Classification Models and Hybrid Feature Selection Method to Improve Crop Performance

U.Muthaiah, S.Markkandeyan, Y.Seetha

Abstract— In this paper classification models and hybrid feature selection methods are implanted on benchmark dataset on the Mango and Maize. Particle Swarm Optimization–Support Vector Machine (PSO-SVM) classification algorithm for the selection of important features from the Mango and Maize datasets to analysis and also compare with the novel classification techniques. Various experiments conducted on these datasets, provide more generated rules and high selection of features using PSO-SVM algorithm and Fuzzy Decision Tree. The proposed method yield high accuracy output as compared to the existing methods with minimum Error Rate and Maximum Positive Rate.

Keywords: Classification, Feature selection, PSO-SVM, Decision Tree

I. INTRODUCTION

Starvation is forcing people to cultivate land that is inappropriate for agriculture and which can only be converted to agricultural use through more pains and costs. Every nation is known for its core competence. Agriculture[4] is main source for indian people. Still, it only accounts for 18 per cent of the total is Domestic Product. Due to urbanization, it is going to be a big challenge to yield food for more people with a smaller amount of land and water. The present research primarily on to improve the crop productivity by considering climate change and landscapes. Agricultural area comprises with spatial inconsistency in weather, soil type, and topography.

Agriculture productions makes impact on overall economic and social wellbeing n India. Farmers are backbone of any country economy, because of a large population lives in rural locations and it is directly or indirectly dependent on agriculture for a living. Main source of income comes from farming forms for the farming community.

The requirements for harvesting crop are sufficient water and funds to buy seeds, pesticides, fertilizers, labor etc. The farmers mostly manage the required funds by compromising on other required expenses, and if it is still insufficient they resort to credit from sources such as banks and private financial.

In india, Mango is cultivated in kharif and Rabi sessions. Mango is one of the most cultivated grain crop in India and in Asian counteris. South India consumes more mango than any part of india. Agriculture is the main profession of about 62 per cent of the people in Tamil nadu. Mango is major food of state about 77 per cent of food grain production. Always payment is dependent on the success of the crop. When the crop fails due to several factors, like improper, bad weather pattern, soil type, excessive and premature application of manures, insecticides, adulterated seeds and pesticides etc. The Maize productions also have not kept pace with demand n the recent year. The Maize in india is the third most important food grains after rice and wheat. It is cultivated mainly in kharif sessions and contributes 100 million to agricultural GDP at current prices. Maize can be grown in variety of soils ranging from loamy sand to clay sand.

The farmers are dealing with seasonal variability in short term is very crucial. The relevance and quality agricultural dataset is vital for farmers who requires accurate predictions of crop yield to help make strategic decisions [1]. The combination of Principal Component Analysis (PCA) is for preprocessing and a modified Genetic Algorithm (GA) is used to get crop yield. By reducing number of features, computational cost and time would be reduced [5], [6]. In such way, the classification models for agricultural corps can be efficient and characterization is also improved. The PCA-GA data mining mechanism will be applied for agricultural crops dataset to classify key attribute combinations and characteristics that determine crop performance.The high resolution spatial simulations have been performed by incorporating high resolution datasets with agro ecosystem models. The climate change over an agricultural landscape which describes the Impact on crop productivity and Impact on Soil organic carbon due to combined effect of soil erosion and climate change [2].Datamining techniques[10] like Classification and prediction are very important tasks to find the useful information in order to help farmers by using agricultural datasets. Variety approaches were implemented to build classification and prediction models, including the support vector machines (SVM). Feature subset selection plays important role n building classification systems.

When reduce the number of input features in a classifier produce a good predictive and less computationally intensive model [2]. By selecting

U.Muthaiah, CSE, Sri Shanmugha College of Engineering and Technology, Sankari, Salem-637304, Tamil Nadu, India
(E-mail: muthaiahu54@gmail.com)

Dr.S.Markkandeyan, Sri Shanmugha College of Engineering and Technology, Sankari, Salem-637304, Tamil Nadu, India
(E-mail: drsmkupt@gmail.com)

Dr.Y.Seetha, Dept of Physics, Associate Professor, Aurora's scientific Technological & Research Academy, Hyderabad, Telangana, India.
(E-mail: yseetha10@gmail.com)

appropriate feature subset, classification process can be realized easier. The appropriate parameter setting can improve the SVM classification accuracy. For designing SVM, must select a kernel function, set the kernel parameters and determine a soft margin constant C. Therefore, SVM classifier based on PSO s applied to produce better results by comparing with the existing methodologies. SVM takes set of nput data and predicts, two possible classes comprises the nput, making the SVM a non probabilistic binary classifier. The linear separating hyper plane s found by maximizing the margin between two classes n this space[9]. Therefore, the complexity of the separating hyper plane depends on the nature and the properties of the used kernel. in this paper, the following sections are organized. Second Section explains the Literature survey. Third Section explains proposed methodologies. Fourth Section explains the results and discussion finally.

II. RELATED WORK

Author [3] proposed an expert data mining techniques depends on PCA-Genetic Algorithm, t has applied to describe crops n agricultural field. The technique exemplifies the progresses to classification difficulties by applying preprocessing method called Principal Components Analysis and a improved Genetic Algorithm optimize the process. The fitness process n GA was revised for that reason by means of well-organized distance measures. The PCA and Genetic algorithm mixed data mining method, using different data sets related agricultural filed, creating data mining classification models and create relationships understandable. The outcomes from experiments give an idea about enhanced classification results and produced representations for the corps n agricultural field. The outcome based domain model can give benefits to researchers and farmers n the agricultural filed. The produced classification patterns may also be demoralized and with desire have as a featured into a decision support system[11]. The experiments produced the results by the PCA-Genetic Algorithm was good and enhancing the data mining process and also producing classification models and procedures for agricultural crops characterization. The identified attributes would optimization characteristics of the Genetic Algorithm during the process..

In this paper the author [8] was used as a tool of data mining for doing approximation n crop yield. The study of new data mining methods an The study made assessed new data mining methods and it was implemented to variety of attributes to create when we found new meaningful relations . The efficient data mining methods can be implemented and assesed using suitable data to resolve larger agricultural problems.

It is used to reduce dimensionality without losing correctness. The pca is implemented to find sparse data and creating improved association rules. It is a multivariate methods that produce the data table with observations are described by nter correlated dependent variables. This method aims s transform the data nto

new orthogonal variables which s a principal components[9].

The EPIC method s used to model the climate change impact by the responsive way. Discussed about Carbon dioxide enhancement, management and crop relations. This paper discussed about discussed about efficient algorithm for classification of multiple classes. Better understanding of soil nature could improve yield n farming, biodiversity, reduce fertilizers and better soil management system for private and public sector Final paper s prepared as per journal the template Contents of the paper are fine and satisfactory. Author (s) can make rectification in the final paper but after the final submission to the journal, rectification s not possible. ANN, ID3, the k- means, and the k-NN [11] and support vector machines applied n the agriculture filed for classification. Production scheduling is important n preparing supply chain management. Marketing task would be used to crop production evaluations.

III. PROPOSED METHODS

Mango dataset s taken as training set for the process. In PSO, a particle considered as each solution of the optimization problems. It is a current vaiable (Ppbest(s)) and good positions (Pgbest(s)). PSO is locate the position and velocity as follows

Vide (s+1) = WVi (s) + a1 ∗ q1 ∗ (P$_p$best (s) Xide (s)) + g1 (1)

g1= a2∗q2∗ (P$_g$best (s) − Xide (s))   (2)

Xi (s+1) = Xide (s) + Vide (s)

Where $= 1,\ 2,\ ...N,\ and\ d = 1,\ 2,...D$, It s referred the dimension of solution space. The (1) and (2) equations, the factors *a*1 and *a*2 are larger constants, random numbers are q1 and q2 uniformly distributed the interval [0, 1], *Vide* ∈ Vminimum, Vmaximum], where Vminimum and Vmaximum for maximum velocity to optimize the process. It is to regulate PSO and avoid growing fast. w s parameter for inertia weight to balance the global and local search abilities. Use PSO-SVM Algorithm for the Selection of important features.

IV. RESULT AND DISCUSSION

The Table 1 shown below s the analysis and comparison of different methods, implemented for doing classification on Maize and Mango datasets. The Existing Methodology implemented by using PCA-GA to select the feature and classifier used for the classification, which can be a final decision. But here, a more improved PSO-SVM based Feature Selection method if it is used for selecting the features from that dataset and then Fuzzy Based Decision Tree is implemented as classifier for doing classification on dataset.Authors should consider the following points:

The Methods when implemented on Various

Datasets on Various Techniques it is concluded that the proposed methodology outperforms with more accuracy as compared to other previous methodologies.

The table 1. shown below is the analysis and comparison of different techniques implemented for the classification of Maize and Mango Datasets. The Existing methodology implemented uses PCA-GA for selecting feature and classifier is used for the classification of final decision. But here a more improved PSO-SVM based feature selection method is applied to get accurate results to optimize the performance.

Table 1. Comparison of Accuracy of Various Techniques

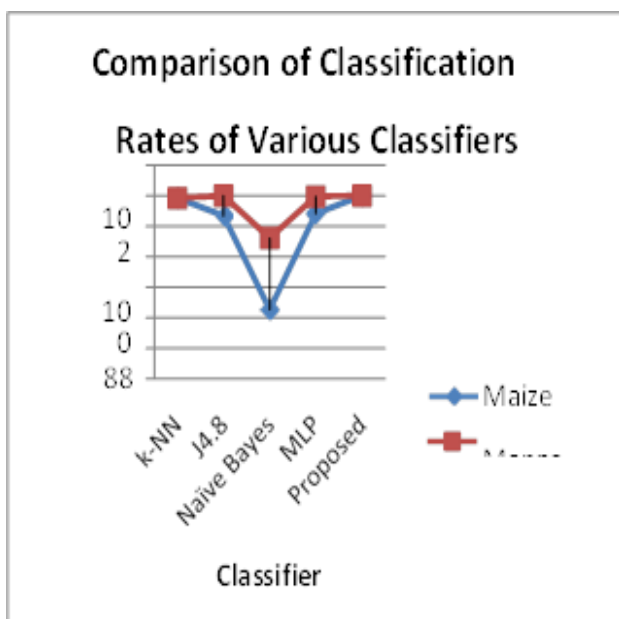| Classifier	Maize	Mango
k-NN	99.87	99.87
J4.8	99.66	98.90
Naïve Bayes	92.53	97.22
MLP	98.82	99.95
Proposed (PSO-SVM)	99.95	99.93

Figure 1. Analysis of Accuracy on Various Dataset

The table below shown the number of reduced features comparison with SVM without PSO and with PSO. In this process, the proposed methods s working good to select suitable features.

MapReduce Based Classification: Mapper function and Reduce function are used. First the mapper method is used to process line by line. This method is processing the dataset and generating many small chunks of data. key/value pair is used for generating ntermediate key/value. Reduce stage is a combination of the shuffling and merging s done n reduce stage. The reduce phase is merging the same key intermediate values. Finally, the output is stored in HDFS (Hadoop Distributed File System).

Table 2. PSO- SVM on Maize Dataset

Algorithms	Original Set of Features	Reduced Features subset	Accuracy (%)
Without PSO-SVM	30	-	80.50
With PSO-SVM	30	6	99.95

Table 3. PSO - SVM approach on Mango Dataset

Algorithms	Original Set of Features	Reduced Features subset	Accuracy (%)
Without PSO-SVM	22	-	82.45
With PSO-SVM	22	4	99.93

Table 4. Various Feature reduction approach for Dataset

Datasets	PCA-GA	k-NN GA	J4.8-GA	Naïve Bayes-GA	Proposed PSO-SVM
Maize	58	4	21	17	2
Mango	41	5	18	26	1

V. CONCLUSION

The result analysis shows the proposed method performance. The proposed method applied to provide more accuracy for the classification using Maize and Mango Datasets. IAS IA result of experiments conducted, the proposed method obtained accuracy rate of 99.95% IBY more generated rules and high selection of features. The proposed method also provides less Error Rate and High Positive Rate. Although the methodology applied here provides efficient results as compared to the other existing techniques, but further enhancements can be done related to the computational time of the methodology as well as reducing the rules generated.

REFERENCES

1. Shang, L; Zhou, Z; Liu, X. Particle Swarm Optimization- based feature selection n sentiment classification.I Soft Computing 2016, 20, 3821-3834. [CrossRef]
2. Feature selection. Available online :(accessed on 2 Januaryi2017) https://en.wikipedia.org/wiki/Feature_selection 3 .Mehreen Naz, Kashif Zafar, and Ayesha khan. Ensemble Based classification of Sentiments using forest Optimiza tion algorithm, 21 May 2019.
3. Drew, J. Operating In a Change Environment NEAR/Drought Reform, Retrieved from (2010).

Retrieval Number: K105209811S219/2019©BEIESP
DOI: 10.35940/ijitee.K1052.09811S219

325

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

4. IJ. Yang and V. Honavar, "Feature subset selection using a I genetic algorithm," IEEE Intelligent Systems TheirApplications,vol.13,no.3,pp.44-49, March/April 1998
5. D. E. Goldberg, Genetic Algorithms n Search, Optimization, and Machine Learning, Reading Menlo Park: Addison-Wesley, vol. 412, 1989.
6. Geraldin B. Dela Cruz, Bobby D. Gerardo, and Bartolome T. I Tanguilig, "Agricultural Crops Classification Models Based on PCA-GA Implementation n Data Mining" International Journal of Modeling and Optimization, Vol. 4, No. 5, October 2014.
7. A. Raoranne and R. V. Kulkarni, "Data Mining: An effective tool for estimation n the agricultural sector," International Journal of Emerging Trends and Technology n Computer Science, vol. 1, no. 2, pp. 75- 79, July-August 2012
8. S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh, "Data mining Techniques for Predicting Crop Productivity –A review article", International Journal of Computer Science and technology, march 2011
9. Dr. D. Ashok Kumar, N. Kannathasan, "A Survey on Data Mining and Pattern Recognition Techniques for Soil Data Mining" IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 1, May 2011.
10. S Usmin, M Arockia Irudayaraja, U Muthaiah. Dynamic placement of virtualized resources for data centers in cloud, International Conference on Information Communication and Embedded Systems (ICICES2014), 2014/2/27, 1-7, IEEE.
11. U Muthaiah, DRS CHITRA. Dynamic Bin Packing for Resource Allocation in the Cloud data center, 2016/3, Volume 1,issue 3, IJORAT.
12. Muthaiah U, Dr.S.Chitra. Predictive Analytics Tiling Methodology Based Data Mining On Farming Crops In Tamilnadu, ISSN: 0976-2876 (Print),ISSN: 2250-0138(Online)Indian J.Sci.Res. 17(2): 89-92, 2018.
13. P Priya, U Muthaiah, M Balamurugan, Predicting yield of the crop using machine learning algorithm, INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY,2018,Volume 7, issue 4,1-7,

VI. AUTHORS PROFILE

Mr. Muthaiah U is currently working as a associate professor at the department of Computer science and Engineering in Sri Shanmugha College of Engineering and Technology, Salem, Tamilnadu, India. He has completed B.Tech IT from Mepco Schlenk Engg., college under Anna University, Chennai, Tamil Nadu and Post Graduate from Anna University Trichy and pursuing part time Ph.D from Anna University Chennai. He is having 13.5 years of experience in Teaching and also looking after corporate relations. He is member of CSI, IEEE and Cyber Security. His Research area include Bigdata analytics, Machine learning, Pest classification in agriculture field, Deep learning. He has over 15 national and international journal publications, book chapters and conference papers. He is also a Guest Editorial Board Member of Journal of Supercomputing (SUPE)-Springer, and the Journal of Big Data (Springer).

Dr.S.Markkandeyan received PhD degree in Computer Science and Engineering from Anna University, Chennai. He is currently Professor in Computer Science Engineering Department at Sri Shanmugha College of Engineering and Technology, Sankari.He has published 7 papers in International Journals and 14 papers in National and International conferences. His research interests include machine learning applications, data mining and web mining, MANET. He is a life member of Computer Society of India, Indian Society for Technical Education.

Dr. Y Seetha has completed her Post Graduation in Electronics from Hyderabad Central University.

She worked as a lecturer at graduate level for 12 years at Kasturba Gandhi Degree and P.G College, Hyderabad. She pursued her PhD in Opto- Electronics sensors from Jawarharlal Nehru Technological University, Hyderabad in the year 2019. Her interests include sensors , optical gyroscopes , Cavity ring down spectroscopy, exhaled breath gas analysis for sensor applications. Her PhD thesis include Quality factor measurements of prism based optical gyroscope using the technique of Cavity Ring Down Spectroscopy for defence and aerospace applications.

Her academic achievements include school first at 10 th std level. She is 5 th rank holder in the P.G entrance conducted by Osmania University and 8 th rank holder in the P.G entrance conducted by Central University, Hyderabad. She is currently working at Aurora Engineering College as an Associate Proffesor.