The results indicate that the accuracy of Prediction is good with moderate computing resources by using the Bayesian approach.

DISADVANTAGES

- Neural network forecasting models are combined to increase the accuracy of prediction. This conjunction forecasting model is less risky in practice and more intuitive and feasible.
- The models were at first trained on the correlation between previous environmental climate and crop yield rate. Finally the models are compared to justify the accuracy.
- The prediction of corn output in a time series prediction model.
- A little amount of mechanization and tractor plough have gained momentum in recent years.

V. SYSTEM IMPLEMENTATION EXPERIMENTAL RESULT

ANFIS, Naive Bayes, are the classification methods used for time series forecast in this paper. Two groups are separated from the data set for training and for testing the algorithms of classification. In order to implement the classification algorithms, the tool used is Rapid miner data analysis. “Read CSV” operator of rapid miner tool is first loaded for Secondary data retained in CSV file. For classification process only a subset of data is selected from the loaded data. To select a subset from original data, “Select Attributes” are utilized by the operator. The chosen subset is then subjected to “X-Validation” operator. It develops the classification model which is validated by the test data. ANFIS and Naive Bayes are implemented for classification by using “X- Validation” operator. The performance of the classification algorithm is evaluated by using the performance operator. Performance evaluation achieved for both the classification algorithms (existing and proposed) are given in Table 1 and Table 2.

DATASET DISCRIBTION

The historical climate data of the CSV type with a record for every line of text alike to the data for a given month of a year. The crop production data of the CSV type with a record for each line of text belongs to the data for a specific month of a year. In the preprocessing stage, though there are many measured parameters available in the raw climate dataset, the less relevant features responsible for the study are ignored and the important features are only taken into account. Thus the two different types of historical records are preprocessed and combined together so that it can be used for this study. The time series historical data over 100 years is taken into study for this experiment. For the purpose of training and testing, separation of the dataset has been made where 60% of the dataset is used for training and 40% of the dataset is used for testing. By using the classification algorithm, the training dataset is used and models generated. The created model is used on the testing dataset in order to test the accuracy. A representation for the environmental demand of evapotranspiration can otherwise be called “Reference Crop Evapotranspiration". This represents both height of the short green crop and sufficient water level in the soil. In other words

it is a reflection of the energy available to evaporate water and the wind available to transport the water vapor from the ground to the lower atmosphere.

Table 1: Execution time Comparison

ALGORITHM	EXECUTION TIME(In min)
ANFIS	8
Naïve Bayes	11

Table 2: Accuracy Comparison

ALGORITHM	ACCURACY (In %)
ANFIS	96
Naïve Bayes	93

VI. SYSTEM TESTING

System Testing is a type of software testing that is performed on a complete integrated system to evaluate the compliance of the system with the corresponding requirements. System testing detects defects within both the integrated units and the whole system. The result of system testing is the observed behavior of a component or a system when it is tested. System Testing is basically performed by a testing team that is independent of the development team that helps to test the quality of the system impartial. The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, subassemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

OBJECTIVES OF TESTING

- Testing is the process of executing a program with the intent of finding an error.
- A successful test is one that uncovers a discovered the error.

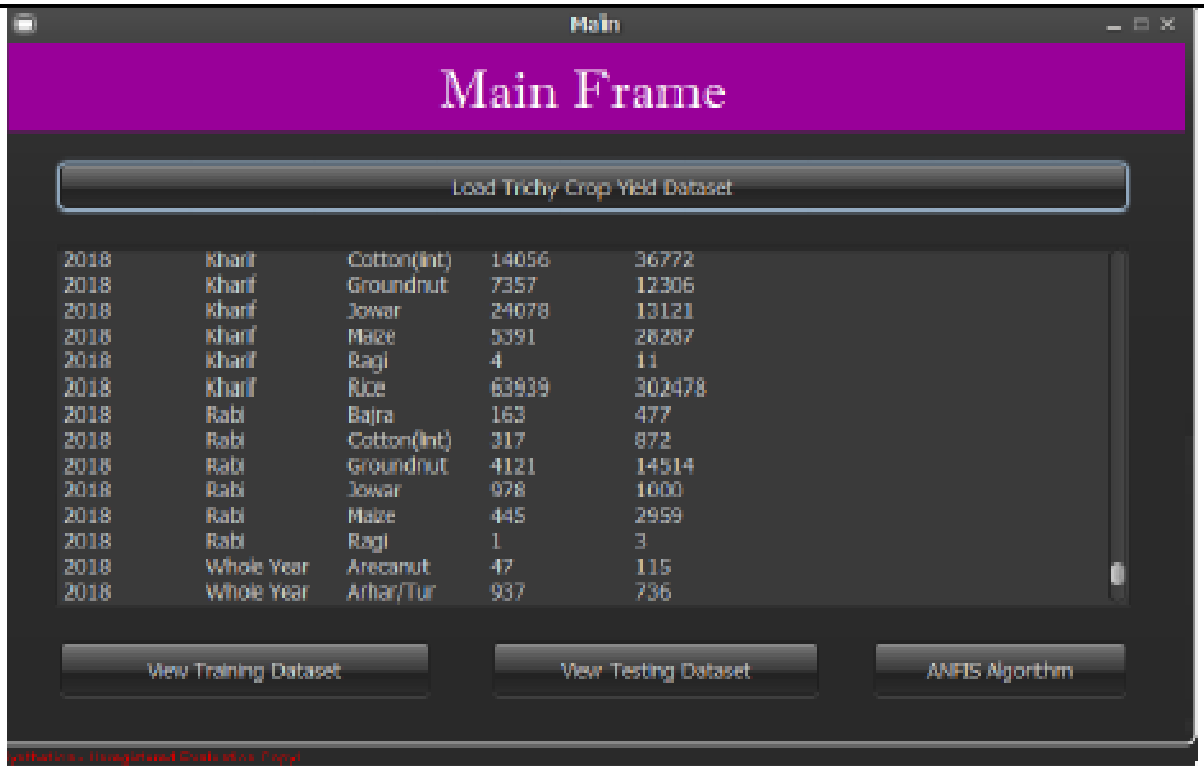

Figure 2 Main Frame Screenshot

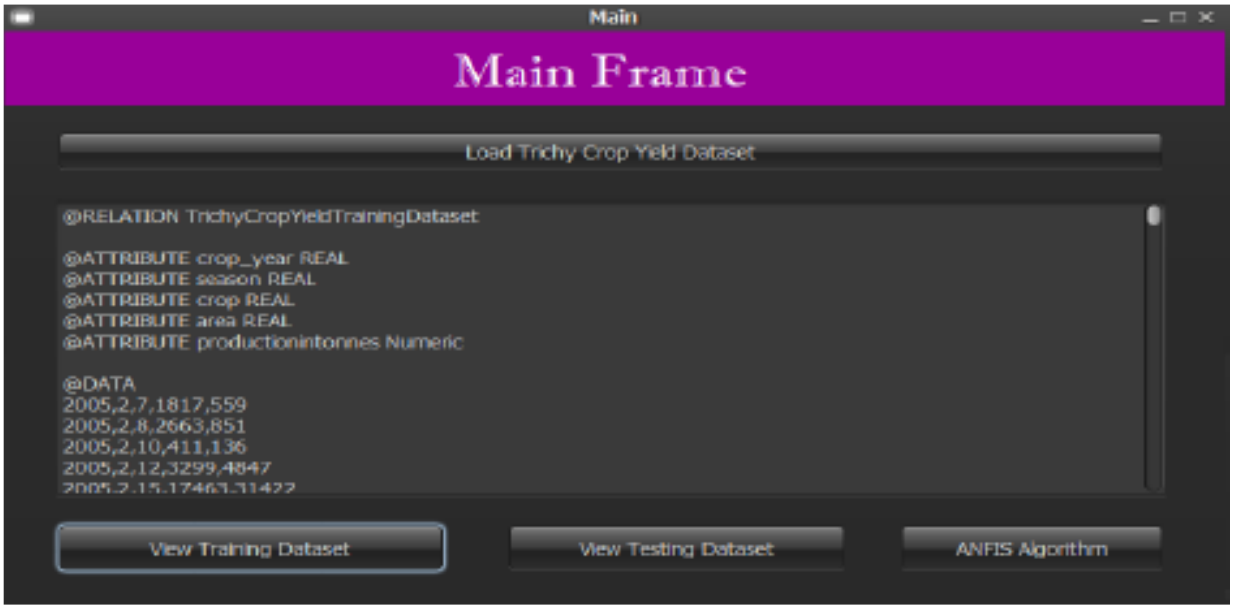

Figure 3 Training dataset Screenshot

Figure 4 ANFIS Algorithm Screenshot

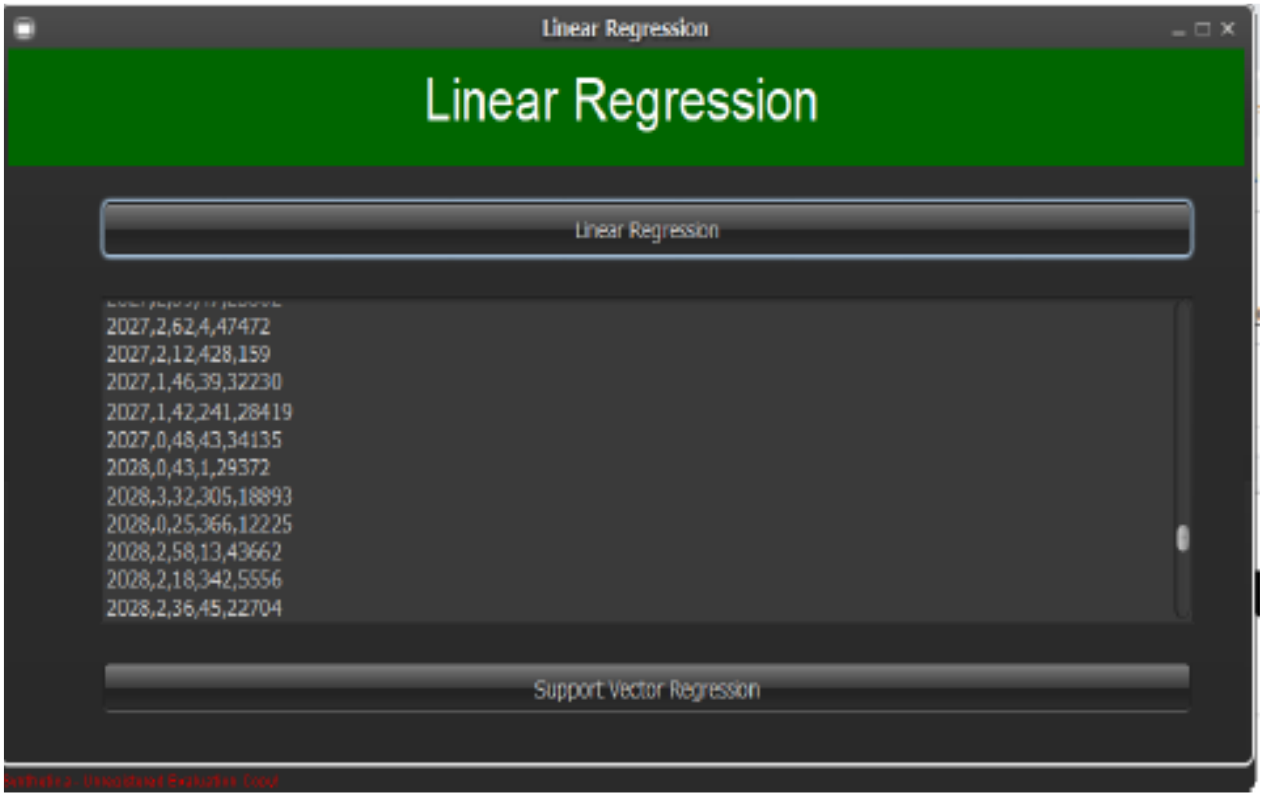

Figure 5 Linear Regression

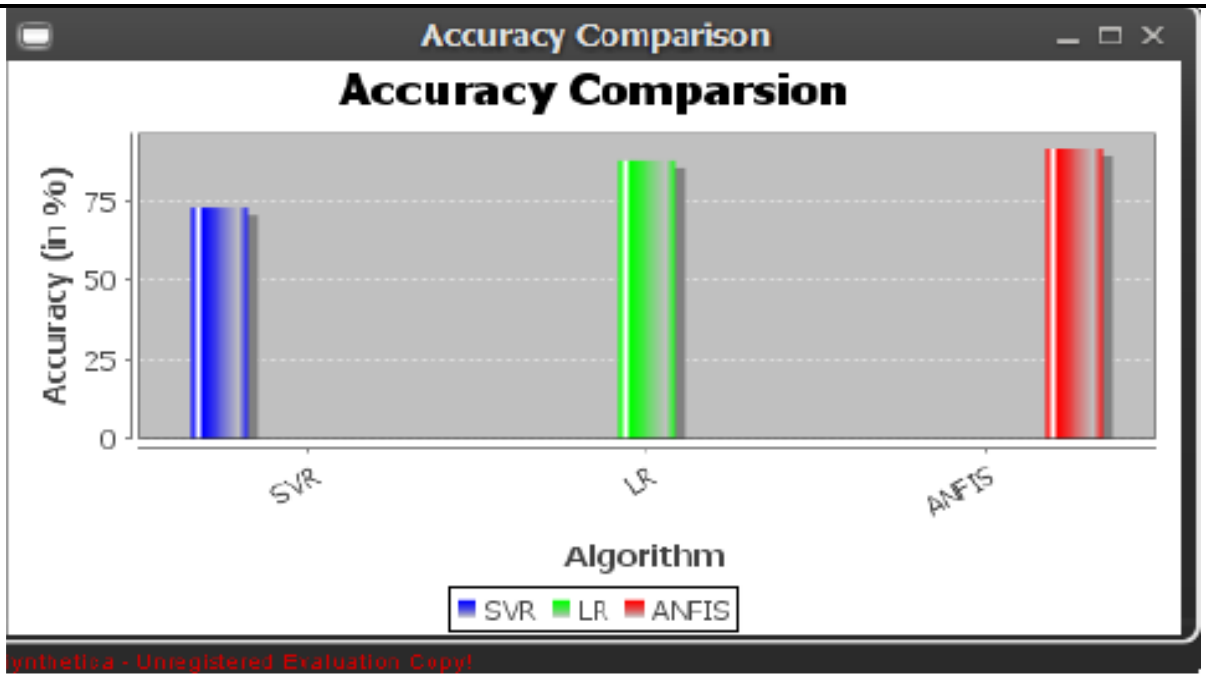

Figure 6 Accuracy Comparison Screenshot

Figure 7 Support Vector Regression

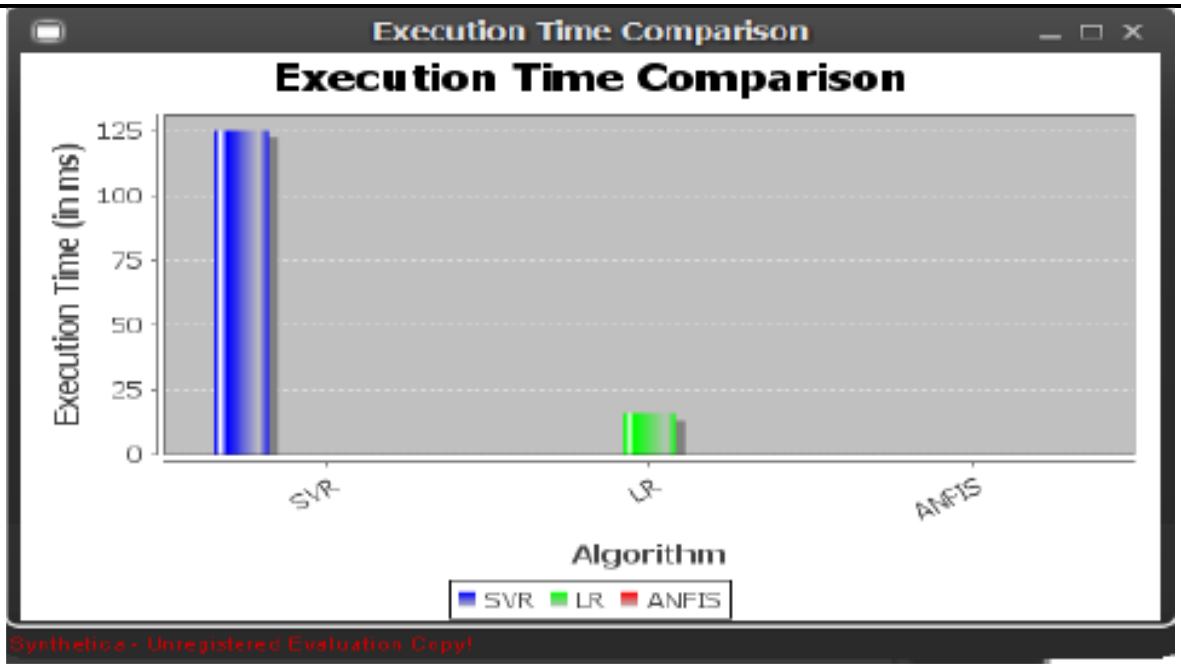

Figure 8 Execution Time Comparison Chart

VII. CONCLUSION

The time series analysis of crop yield prediction is subjected to analysis by the existing techniques ANFIS and Naive Bayes with the proposed technique. It may be concluded from the results that there is good amount of perfection in accuracy of prediction and also good amount of fall in the percentage of accuracy in both the proposed techniques. Future research can enlighten the study whether by changing the technique produces better results or by increasing the input data set for the same technique results change in the findings. Importance of crop prediction is highly needed for agriculture and economy. Continuous research for improving new methods of prediction would be fruitful. This project is a beginning for further research in forecasting.

FUTURE ENHANCEMENT

The obtained result will be helpful for the farmers to know the Yield of the crop so, he can go for the better crop which gives high yield and also say them the efficient use of fertilizer so that he can use only the required amount of fertilizers for that field. This way we can help the farmers to grow the crop which gives them better yield.

REFERENCES

[1] Cheng Zhou, Boris Cule, Bart Goethals “Pattern Based Sequence Classification”, IEEE Transactions on Knowledge and Data Engineering, Vol. 28, No.5, 2016, pp.1285-1298.

[2] B. Tang, H. He, P. M. Baggenstoss and S. Kay “A Bayesian Classification Approach Using Class-Specific Features for Text Categorization”, IEEE Transactions on Knowledge and Data Engineering, Vol.28, No.6, 2016, pp.1602-1606.

[3] P.Samuel Quinan, Miriah Meyer “Visually Comparing Weather Features in Forecasts”, IEEE Transactions on Visualization and Computer Graphics, Vol. 22, No.1, 2016, pp. 389-398.

[4] Y. Chen and Y. Li "Entropy-Based Combining Prediction of Grey Time Series and Its Application”, IEEE International Conference on Intelligent Computation Technology and Automation (ICICTA), 2009, pp. 37-40.

[5] G. Chen, X. Xu, G. Wang and H. Chen "The corn output in a time series prediction model", IEEE International Conference on World Automation Congress (WAC), 2010, pp. 283-286.

[6] A. Dehzangi, K. Paliwal, A. Sharma, O. Dehzangi and A. Sattar "A Combination of Feature Extraction Methods with an Ensemble of Different Classifiers for Protein Structural Class Prediction Problem", IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 10, No.3, 2013, pp. 564-575.

[7] R. Kumar, M. P. Singh, P. Kumar and J. P. Singh "Crop Selection Method to maximize crop yield rate using machine learning technique" IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 2015, pp. 138-145.

[8] M. M. Rahman, N. Haq and R. M. Rahman "Machine Learning Facilitated Rice Prediction in Bangladesh", IEEE Global Online Conference on Information and Computer Technology (GOCICT), 2014, pp. 1-4.

[9] Deepti Gupta, Udayan Ghose, “A Comparative Study of Classification Algorithms for Forecasting Rainfall ", IEEE 4th International conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), 2015, pp. 1-6.

[10] V. B. Nikam and B. B. Meshram, "Modeling Rainfall Prediction Using Data Mining Method: A Bayesian Approach", Fifth International Conference on Computational Intelligence, Modelling and Simulation, Seoul, 2013, pp. 132-13.

[11] T K P Rajagopal, M Venkatesan, 2022, "Energy efficient server with dynamic load balancing mechanism for cloud computing environment", Wireless Personal Communications, vol. 122, no. 4, pp. 3127-3136, (2022).

[12] T K P Rajagopal, M Venkatesan, A Rajivkannan, 2019, "An improved efficient dynamic load balancing scheme under heterogeneous networks in hybrid cloud environment", Wireless Personal Communications, vol.111, no. 3, pp. 1837-1851,(2019).

[13] OM Soundararajan, Y Jenifer, S Dhivya, TKP Rajagopal, "Data security and privacy in cloud using RC6 and SHA algorithms", Networks and Communication Engineering, vol. 6, no.5, pp. 202-205.(2014).

[14] M Kamarunisha, S Ranichandra, TKP Rajagopal, "Recitation of load balancing algorithms in grid computing environment using policies and strategies an approach", International Journal of Scientific & Engineering Research, vol.2 no3, pp. 1-8. (2011).