Automated Modular Data Analysis and Visualization System with Predictive Analytics Using Machine Learning for Agriculture field

Ahish. N [1],Shashikala H K[2], Bharath N[3]
Department of Computer Science and Engineering
School of Engineering and Technology, Jain University
Bangalore Rural, India
[1]ahishnagaraj@gmail.com, [2]shashi.hk85@gmail.com, [3]nbharath96@gmail.com

***Abstract*- Economy of an India is majorly depending on growth of agricultural yields, and its allied agro industry products. Prediction of agricultural yield growth is a most difficult for the agriculture departments across globe. The agricultural yields growth is depending on several factors. In this paper historical data is analyzed and a predictive model was designed. Several Regression models such as linear model, multiple linear model and nonlinear models were tested for an effective prediction, or for forecasting the agricultural yield for a variety of crops. Along with this the crop trade for local farmers is a very complicated and tedious task and can get easily mislead by the system we are proposing helps them to analyze the crop availability and also according to market prices can be able to predict various characteristics of the trade. The proposed method is capable of producing the visual representation after data analysis and provides the prediction results in a visual format. And also the unstructured data analysis is implemented in the system. In the proposed method, the pre-processed input data will be sent to perform a descriptive analysis and a predictive analysis. In the descriptive analysis, the data is analyzed and the summary of the analysis is given as the output.In Predictive analysis, there are steps to be considered for the analysis. At the end summary of predicted results are given as output and summary of both descriptive analysis and predictive analysis is given as final report in visual format.**

Keywords—Agriculture; Machine Learning; Agriculture Industry; Prediction Models; Descriptive Analysis; Predictive analysis.

I. INTRODUCTION

Agriculture is a backbone of India's economy, so it needs a mere support. India has billions of population; out of it around 70% of a total population is living in the rural area. In the India, agriculture is lacking institutional attention like: bank loan support, farmer welfare schemes, and farmers suffer from disasters like ground water level depletion in the rural areas, change in climate, difficulties in predicting monsoon, drought, floods, so farmers are migrating towards the urban areas to search of better jobs.

Agriculture is an important sector which feeds every individual. But the farmers are least taken care compared to other sectors. However, the time has come for the technologies to change it. Since many new problems are arising every day, it's a high time to use emerging technologies to find the solution.

The term data analysis begins with the analysis and the predictive system that is proposed consists of different modules used and each module used for specific purpose like analysis of an agricultural production and weather over the years and detailed visual representation of its progress and current status with appropriate bar graph, bar charts, etc. And also the predicted results from historical data using machine learning algorithms with accuracy.

In agricultural sector, crop prediction is very important and it is measured as an amount of production per unit area in cultivation land. The main reason for less productivity in agriculture are, human factors like lack of training for formers and lack of efficiency in formers, usage of traditional cultivation mechanisms, soil problem, pests, crop diseases, and feeble cattle's, irrigation facilities are inadequate, monsoon unreliability and crops marketing is improper so on. However, to understand the economic development of any country, it is very much required to monitor the crops yield condition. Crops yield is having a direct impact on annual national and international economies, and production prediction will play a important role in food management. Forecast of the agricultural production and weather conditions before harvest will be very much useful for the farmers to plan their activities further. It is very much required to do planning in advance, formulation for importing or exporting decisions, crop procurement, and structure of the price and crop distribution.

Varieties of statistical model were developed by researchers based on the time-series data such as weather variables and crops yield that will suggest pre-harvesting process. Major factor affecting the crops productions are, area availability and weather conditions. The crops simulation models will be helpful for estimation of crops production which is expressed as, function of weather condition and the choice of crops management practice. To develop the better probability model for predicting crops outcome, a predictive

This is an open access article under the CC BY-SA 4.0 license (https://creativecommons.org/licenses/by-sa/4.0/)

model may be used. It is a process which will create, test and validates the model.

There are four different stage sin the predictive data modeling methods:

1. Descriptive Analysis
2. Treatment of Data
3. Data modelling
4. Performance estimation.

One of the predictive data modeling method is Regression Analysis. This method will observe the relation between the dependent variables and the independent variables. This method will be helpful for forecasting through the time-series data and finds the underlying effect among these variables. This technique analyses data, and uses data points to fit a line or curve, and helps for minimizing distance of corresponding data points from the line or the curve.

II. LITERATURE SURVEY

Predictive analysis uses the past information to understand the likely occurrence of the future. For example, what is likely to happen next or what is a customer likely to buy next. Sophisticated models and machine learning are crucial in this area to perform the inductive reasoning needed. Finally, the most difficult type of analytics is Prescriptive analytics. Prescriptive analysis provides an insight on what can be done for increasing the probability of a desired outcome occurrence. Predictive and prescriptive analytics are much more prevalent now than previously. There have been both software and hardware technological advancements which enables more complex analytics to be performed. There are also many more sources of data with high level digitization across the globe, and with the Internet of Things taking shape. These factors together with an exponential improvement in both software and hardware, that a lot of data is available for analytics today than before.

Agricultural growth depends on various parameters(like Nitrogen, Rain, water, Weather and characteristics of the soil, rotation of the crops, Surface temperature, soils moisture etc). The decision tree induction technique is used for analyzing the productivity of the crop due to the variety of climate parameters [2]. An end-user has benefits from the derived rules from Decision Tree method findings. The Regression models, like linear model, multiple linear and nonlinear models were used for predicting effectively or for forecasting the efficient agricultural yields for variety of crops [1].

The ARMA (autoregressive–moving-average) model was proved as more efficient time-series model for effective prediction of crops yield than the smoothing methods like damped trend linear exponential, double exponential and simple exponential methods [3]. The spatial dependencies could be unified into a statistical representation for crops yields [1]. The Bayesian model has increased the flexibility; it will also help to improve the prediction process than the existing least-square method.

The SVM-based prediction model has a capability of implementing several feature-selection techniques from an annual report for 10 years. Also it indicated that the frequency threshold of a document will reduce the feature space effectively and preserves the accuracy of equal classification [7].

III. PROPOSED SYSTEM

The proposed method is capable of producing the visual representation after data analysis and even prediction Results in a visual format. The advanced Machine Learning Algorithms like random forest, Decision Trees and Gradient Boosted Regression Trees can be used. The unstructured data analysis is implemented in the system.

Predictive capability of the system will be improved by more training of the system. System would be designed with different modules such as student analysis module, agriculture data analysis module and so it would be convenient for user.

In the proposed method, Input data given to the system will be pre-processed. Pre-processed Data will be sent to perform a Descriptive Analysis and a Predictive Analysis. In Descriptive Analysis, the data is analyzed and the summary of the analysis is given as the output.In Predictive analysis, there are steps to be considered for the analysis.

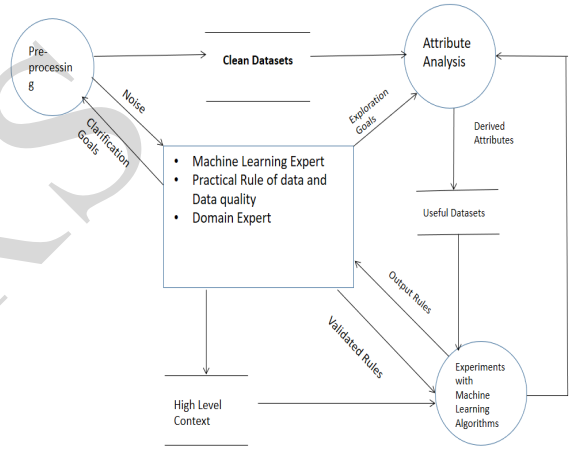


Fig 1. System Architecture

Data for analysis is prepared by splitting the data into training and test Data, then the data Model is generated. The Model generated in deployed to predict the values. The results are collected and evaluated. Finally the evaluated results are compared with the expected value if satisfied, and then stop the analysis, if not satisfied, Repeat from the steps. At the end summary of predicted results are given as output and both summary of descriptive analysis and predictive analysis is given as final report.

In Descriptive Analysis the data count for each district is found out and more data is needed to train the machine learning model for predictive analysis for crop year details seasonal crop details.

After all the analysis and prediction steps is succeeded the results, summaries and plots are displayed on the UI output tab

This is an open access article under the CC BY-SA 4.0 license (https://creativecommons.org/licenses/by-sa/4.0/)

as well as a detailed report is generated using feature in R Markdown. The report can be generated into a html or a word document. The markdown file (.Rmd) file which is later knit into html or a word document. R packages knit and markdown are imported and used to archive the report.

IV. LIMITATIONS OF CURRENT WORK

i. This manner of implementation of systems in terms of modules is rarely found.
ii. The current system only produces the visual representation after data analysis and not the predicted results in a visual format.
iii. The Machine Learning Algorithms like linear regression is used for predictive analytics in most of the systems but these are not that reliable in terms accuracy so advanced random forest for supervised and clustering for unsupervised would be more reliable.
iv. The portfolio generation exists only in few systems which are not that advanced.
v. The unstructured data analysis is not available for current systems.
vi. The Visualization of data is not interactive in nature.

V. CONCLUSION

To conclude final output of the implementation will result in a system with predictive capabilities to determine future values. The proposed method states that the Machine Learning models like linear regression, multiple regressions and logistic regression provide near accurate predictions. And also categorical values can also be predicted and also K-fold cross validation increases the performance of model and provides much more accurate results. These kind of predictive models are very much required to increase the efficiency in Indian agricultural practices. Here only dataset for State of Karnataka is considered. Similarly, other states in India can be considered for prediction.

The same model and automated system can be provided for other fields like Student progress analysis and also small scale industry data analysis only with minimal changes in model and user interface.

REFERENCES

[1] Xin Ying Qiu, Padmini Srinivasan & W Nick Street, “Exploring the forecasting potential of company annual reports,” proceedings of the American society for information science and technology 43(1), pp.1–15,2006.
[2] S. Veenadhari Dr. Bharat Mishra Dr.C D Singh, -‘Crop Productivity Predictive Modelling using decision tree algorithms’, Vol.27– No.7, August 2011, International Journal of Computer Applications(0975–8887).
[3] ‘Agriculture yield prediction using predictive analytic techniques’, 2016 2 nd International Conference on Contemporary Computing and Informatics (ic3i)
[4] Askar Choudhury, James Jones ‘Crop yield prediction using time series models’, Pg.No:53- 68,Journal of Economic and Economic Education Research,Volume 15,Number 3, 2014.
[5] Kao, S Poteet, ‘Text mining and natural language processing-introduction for the secial issue’, SIGKDD Explorations, 2004, vol. 7, Issue. 1, pp.1-3.
[6] Ramesh V, Parkavi, P, and Yasodha P. ‘Performance -analysis of data mining techniques for placement chance prediction-’, International Journal of Scientific & Engineering Research, 2(8),2011,
[7] Cortez, P., & Silva, A. M. G(2008). ‘Using Machine Learning and Text mining to predict school student performance’.
[8] Ayham Omary, Ahmad Wedyan, Ahmed Zghoul, Ahmad Banihani, Izzat Alsmadi (2012) An interactive predictive system for Weather Forecasting .
[9] Zhongnan Zhang, Weili Wu, Yaochun Huang , ‘-Mining Dynamic Inter dimension Association Rules for Local scale Weather Prediction’, , Proceedings of the 28th Annual International Computer Software and Applications Conference (COMPSAC'04), 2004.
[10] Mark Holmstrom, Dylan Liu, Christopher Vo, Machine Learning Applied to Weather Forecasting, Stanford University (December 15, 2016)

IJRS

This is an open access article under the CC BY-SA 4.0 license (https://creativecommons.org/licenses/by-sa/4.0/)