International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-5, January 2020

Paddy Crop Production Analysis Based on SVM and KNN Classifier

Pankaj Bhambri, Inderjit Singh Dhanoa, Vijay Kumar Sinha, Jasmine Kaur

Abstract: ***In earlier times, the people used to fulfill their own requirements by cultivating the crops in their own land regions. In the economy of a nation, an important role is played by the farming sector. A variety of fungal and bacterial infections infect various crops. Reducing the use of insect killers is a prominent demand of sustainable development. The minimum use of pesticides saves environment and increases the quality of crops. To improve the accuracy of paddy production prediction the KNN is implemented for the paddy production prediction in data mining. The SVM classifieris also implemented which is compared with the KNNclassifier. The presented and earlier classifier will be applied in python and it is expected that accuracy will be improved and execution time will be reduced. It is analyzed that KNN performs well as compared to SVM classifier for the paddy production prediction as per the obtained analytic results.***

Keywords: Paddy Crop Production, SVM Classifier, KNN classifier.

I. INTRODUCTION

Data Mining is easy and effective to process the sorted information. Most of the large organizations use data mining such that the issue of including huge amount of data in warehouses can be resolved by applying simple methods. There are several data mining tools applied in the software for generating and automated analysis process. Furthermore, it is important to use historical data such that new information can be extracted. Thus, the use of software is reduced through the automated analysis process. The time and cost can be reduced using this software. Since the application fields can easily use this technology for analysis, all the complex issues are resolved by it [1]. Thus, the major objective of applying this technique is sorting the data available in unorganized manner. The scientific data, military intelligence, business transactions and satellite pictures are collectively used for generating such large rich source of data.

Manuscript received on January 02, 2020.
Revised Manuscript received on January 15, 2020.
Manuscript published on January 30, 2020.
* Correspondence Author
Dr. Pankaj Bhambri*, Dept. of Info. Tech., Guru Nanak Dev Engineering College, Ludhiana, Punjab, India. E-mail: pkbhambri@gndec.ac.in,
Dr. Inderjit Singh Dhanoa, Dept. of CSE, Guru Nanak Dev Engineering College, Ludhiana, Punjab, India. E-mail: cecm.cse.vk@gmail.com,
Dr. Vijay Kumar Sinha, Dept. of CSE, Chandigarh Engineering College, Mohali, Punjab, India. E-mail: inderjit26@gndec.ac.in,
Er. Jasmine Kaur, Dept. of CSE, IKGPTU, Jalandhar, Punjab, India. E-mail:jasminekaur@gmail.com

From such collective raw data, extracting only the required data is necessary. For data mining, knowledge discovery is prominent. Theiterative process is applied here for extracting important information.

The predictive analytics decides the future of data mining. The terms data mining and data extraction seems like similar. However, there is an important difference between these two terms. In data extraction, the data is obtained from one data source and loaded into a targeted database.

Therefore, the data can be extracted from a source or legacy system for its storage into a standard database or data storehouse.

However, in data mining, the unclear or hidden predictive information is extracted from big databases or data storehouses. Further, data mining as knowledge discovery is used to search patterns in data warehouses [2]. Here, computational methods from statistics and pattern identification are utilized by data mining. Therefore, the nature of data mining is defined by the searching of patterns in data. Different data mining tools and techniques are used to build a predictive analytical model. Large databases are accessed to extract data in primary step. Here, the crop yield needs to be forecasted and the ideal condition needs to be identified such that high yield of paddy production can be performed [3]

K-Nearest neighbor is a lazy learner technique. This algorithm depends on learning by analogy. It is a supervised classification method. This classifier is used extensively for classification purpose. This classifier waits till the last minute prior to build some model on a specified tuple as compared to earlier classifiers. The training tuples are characterized in N-dimensional space in this classifier. This classification model looks for the k training tuples nearest to the indefinite sample in case of an indefinite tuple. Then, this classifier puts the sample in the closest class. This algorithm can be implemented easily. This algorithm performs quickly in case of small data sets. However, this algorithm performs slowly on huge amount of data and big size data. This approach is responsive to the value of k [4]. The performance of the classification model also gets affected by this.

Support vector machines classification model gives good performance on unknown data. The maximum margin classification model is the simplest example of this algorithm. This classification model provides solution of the most fundamental classification issue. This issue is known as binary classification with linear separable training data. This classification model finds hyperplane with the maximal margin. In support vector machine, some slack variables are established to manage the nonlinear separable cases. Some training errors could be handled using this phenomenon. This reduces the effect of noise in training data.

Through the selection of uppermost probability, classification is executed. This classifier involves a penalty metric that permit a definite amount of misclassification. This is mainly imperative for non separable training sets [5].

II. RELATED WORK

UraiwanInyaem, et al., (2018) stated that various investigations had been performed in the domain of data mining.

The application of data mining in farming such as the forecasting of rice productivity was very difficult for cultivators. The main aim of this study was to introduce a system using data mining techniques. In this study, decision tree approach and ANN were compared to predict the production of paddy crop for farmer. With the help of these techniques, the production of paddy crop and its value could be predicted by the farmers. This betide was gainful as it improve the finances of cultivators. A mechanism named CRISP-DM was followed by this research work. Using a dataset of cultivators, the data mining techniques were tested for the classification. Further, these four choices were standardized for accuracy. The tested results related that the good access having more accuracy rate could be plummy to farmers.

K. L. Ponce-Guevara, et al., (2017) reviewed some most imperative factors of farming. These factors included humidity, soil moisture, carbon dioxide and intensity level. These factors influenced the photosynthesis process of plants which affected the production of crop in a greenhouse. For decision making within financial and business applications, these two fields provided huge assistance. In presence of massive data, the pattern identification was the prime focus in this approach. There was no particular control of data analytics through a standard with the help of these technologies and methods. Nevertheless, this study provided a set of algorithms to generate descriptive models and a database for information classification and forecasting.

Luminto, et al., (2017) proposed a new multiple linear regression approach. The key purpose of projected approach was to forecast the time of rice farming. In this study, just two variables were used which were not sufficient for forecasting purpose. By analyzing all variable combinations that generated low RMSE values, the forecasting could be done in some specific areas. A suitable technique called multiple linear regressions was utilized for the prediction of one such problem which was based on multiple dependent variables. This technique provided quicker outcomes as compared to other machine learning methods.

Susanto B. Sulistyo, et al., (2017) presented a new scheme for estimating the nutrient content present within wheat leaves. This new approach was fundamentally a computational intelligence vision sensing scheme. In conjunction with committee machine, several DSELMs were incorporated. GA approach was utilized for their development so that the nitrogen content within wheat leaves could be measured. The simulation results depicted that proposed approach showed better performance regarding quality and processing speed in all stages.

Vladimir M. Bure, et al., (2017) presented a novel algorithm to scrutinize the color features of plants occurring within aerophotos. In addition, several tests were carried out for evaluating the appropriateness of produced model. Various conclusions were obtained using this scheme. Therefore, the environmental condition of field was improved using this approach. In view of the simultaneity of assessment regions, the tests were repeated here. One more covariance scrutiny was carried out for obtained linear regressions. The likelihood of data calibration using test regions and classification of plants was scrutinized importantly.

Qiben Yan, et al., (2017) proposed a scalable and private continual and private continual geo-distance evaluation system. This system was recognized as SPRIDE. The main aim of proposed approach was to provide geographic based services. In this study, fresh and realistic modifications based on data segmentation as well as distance forecasting methods were proposed. A real-time private distance assessment was obtained on the large network of agricultural farms because of the implementation of proposed approach. The proposed approach showed seventeen times better improvement in runtime performance in comparison with earlier approaches according to simulation outcomes.

Yolanda. M. et al., (2017) reviewed the estimation of maize crop production using remote sensing and practical models. In this study, the different growing phases of maize crop were anticipated. Therefore, as per the experiential true field values, improved accuracy was estimated. Approximately 14% of accuracy for LAI based prediction model and 97% of accuracy for NDVI based predictive model was estimated. In order to estimate yield and quantity of corn being produced, this method was applied in different countries like Mexico. For the execution of grain imports strategies related to domestic demand, the government officials utilized these estimations.

Yuichiro Maeda, et al., (2017) described the foreseeing of rice crop which was quite beneficial to plan the rice farming timetable. The implementation of XG Boost described the foreseeing system of rice produce using climatic data, farming data and the zone knowledge of paddy rice farm. The climate knowledge included utmost temperature on daily basis, minimal temperature on daily basis and daylight time. These factors were combined to describe the descriptive variables. When climate data was combined in two time gaps, then the finest accuracy of 74.4% was achieved. This data was combined as planting time to heading time and heading time to maturing time. It was analyzed that the knowledge about climate conditions played an effective role in the foreseeing of paddy rice.

Zhihao Hong, et al., (2016) presented a data-driven approach for providing PA solutions for data gathering and data modeling models. On the basis of same soil as well as climatic qualities collected by sensor node, the soil moisture was forecasted by the structure for future. Within the Illinois past data, the evaluations were made for proposed technique as machine learning algorithms required big data size. During the prediction of soil moisture for upcoming two weeks, about 15% of error rates and 95% of correlations were obtained among the forecasted values and the authentic values of nine diverse places.

III. PROPOSED WORK

On the basis of earlier historical data, Predictive analytics predicts the future events by implementing sophisticated techniques such as machine learning. Different methods like filtering, data correlation etc. are used for the collection and transformation of historical data.The methodologies of paddy production prediction analysis are based onclassification.In base paper, prediction analysis is performed using data mining techniques such as J48, LAD tree and LWL. In the existing approach technique KNN and SVM classifier is applied for the prediction analysis.

The KNN classification method can be applied over the small datasets, it is analyzed that when the size of the dataset get increased accuracy get reduced.This classifier can be divided the data into two parts training data and testing data. Then processing the training set for the classification. In this study we also implemented SVM classifier for paddy crop prediction.

There are different steps of research methodology can be explained:

- **Data Pre-Processing:**Data Pre-processing is the first step of research methodology. In this work, dataset is authorized from the UCI repository. In this step, the input value data is being noise free. This means that error values are segregated from the dataset.
- **Feature Extraction:** The second step is Over-fitting should be avoided as an important objective of feature selection. The performance of model can be improved. This process can be gives the relation between each and every features of the data with the predestined target data set.
- **Classification:** The SVM and KNN classification can be implemented on given dataset. The paddy production prediction will be originate by classification stage. SVM is mainly imperative for non separable training data sets. Some slack variables are established to manage the nonlinear separable cases. Some training errors could be handled using this phenomenon. This classifier waits till the last minute prior to build some model on a specified tuple as compared to earlier classifiers. KNN algorithm can be implemented easily. This algorithm performs quickly in case of small data sets. However, this algorithm performs slowly on huge amount of data and big size data. This approach is responsive to the value of k.

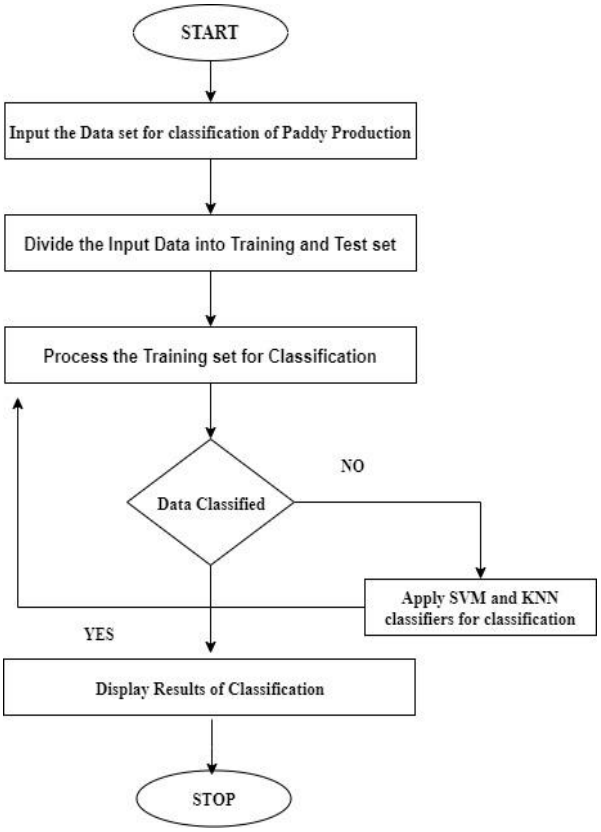

Figure 1: Proposed Flowchart

IV. EXPERIMENTAL RESULTS

This research work is implemented in Python and the outputs are appraiseby encounter the introduced and existing classifiersin terms of different attributes.

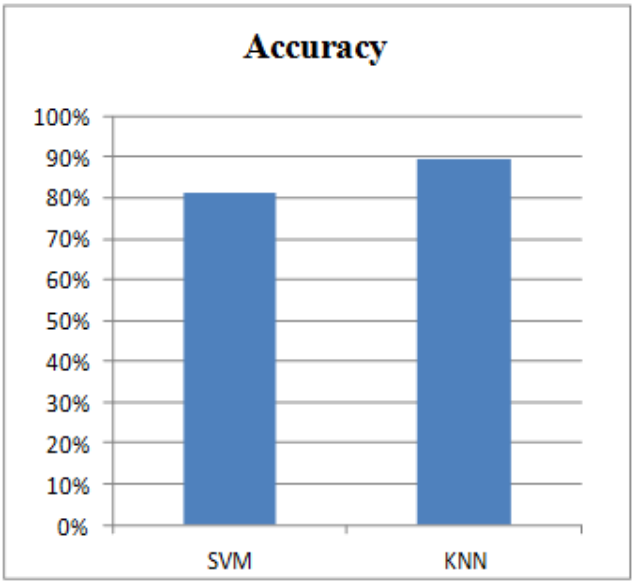

Figure 2: Accuracy Comparison

Figure 2 shows the accuracy of SVM and KNN classifier is compared for the paddy crop production prediction analysis. It is analysed that accuracy of KNN is high as compared to SVM.

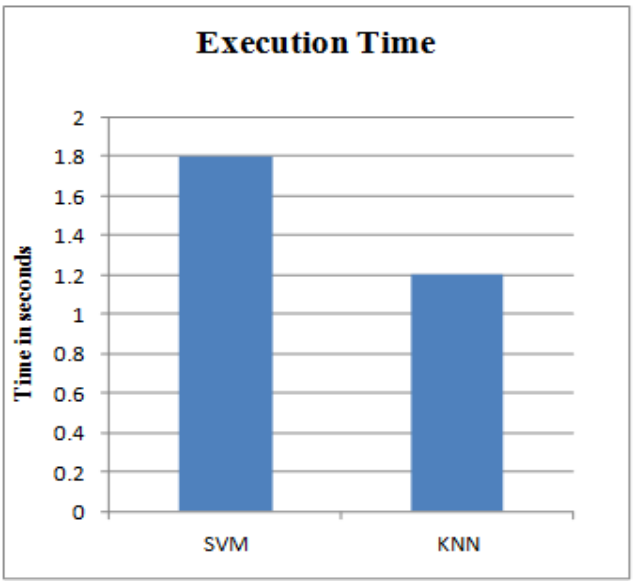

Figure 3: Execution Time Comparison

Figure 3 shows that the execution time of SVM and KNN classifier is compared for the prediction analysis. It is analysed th/at execution time of KNN classifier is less as compared to SVM.

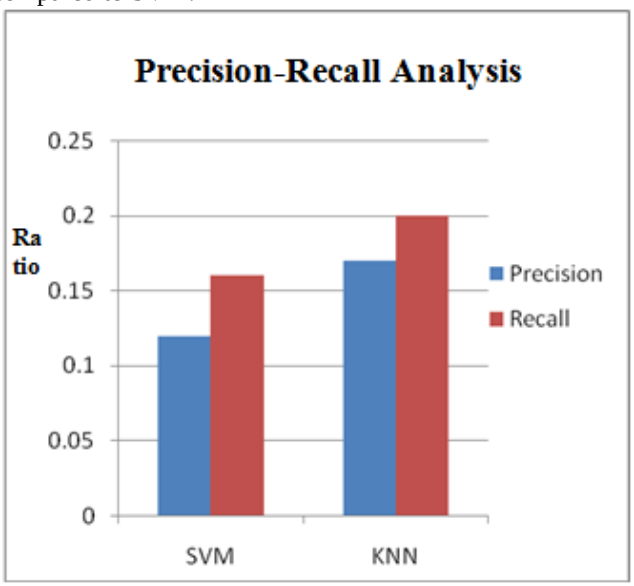

Figure 4: Precision-Recall Comparison

Figure 4 shows the comparative analysis of SVM and KNN in terms of their precision-recall values. The output shows that in comparison between SVM and KNN, the values of precision-recall are higher for KNN.

V. CONCLUSION

Data mining is the process of extraction of relevant data from the huge dataset. The process used for the conversion of data records into set of classes is called classification. The classification is generally of two types. These types are supervised classification and unsupervised classification. The customers do not predict set of cases in unsupervised classification. In classification, a certain result reliant on a given input is predicted. The earlier present data should be extracted for predicting results. Based on this data the records are classified. There are two categories of data sources. These categories are training set and test set. The already classified data is included in the training set. In order to perform classification, this data is utilized as a reference. The attributes are used to predict outcomes. Further, the test data is provided to the algorithm. In this work, the SVM algorithm and KNN algorithm is implemented for the prediction analysis. The performance of SVM and KNN is compared in terms of accuracy, execution time, precision and recall. It analyzed that KNN has high accuracy and less execution time as compared to SVM.

REFERENCES

1. G Rub, “Data Mining of Agricultural Yield Data: A Comparison of Regression Models”, In Conference Proceedings, Advances in Data Mining – Applications and Theoretical Aspects, P Perner (Ed.), Lecture Notes in Artificial Intelligence 6171, Berlin, Heidelberg, Springer, pages: 24-37, 2009.
2. Rajagopalan B, Lall U, “A K-Nearest Neighbor Simulator for Daily Precipitation and Other Weather Variables”, Water Resources Research, volume 35, Issue 10, pages: 3089-3101, October 1999.
3. Yethiraj N. G., “Applying Data Mining Technqieus in the Field of Agriculture and Allied Sciences”, International Journal of Business Intelligents, Volume 01, Issue 02, December 2012.
4. Tripathi S, Srinivas V V, Nanjundiah R S, “Downscaling of Precipitation for Climate Change Scenarios: A Support Vector Machine Approach”, Journal of Hydrology, pages: 621-640, November 2006.
5. Camps-Valls G, Gomez-Chova L, Calpe-Maravilla J, Soria-Olivas E, Martin-Guerrero J D, Moreno J, “Support Vector Machines for Crop Classification using Hyper Spectral Data”, Iberian Conference on Pattern Recognition and Image Analysis, pages: 134-141, January 2003.
6. UraiwanInyaem, “Construction Model Using Machine Learning Techniques for the Prediction of Rice Produce for Farmers”, InIEEE 3rd International Conference on Image, Vision and Computing (ICIVC), June 2018.
7. K. L. Ponce-Guevara, J. A. Palacios-Echeverrıa, E. Maya-Olalla, H. M. Dom´ınguez-Limaico, “GreenFarm-DM: A tool for analyzing vegetable crops data from a greenhouse using data mining techniques (First trial)”, In IEEE Second Ecuador Technical Chapters Meeting (ETCM), October 2017.
8. Luminto, Harlili, “Weather Analysis to Predict Rice Cultivation Time Using Multiple Linear Regression to Escalate Farmer’s Exchange Rate”, InInternational Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA),August 2017.
9. Susanto B. Sulistyo, Di Wu, WaiLok Woo, S. S. Dlay, and Bin Gao, “Computational Deep Intelligence Vision Sensing for Nutrient Content Estimation in Agricultural Automation”, IEEE Transactions on Automation Science and Engineering, Volume: 15, Issue: 3, July 2017.
10. Qiben Yan, Hao Yang, Mehmet C. Vuran, Suat Irmak, “SPRIDE: Scalable and Private Continual Geo-Distance Evaluation for Precision Agriculture”, InIEEE Conference on Communications and Network Security (CNS), October 2017.
11. Vladimir M. Bure, Olga A. Mitrofanova, “Analysis of Color Characteristics of Plants Using Aerial Photography”, Constructive Non-smooth Analysis and Related Topics (dedicated to the memory of V.F. Demyanov) (CNSA), May 2017.
12. Yolanda. M. Fernandez-Ordoñez, J. Soria-Ruiz, “Maize Crop Yield Estimation With Remote Sensing and Empirical Models”, In IEEE International Geoscience and Remote Sensing Symposium (IGARSS), July 2017.
13. Yuichiro Maeda, TaichiGoyodani, ShunsakuNishiuchi and Eisuke Kita, “Yield Prediction of Paddy Rice with Machine Learning”, In Int'l Conf. Par. and Dist. Proc. Tech. and Application, July 2017.
14. Zhihao Hong, Z. Kalbarczyk, and R. K. Iyer, “A Data---Driven Approach to Soil Moisture Collection and Prediction”, In IEEE International Conference on Smart Computing (SMARTCOMP),May 2016.