Wheat Production Analysis based on Naive Bayes Classifier

Jasmine Kaur
CSE Department
Guru Nanak Dev Engineering College,
Ludhiana, India

Dr. Pankaj Bhambri
IT Department
Guru Nanak Dev Engineering College,
Ludhiana, India

Kapil Sharma
CSE Department
Guru Nanak Dev Engineering College,
Ludhiana, India

Abstract: ***The most important culture being followed in India since ancient times is agriculture. The crops were cultivated by the people in ancient times within their own land areas such that they could fulfill their own requirements. India is a farming nation. Crop production analysis is one of the applications of prediction analysis. This study is related to paddy production. To improve accuracy of the paddy production, the hybrid classifier will be designed based on k mean clustering and Naive Bayes classifier. The presented and earlier algorithms will be applied in python and it is expected that accuracy will be improved with reduction in execution time. The performance of SVM, KNN and Naïve Bayes is compared for the wheat production prediction. Naive Bayes is the best classifier for the wheat production prediction as per the obtained analytic results.***

KEYWORDS: Paddy crop Production, Naive Bayes Classifier, SVM Classifier, KNN classifier.

I. INTRODUCTION

The technique in which the hidden patterns are identified from datasets of large size and including massive volume of raw data is called data mining. Since this technology provides only the necessary data, it is helpful in saving time. The applications like military or medical fields are among the most common data mining applications. In almost every application, huge amount of data is being generated that must be analyzed and sorted regularly such that it can be stored in arranged form [1]. To ensure that the collected data can be used in other applications, this analysis is important. Initially, collected data is referred as raw data. Information relevant to one particular topic is available here and data mining is applied to extract it from other unnecessary data. This approach can be known as data archeology, knowledge mining or knowledge extraction by various authors. The satellites and computers are used to collect this large amount of data. Since there is different kind of data available, it includes computer and mass digital storage [2].

Within every field, huge amount of data is being generated regularly. Thus, new technologies are being developed such that this data can be stored in a better way and can be sorted accordingly to be used in near future applications. The technologies such as structured databases and Database management system (DBMS) have been applied for handling such problems [3]. These technologies are applied in applications were large datasets are involved and some important information needs to be extracted from them. There are several statistical trends and techniques that range from predictive modeling to machine learning to perform prediction analysis. The predictions related to unknown future events can be made can be generated to process and analyze the historical data using data mining. It is possible to recognize the various risks and opportunities of a certain business aspect of prediction analysis by exploiting the patterns identified in historical business data [4]. For providing risk or threat assessment, the relationships among various factors are capture. Important decision making steps are used such that the business can be guided. For predictive modeling and forecasting, it is important at times to describe prediction analysis. Pathogen that is known as an agent is responsible for causing paddy crops disease. The diseases are seen on the leaves or stems of plant in most of the scenarios. Thus, an important role is played by the successful cultivation of paddy crops by recognizing the plants, leaves and stems. Further, the percentage of disease incidence, the symptoms of pest attack and types of diseases are also identified here. It is important to ensure that a diseased image of paddy crop leaf is processed [5]. Also, among various diseases, the currently identified disease needs to be classified.

Finally, for the particular disease, possible treatments are provided by the farmer. Artificial Neural Network is only termed as neural network is an arithmetical system or computational system. This classifier is dependent on biological neural system. Generally, it is a flexible model that alters its configuration on the basis of exterior or interior information [6]. This information streams through the network for the period of learning. Support vector machines are mainly binary classification algorithms. This is a classification model which is obtained from statistical learning theory. Genetic algorithm are dependent on natural genetics. These genetics offer powerful exploration abilities in complex spaces and hence provides a suitable approach to issues which require capable and effectual search procedures. Every answer is attained through an encoding/decoding method [7]. This mechanism enables the representation of solution as a chromosome.

II. RELATED WORK

A. M. P. B. Samarasekara, et al., (2018) stated that prevalent surplus of Sri Lankan paddy farming was rice fodder. The rice straw had not applied to an effectual value addition [9]. However, it could be transformed into a high value produce since it contained large quantity of fiber. Thus, rice straw was advantageous in several ways. It was used to discover different types of rice fodder for providing an adequate fiber produce. The results of proposed technique were compared with the produce of traditional three-step wet chemical technique for the validation of proposed technique. The presented approach showed good accuracy level for all four classes of rice straw. The proposed techniques showed less than 5% percentage variations for all four classes of raw straw. The rice fodders of Suwandel class provided maximum fiber produce.

AshkanNabavi-Pelesaraei, et al. (2018) proposed a new scheme for predicting paddy production. In this scheme, two AI methods were applied [8]. These methods were identified as artificial neural networks (ANNs) and adaptive neuro fuzzy inference system (ANFIS). These two approaches were incorporated for generating an energy output and climate effect forecasting of paddy produce in Iran. For predicting energy output, ANN model with 12-6-8-1 configuration was found finest. The tested results depicted that ANFIS approach proved to be extremely advantageous to the managers during the prediction of energy output and environmental indicators of farming production models. This approach provided high computational rapidity as well as accuracy.

M. R. S. Muthusinghe, et al., (2018) presented a paradigm which was focused on intelligent cultivation theory for rice crop [10]. The presented model followed some modules such as a forecasting unit for predicting rice crop and a forecasting unit for predicting paddy requirement .In this study, two machine learning algorithms had been utilized to develop forecasting models. About 78% of accuracy had been provided by crop forecasting module for the train score with 0.04 MSE value. On the other hand, this model also provided about 75% of test score with 0.11 MSE value. In future, more tests would be executed for improving the precision level of presented forecasting system. Moreover, in future, larger data sets would be utilized.

UraiwanInyaem, et al., (2018) compared decision tree approach and neural network approach (ANN) to predict the production of rice crop for cultivators [11]. With the help of these techniques, the production of rice and its selling price could be predicted by the cultivators. This phenomenon was quite advantageous as it increased the income of farmers. A procedure named Cross-industry standard process for data mining (CRISP-DM) was followed by this research work. Using a dataset of farmers, the machine learning algorithms were tested for the classification of model prototype. Four alternatives were utilized to measure the performance of proposed approach. Further, these four alternatives were standardized for accuracy. The tested outcomes depicted that the finest approach having maximum accuracy rate could be beneficial to cultivators.

Abishek.B, et al., (2017) stated that the prediction of effective rainfall and harvest water was extremely difficult [12]. Some aspects like temperature and humidity should be recognized for providing a detailed and careful scrutiny. In order to predict quantity of rainfall and crop water needs of some particular area, this approach was implemented. Different problems occurring in irrigation of crops within some specific area were also eliminated using this methodology.

A. K. Mariappan, et al., (2017) stated that rice was the popular foodstuff of Tamil Nadu over the years. Various varieties of rice were produced by this state. Seventeen river basins distributed all through this state were the main source of water sprinkling [13]. The data of river basin provided support in farming alternation, planning and plant selection. The result of crop production was based on different features like soil, environment, compost and irrigation. All these factors were important for farming. In this study, rice crop and its production in this state was analyzed. A model that directed the forecasting of rice production on the basis of certain parameters was structured in this study. This model

influenced the production of rice crop. This model was able to facilitate officials for making truth based decisions.

III. PROPOSED WORK

The approach which is used to foresee future probabilities from current data is known as prediction analysis. The methodologies of prediction analysis are dependent on clustering and classification. In base paper, prediction analysis is performed using neural network. The classification algorithm accepts clustered data as input. This algorithm divides dataset into two segments called testing and training. The data is classified into various classes with the help of KNN classification model. The k-mean clustering algorithm computes central points by using numeric mean of the whole dataset. This can decrease the accuracy of prediction analysis. In case of difficult data set, the establishment of association among the attributes of the dataset becomes tricky. In this study, the KNN classification model is implemented. This classifier is used to classify wheat produce in various classes. The classification accuracy could be enhanced by replacing this classification model with some other classification model.
Different stages of research procedure are explained:

- **Pre-Processing:** Pre-processing is the initial stage of research procedure. In this process, dataset is occupied from data gathered from the UCI warehouse. In this stage, the input data is being cleaned. This implies that missing values are separated from the dataset.
- **Feature Extraction:** The second stage implements feature extraction process. This process establishes relation among each feature of the data with the predefined target. Implementation of feature extraction methodology will make feature identification simple.
- **Classification:** The Naive Bayes classification model will be implemented on the given data in final stage. The yield of prediction will be generated through classification process. Naive Bayes is a subset of Bayesian decision theory which is growing popular these days. This algorithm is implemented in important applications as it needs minimum storage and contains quick training procedure. This implies that this approach can be applied on big data sets. The interpretation of this classification process is extremely simple. Even the untrained clients using this algorithm can understand classification procedure.

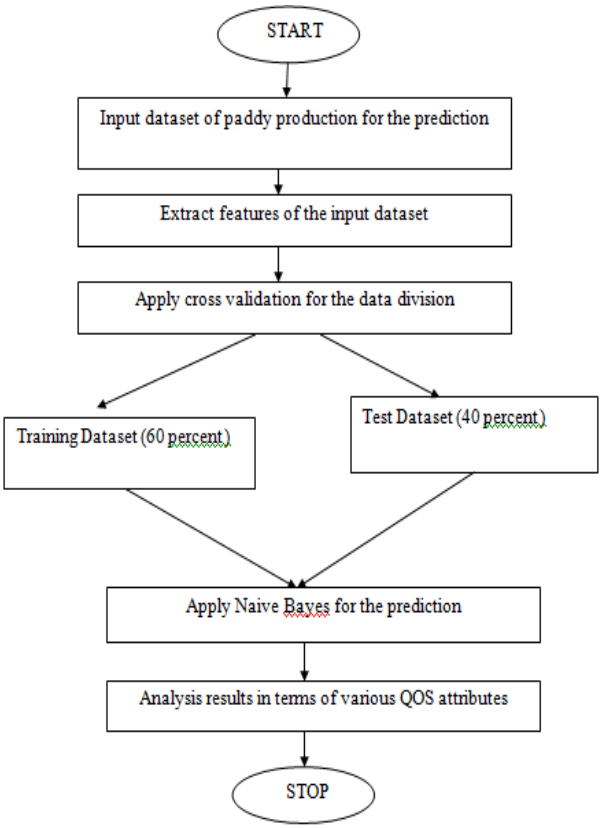

Figure 1: Proposed Flowchart

IV. EXPERIMENTAL RESULTS

The proposed research is implemented in Python and the results are evaluated by comparing proposed and existing algorithms in terms of different parameters.

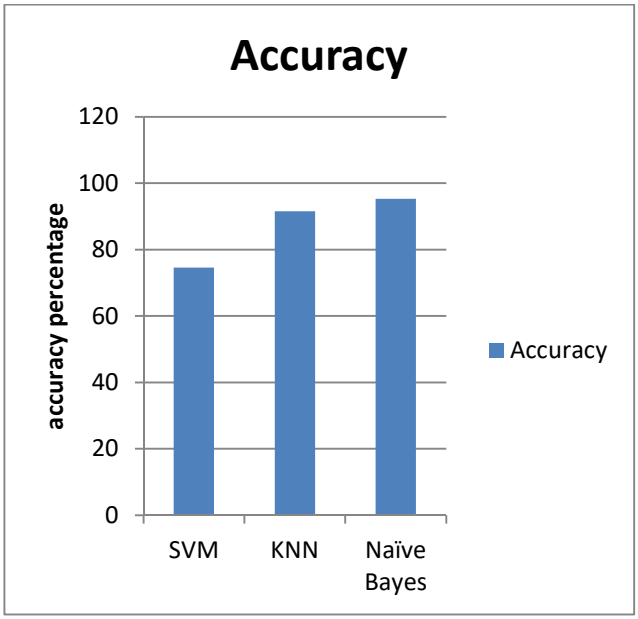

Figure 2: Accuracy Comparison

Figure 2 shows the comparative analysis of accuracy of three classifiers known as SVM, KNN and Naive Bayes in terms of predicting the paddy production. The results show that in comparison to the other two classifiers, the Naive Bayes classifier has maximum accuracy.

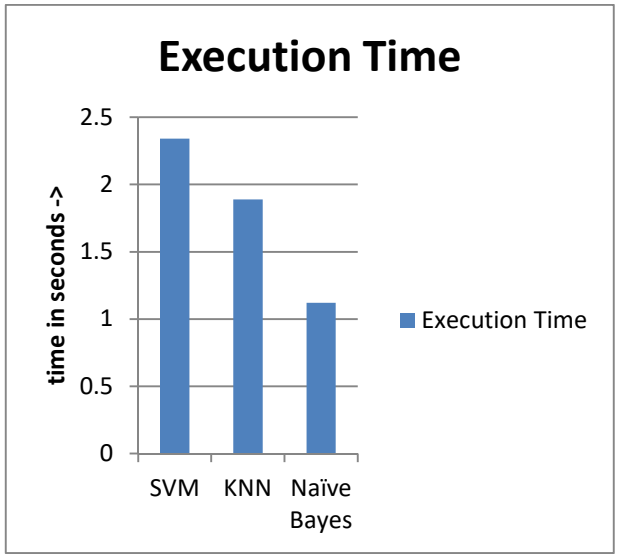

Figure 3: Execution Time Comparison

Figure 3 shows the comparative analysis of Naive Bayes, SVM and KNN classification models. The results show that in comparison to the other two classifiers, Naïve Bayes provides less execution time.

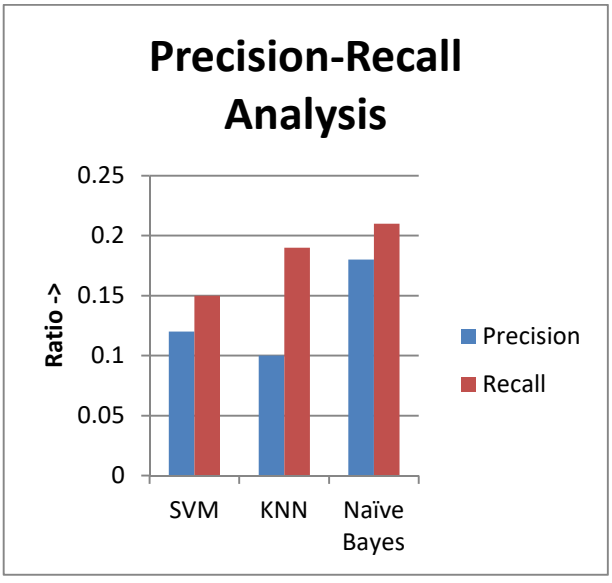

Figure 4: Precision-Recall Comparison

Figure 4 shows the comparative analysis of SVM, KNN and Naive Bayes in terms of their precision-recall values. The results show that in comparison to the other two classifiers, the values of precision-recall are higher for Naive Bayes.

V. CONCLUSION

The technology of data mining technique identifies relevant information from rough dataset. By calculating the similarity patterns within the input dataset, same and different data is grouped. In earlier times, the people used to fulfill their own requirements by cultivating the crops in their own land regions. Therefore, the living survival of certain group of people completely relies on the cultivation. The human beings, animals and birds depend upon the natural crops cultivated on lands. A healthy and welfare life is led on by consuming the greenish products cultivated on the lands. To analyze the data of various viewpoints and summarize it into important information, data mining is applied. This work aims to predict the paddy production using Naïve Bayes classifier. The SVM, KNN and Naive Bayes classifier are compared at the end. It is seen here that the paddy production is predicted here with approximately 95.34% accuracy by applying Naive Bayes algorithm.

REFERENCES

[1] 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.

[2] D Ramesh, B Vishnu Vardhan, “Analysis of Crop Yeild Prediction Using Data Mining Techniques”, *IJRET: International Journal of Research in Engineering and Technology, Volume: 04 Issue: 01,* Jan-2015.

[3] DakshayiniPatil, Dr. M .S Shirdhonkar, “Rice Crop Yield Prediction using Data Mining Techniques: An Overview”, *International Journal of Advanced Research in Computer Science and Software Engineering,* Volume 7, Issue 5, May 2017.

[4] AnithaArumugam, “A predictive modeling approach for improving paddy crop productivity using data mining techniques”, *Turkish Journal of Electrical Engineering & Computer Sciences,*pp- 4777 – 4787, October 2017.

[5] Noor Ayesha, “A Study of Data Mining Tools and Techniques to Agriculture with Applications”, *International Journal of Trend in Research and Development*, pp: 1-4, March 2017.

[6] Nishchol Mishra, Dr.SanjaySilakari, “Predictive Analytics: A Survey, Trends, Applications, Oppurtunities& Challenges”, *International Journal of Computer Science and*

Information Technologies, Vol. 3, Issue 3, pp- 4434-4438,July 2012.

[7] 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.

[8] Ashkan Nabavi-Pelesaraei, Shahin Rafiee, Seyed Saeid Mohtasebi, Homa Hosseinzadeh-Bandbafha, Kwok-wing Chau, "Integration of artificial intelligence methods and life cycle assessment to predict energy output and environmental impacts of paddy production", In*Science of the Total Environment,*Volumes631–632,pp-1279–1294, August 2018.

[9] A. M. P. B. Samarasekara, M. P. A. Nanayakkara, W. G. A. Pabasara, D. A. S. Amarasinghe, L. Karunanayake, "Novel Thermogravimetry Based Analytical Method for Cellulose Yield Prediction of Sri Lankan Rice Straw Varieties",In*Moratuwa Engineering Research Conference (MERCon),* Pages: 185 – 190, June 2018.

[10] M. R. S. Muthusinghe, Palliyaguru S. T. W. A. N. D. Weerakkody, A. M. HashiniSaranga, W. H. Rankothge, "Towards Smart Farming: Accurate Prediction of Paddy Harvest and Rice Demand", In*IEEE Region 10 Humanitarian Technology Conference (R10-HTC),*December 2018.

[11] UraiwanInyaem, "Construction Model Using Machine Learning Techniques for the Prediction of Rice Produce for Farmers", In*IEEE 3rd International Conference on Image, Vision and Computing (ICIVC),* June 2018.

[12] Abishek.B, R.Priyatharshini, AkashEswar M, P.Deepika, "Prediction of Effective Rainfall and Crop Water Needs using Data Mining Techniques", In *IEEE International Conference on Technological Innovations in ICT For Agriculture and Rural Development (TIAR 2017),*April 2017.

[13] A. K. Mariappan, J. Austin Ben Das, "A paradigm for rice yield prediction in Tamilnadu", *IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR),* Pages: 18 – 21, April 2017.