2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC)

A Methodology for Crop Price Prediction Using Machine Learning

1st G Thapaswini
Department of CSE
Saveetha School of Engineering
Saveetha Institute of Medical and Technical Sciences
Chennai, India
thapaswini7@gmail.coms

2nd M. Gunasekaran
Department of CSE
Saveetha School of Engineering
Saveetha Institute of Medical and Technical Sciences
Chennai, India
gunasekaranm.sse@saveetha.com

***Abstract*—Every nation on the globe depends heavily on agriculture. Numerous issues with agriculture still need to be resolved, such as yield recommendations and crop price forecasting based on market value. This might boost farmers' livelihoods and increase yield as well. Modern technologies of today have also been demonstrated to be capable of producing an automated advanced application for a diverse range of purposes. Machine learning models namely, decision tree and neuro-evolutionary algorithms are used in the proposed work to provide an improved crop productivity and pricing estimate. The paper provides a comprehensive analysis on the machine learning based techniques/algorithms used in the agricultural segment. The outcomes of the proposed system were demonstrated and illustrated how far it is efficient. As an extension to this work, an automated price recommendation system using genetic algorithm shall be designed in the future.**

Keywords—Farming, Crop yield, price recommendation, machine learning, decision tree, neuro-evolutionary algorithm.

I. Introduction

Being one of the major sources of food, agriculture is one of the most pressing societal issues. We all know that the Indian economy is based on agriculture. In India, agriculture is a significant industry. The nation's 1.3 billion inhabitants are fed by agriculture, which occupies more than 60% of its land. Both animal and plant cultivation is referred to as agriculture. India's civilization developed as a result of agriculture. In order to grow crops, we need soil. Soil is an important element in agriculture. A healthy soil is necessary for effective food production. It offers support, water, oxygen, and vital nutrients to the roots. All plants utilized in food production are grown on soil, which also serves as the system's base. Various types of soil are present in India. Numerous countries currently suffer from food shortages or scarcity as a result of growing populations [1]. An increasing population, unpredictable weather, soil degradation, and an environmental change all together necessitate solutions to ensure timely and consistent crop production and output. Additionally, it must help increase the sustainability of crop production [2]. These conditions imply that evaluating land, safeguarding crops, and forecasting agricultural production are most essential to the yield of food worldwide [3]. The nation's economic decision-makers must therefore rely on an accurate crop production projection in order to obtain practical exports and imports recommendations for enhancing the safety of the country's food supply.

However, predicting agricultural production is difficult because of many complex aspects. In general, a variety of factors influence crop production, including terrain, soil grade, pest problems, genotype, the performance and availability of water, weather patterns, harvesting technique, as well as other factors [3] and [4]. Progressions and methods for predicting crop yields are fundamentally nonlinear and time-specific [5]. These tactics are especially complicated because they take into account a wide variety of intersected characteristics that persist described and impacted by outside influences and non-arbitration [6], [7]. Farmers used to forecast crop output based on their past experiences and reliable historical data and then make important cultivation decisions in accordance with the prophecy. Furthermore, in latest decades, the emergence of innovative technologies, encompassing crop simulation process using machine learning, had proved to predict production greater precisely, in combination with the capability to identify a significant amount of information utilizing high-performance processors [8-11]. Currently, several studies show that using machine learning algorithms has a larger potential than using classical statistics [12] and [13]. In the domain of artificial intelligence known as machine learning, computers can be trained without specific coding. Such strategies surpass farming techniques that further can be both non-linear as well as linear, by ensuring a substantial predictive accuracy [14]. The tactics are taken from the learning process used by the agricultural machine-learning system. These techniques necessitate performing a firm job while being trained and using training materials. The model assumes that when the training phase has been completed successfully, the data should be tested. Over the past 15 years, a number of insightful, A number of published novel investigations and analyses on agricultural crop yield along with oil palm production forecasting. Paper [15] offered a complete analysis of applications for remote sensing for palm oil farming methods, covering yield prediction, tree numbering, activity recognition, age estimates, disease and pest detection, aboveground biomass and carbon production assessment, including age assessment. The conceivable research gaps are identified and offered concepts for feasible solutions. Unfortunately, the evaluation did not specifically include the algorithms employed to forecast palm oil production. The authors looked at the primary techniques applied recently to create official statistics, including remotely sensed, surveys, including their interaction with meteorology, managerial, or some other data [16]. They also discussed the possibility for research to enhance current crop production forecasting approaches and the prospect for prediction ambiguity in their research. This article has explored numerous agricultural crop yield forecasting techniques throughout a broad geographic region, however it has not included machine learning techniques those are frequently applied to predict production of crops. This article might also not be beneficial to studies with respect for exact crop yields forecasting methods for a particular crop. [17] offered a great systematic review that examined a variety of characteristics and prediction systems. Instead of critical analysis, research gap examination, and recommendation, the essay focuses more on material abstraction. An evaluation of

2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC) | 978-1-6654-9111-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICMNWC56175.2022.10031852

the measure of nitrogen levels employing machine learning [8]. They arrived to the conclusion how cost-effective agriculture solutions might be produced through the quick growth of detection and ML technology. An assessment of crop management forecast machine learning models primarily focused on weather patterns [15]. It advises undertaking a thorough search for further crop yield parameters. Additional investigation was carried out in [18], in which the researchers focused at the body of research on the production of palm oil though a physiologic plant. This analysis was conducted to have a comprehensive overview of the causes of the productivity deficiencies in palm oil. An investigation on the machine learning techniques in agricultural productions were presented [19]. In order to conduct their assessments, the books on land management, crop monitoring, animal monitoring and control, and water management are incorporated. An analysis on the techniques deployed to determine fruit ripeness further to forecast production and harvesting period with the best accuracy were illustrated [20]. According to the discussions above, a thorough review paper that attempts to address the inadequacy of retrospective studies is still required. Therefore, the two main contributions of this work are to recommend crops that produce higher yields and to recommend crop pricing depending on a variety of parameters in order to improve farmers' economies.

II. Literature Survey

All economies are based on agriculture. Agriculture has long been regarded as the primary and dominant traditions that exits everywhere. There are several techniques to boost and enhance crop productivity and excellence. Data mining can also be used to forecast crop prices. In essence, data mining is the method of examining data from a variety of aspects and turning it into helpful knowledge. Predicting crop prices is a major agricultural concern. Every farmer tries to determine what proportion of the price he anticipates will really be paid. In the past, price forecasting was done by examining farmers' prior experience with a certain crop. For making decisions on the forecasting of crop prices, accurate knowledge of crop yield history is essential. As a result, this research suggests a strategy for forecasting agricultural prices. The process of examining data sets to gain insight about the facts they contain, sometimes through the use of sophisticated software and technology is known as data analytics. In the past, farmers' prior experience with a crop and land was taken into account when predicting crop prices. Farmers are compelled to grow an increasing number of crops, though, because of how quickly the environment is changing day by day. Due to the existing circumstances, many of them are unaware of both the advantages they receive by cultivating them and the potential losses they may sustain. To find patterns in the data and then process them, the proposed system employs prediction techniques based on machine learning namely XGBoost, Neural Nets, Clustering, Logistic Regression, and Decision Trees. This in turn will aid in determining the crop's target price. A farmer should constantly make the proper crop selection while taking environmental factors into account because crop selection is one of the most essential factors that significantly impacts the final yield. It can be challenging to decide which crop is best for a particular farm because there are so many different factors that might affect the production. Experts are regularly called upon to help farmers with crop recommendations or crop selection, however many farms cannot afford or have access to this alternative due to its time and cost requirements. The farming sector is confronted with substantial obstacles, including food supply, productivity, and sustainability due to the enormous population growth and sporadic weather variations. Despite their expertise in crop cultivation, farmers in rural areas lack access to a wide range of scientific and technological knowledge. Climate change and its consequences, such as extreme weather, are one of the major obstacles to a nation's ability to ensure its food security.

A subtype of artificial intelligence called machine learning enables computers to learn from data without the need to be explicitly processed. Machine learning is now more useful because to big data technology. A considerable amount of information that is collected quickly from multiple sources is simply referred to as "big data." Essentially, machine learning is about developing mathematical approach to better understand data. The three primary kinds of difficulties that machine learning is used to address are supervised problems, unsupervised challenges, and reinforcement concerns. The various categories are depicted in Fig.1. The empirical parameters for machine learning consist of a tuple comprising (x,y), in which denotes the so-called independent traits while y indicates the dependent variable or endpoint. Every time the factor y is known prior to the actual training, investigators experience a supervised learning problem. Regression and classification approaches are the two most common issues in supervised learning. Among the difficulties with regression is estimating land values. When some variable y was not known during training, unsupervised learning has been used. Many clustering problems have been resolved using their methods.

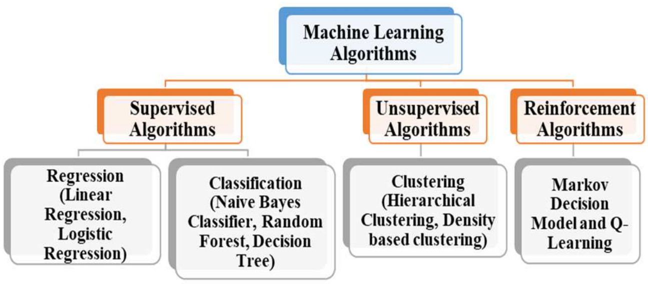

Fig. 1. Classification of Machine Learning Algorithms

Crop productivity projections typically involves the use of statistical models, which is laborious and time-consuming [21]. Big data's introduction in this decade has increased the adoption of more sophisticated analysis methods like machine learning. In accordance with the study challenge as well as strategies, the machine-learning model might be generally descriptive or prescriptive [22]. The three main types of difficulties that machine learning is employed to solve are supervised learning, unsupervised learning issues, and reinforcement difficulties [23-25]. Long-term changes in local, regional, or globe temperatures or weather patterns are referred to as climate disruption. Climate disruption makes it more difficult to reduce emissions of greenhouse gases and prevent global warming [26].

Machine learning with pragmatic agricultural modelling concepts were coupled to produce a machine learning paradigm for major agricultural productivity prediction [27]. Recurrent neural networks (RNN) as well as convolutional neural networks (CNN) are coupled in an unique multilevel deep learning method to incorporate spatial and temporal data for crop productivity prediction [28]. Paper [29] proposed a preliminary study to demonstrate the effect of merging crop simulation using machine learning on the improvement of maize output forecasts in the US region. Another investigation into the forecasting of potato tuber output were performed employing the four machine learning techniques [30]. It will also be intriguing to think about methods involving Big Graphs [31–33] and information obtained from smartphone sensors. The 1-2.5 degree Celsius temperature increase predicted for 2030 is anticipated to have significant consequences on crop yields [35] as it permits modifications to photosynthesis, accelerates plant respiration, and impacts pest population growth. "No hunger" is one of the objectives slated for accomplishment by 2030, along with "developing sustainable agriculture" [36]. Sustainable agriculture supports gender equality, poverty alleviation, small-scale farmer empowerment, and national economic prosperity. The current situation is worrying. Countries would impose continuous agricultural and food production techniques in order to guarantee sustained access to nutrient-dense food for everyone [37].

To achieve these objectives, agricultural observance must be timely and cost-effective. Agricultural production quantification is essential in this context for assessing and strengthening advanced cognitive capacities like coverage for agriculture, forecasting the financial markets, and addressing worries about food and nutrition security [38]. In view of the rapidly advancing technology, the current study aims to alter the procedure and boost crop yield [40] employing machine learning algorithms [39] and control mechanisms. In the earlier, agricultural yield estimates were made using machine learning approaches, particularly multivariate and linear regression[41]. The paper suggests improved machine learning techniques rely on specialised ensemble methods like gradient boosting, random forest, selection operator (LASSO) regression, and least absolute shrinkage. Their objective is to develop a web app that informs landowners and consumers about the quantity of crop production that will be produced based on the specific input, as well as the interrelations between the production and the various independent variables. A framework based on convolutional neural network-recurrent neural network is addressed [42]. Several models, such as deep fully connected neural networks, random forest (RF). The Forecasting was done for the US Corn Belt region for the years 2016, 2017, and 2018. Three types of variables, including soil, weather, maintenance, and the accuracy for corn and other crops, were used to determine the results 87.82% and 87.09%, respectively, were for soybean. To predict the yields, a classifier based mechanism random forest technique was used. According to this work, a web-based graphical interface was created to allow farmers to estimate crop yields before planting. The dataset includes information on Maharashtra's crop production, where the study was carried out. For yield forecasts, physical elements were applied as opposed to fuzzy models. Reports on the annual forecast evaluation. An experimental investigation for agrarian prediction was conducted in order to focus on making predictions the yield of “bajra,” or millet harvest, using appropriate computational models such as regression with time-series models [43]. Auto-regressive integrated moving average models (ARIMA) and an exogenous variable prediction model (ARIMAX) were employed. The ARIMAX model produced the best results. In comparison to the “bajra” regression time series approach. The use of ML to forecast crop yield was suggested. For agricultural yield, they employed layered regression, productivity, with soil nutrients as a secondary factor. The smallest errors for the LASSO, kernel ridge, and efficient neural network (ENeT) methods were 4%, 2%, and 1%, respectively. Natural selection-based genetic algorithms can provide efficient predictive model [44, 45].

A. *Machine Learning Techniques*

The Regression Model [46] portrays an association between independent and one or much more dependent variables. Using data and regression techniques to reduce loss or noise or error (RMSE or MSE), learning can be accomplished in a machine learning framework. Multiple independent variables were found to be more effective and trustworthy than one independent variable when the MLR analysis was applied in these applications [47]. Sets of adaptable splines with distinct slopes are created from the learning data. When used for classification and regression, the K-NN gives close neighbors greater weight when making predictions. Almost as outcome, they are much more indicative of the overall mean than their distant neighbours. The distance between two neighbors can be calculated using a number of distance formulas, including the Euclidean, Manhattan, and Minkowski ones. To categorise the data points, the binary classifier support vector machine generates a linear hyperplane that separates. Support vector regression, which can be employed to resolve regression problems, varies significantly from SVM. Decision Tree is a supervised machine learning method that can be used for regression as well as classification. The graph is composed of three nodes: a leaf node, a decision node with both inbound and outbound edges, as well as a root node with no incoming edges. A nonparametric study using such data, and produces binary trees using the continuous and discrete features [48]. Information in the CART gain ratio, the Gini Diversity Index (GDI), and also gain splits the attributes using. Random forest is an efficient way for yield estimation that has been used in agricultural research. It generates a wide range of regression trees for evaluating regression from a large amount of decision trees. The Random forest performs more accurately than any other decision tree, and the randomization allows the single decision tree to make up for the bias [49]. Unpruned decision trees are used in the ensemble model known as extremely randomised trees (extra tree), which is similar to random forest. The same amount of trees and variables are used to divide the nodes into groups as in RF. The most popular machine learning method for predicting agricultural yield is an ANN [50], that could also denote a complicated nonlinear relationship between inputs and outputs. An input layer, a hidden layer, and an output layer make up the total of three layers. An ANN's performance is influenced by many factors, including the quantity of nodes in the learning rate, training tolerance and the hidden layer. The learning statistic illustrates how significantly the rating will change over the span of many iterations in order to achieve the projected value within an acceptable range of the observed data.

B. *Deep Learning Techniques*

Deep learning is a subset of a wider range of machine learning techniques that integrate artificial neural networks and representation learning. A typical optimization model usually results in a localised rather than a global optimum state since the standard ANN is a local minima problem. Furthermore, a complexity problem brought on by overfitting occurs infrequently in generic machine learning models. The nonlinear operations in the activation layer could enhance CNN's capacity to perform nonlinear fitting. CNN in general updates weight using BP similarly to BPNN. Classification is successfully accomplished using an LSTM network. In addition to forecasting the time-series data and sequence data. It encompasses of input, output, and forgetting gates that used to regulate the preceding status filtering. This Structure seeks to obtain influential prior statuses rather than the most recent ones, to the present. The specific LSTM models' methodology and network structure can be used [51].

C. Miscellanous Algorithms

Other than core categories, numerous classification and regression methods, including as XGBoost, MARS, ridge regression (RIDGE), Gradient boosting TOMGRO and elastic net techniques, have been successfully applied in crop yield prediction. Crop yield prediction has also made use of hybrid techniques including CNN-RNN, CNN-LSTM, and MLR-ANN. A simple and precise estimating technique is needed by agricultural management to estimate rice yields throughout the arrangement process. Comparing the performance of ANN models with various formative parameters, and consider the compatibility of several direct relapse schemes with ANN models. In conclusion, Regression Neural Networks (GRNN) technique is used to predict the creation of agricultural produce. They considered GRNN to be a respectable tactic for predicting grain production in provincial districts. The GRNN model was identified as being appropriate for multi-target, multi-variate, and non-straight estimating. In contrast, planting, particularly the application of the ideal amount of treatment, has a substantial effect on crop production of certain other categories. The expert displays his evaluation of the crop forecast's updated k-Means grouping calculation. Their findings and evaluation showed a correlation between modified k-Means with k-Means bunching computation, and they discovered that the tweaked k-Means had produced the most high-caliber groups, the correct yield forecast, and the most extreme precision count. A technique for analysing time on parametric, layout information is to cut out key measurements and other informational characteristics. A method to predict future qualities based on recently viewed qualities is time arrangement anticipating. A novel idea for harvesting productivity under typical meteorological conditions was put forth, and a predicted model was devised by using time arrangement techniques to historical yield data. Recent years have seen a dominance of machine learning (ML) algorithms that have grown out of the data science paradigm. Forecasting time series in the financial and economic sectors has been done using it. Numerous empirical studies have demonstrated that when forecasting various financial assets, ML techniques outperform time series models. [52] provides a comparison of statistical models and machine learning strategies. These are all nonparametric, data-driven strategies that identify stochastic dependencies in the data. According to reports, ANNs perform better than traditional statistical techniques like linear regression and Box-Jenkins procedures. Technologies based on machine learning and deep learning have been shown to be effective solutions for time series prediction issues with strong seasonality. It has been demonstrated that neural networks are more accurate in forecasting agricultural prices than statistical methods. A well-written description of methods for forecasting time series using GRNN and utilising their inherent qualities to produce quick, extremely accurate forecasts may be found in [53]. Hog price predictions were made using support vector regression. Crop yield predictions using Random Forests regression can be found in some theoretical time series facts modelling developments.

III. Proposed Work

The proposed work includes machine learning algorithms namely decision tree and neuro-evolutionary algorithm. The overview illustration of the proposed framework is illustrated in Fig. 2.

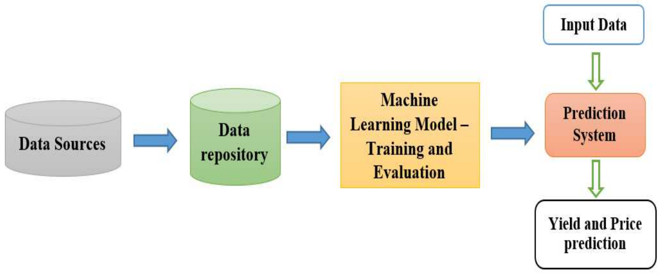

Fig. 2. Overview of Crop Yield Prediction System

A. Data Pre-processing

When using the ML algorithm, three datasets that were acquired as raw data must be processed. The data that are gathered are frequently missing in specific behaviours or trends, inconsistent, and incomplete. They are also probably full of mistakes. In order to use them for the model, they are pre-processed into a format after being gathered, using pandas in python to visualise and analyse massive amounts of data, removing any unnecessary columns. Mean values are used to fill in empty columns.

B. Decision Tree Technique

A technique which can be applied to resolve the classification problems as well as regression problem. This is basically a supervised category. The two main duties are price prediction and profit prediction. The factors taken into consideration for price forecast include Rainfall, Minimum Support Price (MSP), Maximum Trade, and Production. Crop price, harvest, cultivation expenses, and grain charges are all factors considered in profit projections. To forecast crop prices, we employ the Naive Bayes algorithm, a machine-learning classification technique. To forecast the crop's profit, we use K-Nearest Neighbor (KNN), a supervised learning based algorithm. The representative has the option of selecting the harvest and the prediction methods, as well as providing the necessary model parameters. The optimization techniques work behind the scenes to provide better prior knowledge of the outcome and produce the yield to the superiors, who would then converse it to the farmers. Like a outcome, the System offers Ranchers with a forecast, increasing their profitability and, as a result, the national economy.

C. Neuro Evolution Algorithm

The genetic algorithm, which aids in optimization, is one of the many algorithms that are supported by ANN. The evolution of natural organic systems was used to create the genetic algorithm known as the neuroevolution algorithm. The flow of the process is illustrated in Fig. 3. This algorithm's methods are more suited to handling a variety of prediction issues. To build ANN systems that can do the essential tasks, it is helpful to investigate knowledge that modifies the design process. The neuroevolution algorithm is used to do out unusual actions that result in inconsistent environments and enhance learning. The far more popular application for adaptable and nonlinear supervision of physical objects is reinforcement learning, and it can be incorporated.

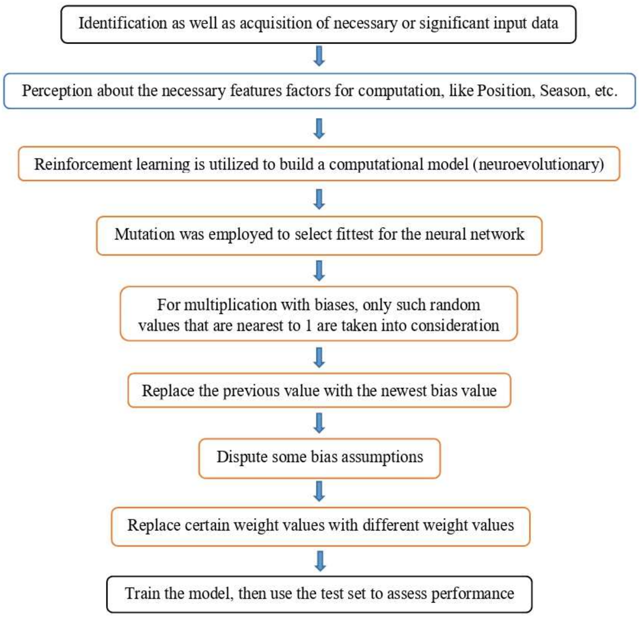

Fig. 3. Flow of Algorithm

Neuro evolutionary algorithms are generated by using mutation strategies that involve whether in adding or removing values from altered weights.

IV. Results and Discussions

The proposed framework is evaluated with a sample dataset from Kaggle. Entropy and Information gain are most important in finding the root node and to arrive at the decision. Neuro evolutionary algorithms are widely important in computing the price of the crop based on the market prices. The SSE is the most widely used splitting metric for regression.

$$SSE = \Sigma_{iss_{d_1}} (b_i - b_1)^2 + \Sigma_{iss_{d_2}} (b_i - b_1)^2 \quad (1)$$

SSE – Sum of Squared Error

s_{d1} and s_{d2} – groups of dataset

b_i and b_1 – average of the samples of the groups of the dataset

$$F_f = {}^{1}/_{|a+b+c-t|} \quad (2)$$

Based on the analysis, the better crop yield shall be arrived using random forest technique and also regression analysis is more important since regression analysis is a method of predictive technique that looks into the connection between a dependent (target) and an independence (s) factor (predictor). For instance, if the crop is a seasonal crop, its maturity will be classified as Early-mid maturing, Mid-maturing, or Default. The harvesting period may also be forecast, allowing farmers to obtain a higher yield and higher profit.

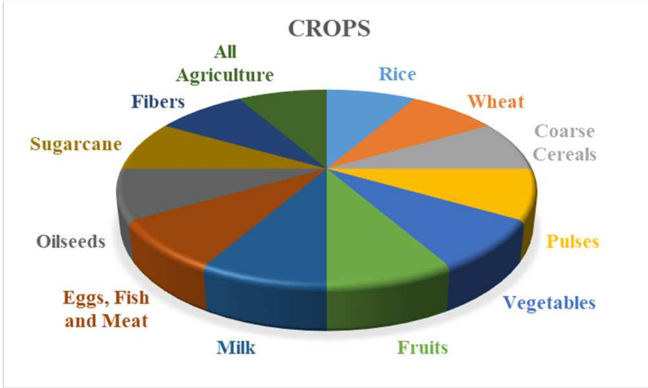

Fig. 4. Various Crop Yields

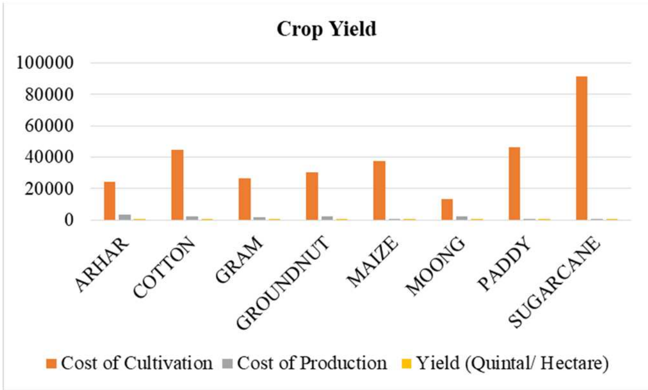

Fig. 5. Crop Yields

The sample datatset comprises of the information on the crops yield in a district from andhra pradesh, Fig.4 illustrates the vaious categories of crops and yield is depicted in Fig.5.

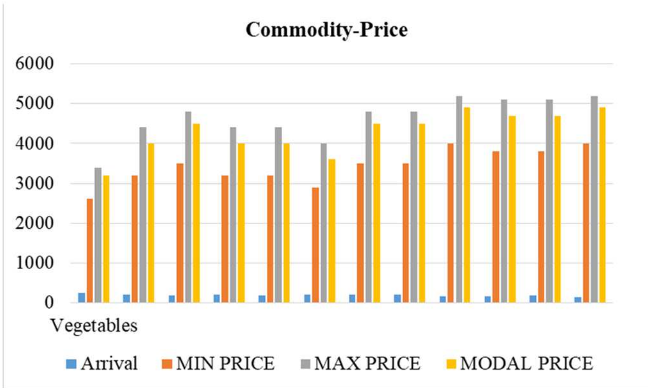

Fig. 6. Commodity Price Estimation

With the market demands, the prices are predicted based on the attributes feeded. This suggests a price that is effective for the crops namely tomato, brinjal, cabbage. The prices given in Fig.6 are estimated per quintal. Metrics for evaluation can be used to describe a model's performance. Assessment metrics are crucial because they can distinguish between both the outcomes of different learning models. Some performance measures used to determine the efficacy of the regression model include mean absolute percentage error (MAPE), mean absolute error (MAE), determination coefficient (R-squared), root mean square error (RMSE), and mean squared error (MSE). The MAE is defined as an arithmetic mean of both the absolute variability between the predicted inference and the actual observation in order to find the average significance of the inconsistencies with a given array of projections. The performance of the estimator is assessed using the MSE, which measures how closely the regressor line adheres to the dataset's points. RMSE measures the accuracy of the information. Focused on finding the best fit line and using it for estimating the standard deviation of the residuals or projected inaccuracy. For gauging the precision, use the determination coefficient, showing the accuracy of the regression model's fit, which demonstrates how. The created framework outperforms the existing one. The model's predictions differ from the actual findings. Machine learning-based classification methods for crop yield predictions are evaluated using precision, accuracy, recall, specificity, sensitivity along with F1 Score. Precision rate for the computed model is 86% and the recall rate is 87% along with an F1 score is 86.5%. The most popular and efficient statistic for classification issues is classification accuracy.

V. CONCLUSION

To nourish the growing global population, modern agriculture technology must be incorporated. Aside from that, agronomists require a proper framework in time which will enable people to foresee crop production so that they may effectively implement strategies to maximize crop yields. A machine learning based frameworks provide a clear description of the process by analyzing large amounts of data and interpret the findings. These technologies are used to construct models that describe the relationships between components and behavior. We propose a methodology for predicting corn crop production based on historical information. The approach utilizes a prediction model which is based on several parameters using machine learning techniques. The market rate of a crop reveals the current supply of housing for that harvest from the perspective of farming business. The agronomical timeline could be enhanced to generate revenue by tracking and forecasting market price. Crops are viewed as a sample, and the factors used to forecast crop costs encompass weather, current cost, position, demand metrics, and nutrient status. Empirical analysis revealed that the proposed framework was efficient and precise in making predictions farm prices. An automated and intelligent crop yield and price recommendation shall be developed with genetic algorithm in future.

REFERENCES

[1] Elavarasan, Dhivya, and PM Durairaj Vincent. "Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications." *IEEE access* 8 (2020): 86886-86901.

[2] Huang, Jianxi, et al. "Assimilation of remote sensing into crop growth models: Current status and perspectives." *Agricultural and Forest Meteorology* 276 (2019): 107609.

[3] Holzman, Mauro E., et al. "Early assessment of crop yield from remotely sensed water stress and solar radiation data." *ISPRS journal of photogrammetry and remote sensing* 145 (2018): 297-308.

[4] Singh, Arti, et al. "Machine learning for high-throughput stress phenotyping in plants." *Trends in plant science* 21.2 (2016): 110-124.

[5] Whetton, R., Zhao, Y., Shaddad, S., & Mouazen, A. M. (2017). Nonlinear parametric modelling to study how soil properties affect crop yields and NDVI. *Computers and electronics in agriculture*, *138*, 127-136.

[6] Dash, Y., Mishra, S. K., & Panigrahi, B. K. (2018). Rainfall prediction for the Kerala state of India using artificial intelligence approaches. *Computers & Electrical Engineering*, *70*, 66-73.

[7] Wieder, W., Shoop, S., Barna, L., Franz, T., & Finkenbiner, C. (2018). Comparison of soil strength measurements of agricultural soils in Nebraska. *Journal of Terramechanics*, *77*, 31-48.

[8] Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and electronics in agriculture*, *151*, 61-69.

[9] Basso, B., & Liu, L. (2019). Seasonal crop yield forecast: Methods, applications, and accuracies. *advances in agronomy*, *154*, 201-255.

[10] Shahhosseini, M., Martinez-Feria, R. A., Hu, G., & Archontoulis, S. V. (2019). Maize yield and nitrate loss prediction with machine learning algorithms. *Environmental Research Letters*, *14*(12), 124026.

[11] Shahhosseini, M., Hu, G., & Archontoulis, S. V. (2020). Forecasting corn yield with machine learning ensembles. *Frontiers in Plant Science*, *11*, 1120.

[12] Rehman, T. U., Mahmud, M. S., Chang, Y. K., Jin, J., & Shin, J. (2019). Current and future applications of statistical machine learning algorithms for agricultural machine vision systems. *Computers and electronics in agriculture*, *156*, 585-605.

[13] Elavarasan, D., Vincent, D. R., Sharma, V., Zomaya, A. Y., & Srinivasan, K. (2018). Forecasting yield by integrating agrarian factors and machine learning models: A survey. *Computers and Electronics in Agriculture*, *155*, 257-282.

[14] Johnson, M. D., Hsieh, W. W., Cannon, A. J., Davidson, A., & Bédard, F. (2016). Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. *Agricultural and forest meteorology*, *218*, 74-84.

[15] Chong, K. L., Kanniah, K. D., Pohl, C., & Tan, K. P. (2017). A review of remote sensing applications for oil palm studies. *Geo-spatial Information Science*, *20*(2), 184-200.

[16] Young, L. J. (2019). Agricultural crop forecasting for large geographical areas. *Annual review of statistics and its application*, *6*, 173-196.

[17] Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, *177*, 105709.

[18] Woittiez, L. S., Van Wijk, M. T., Slingerland, M., Van Noordwijk, M., & Giller, K. E. (2017). Yield gaps in oil palm: A quantitative review of contributing factors. *European Journal of Agronomy*, *83*, 57-77.

[19] Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, *18*(8), 2674.

[20] Li, B., Lecourt, J., & Bishop, G. (2018). Advances in non-destructive early assessment of fruit ripeness towards defining optimal time of harvest and yield prediction—A review. *Plants*, *7*(1), 3.

[21] Bali, N., & Singla, A. (2021). Deep learning based wheat crop yield prediction model in punjab region of north india. *Applied Artificial Intelligence*, *35*(15), 1304-1328.

[22] Romero, J. R., Roncallo, P. F., Akkiraju, P. C., Ponzoni, I., Echenique, V. C., & Carballido, J. A. (2013). Using classification algorithms for predicting durum wheat yield in the province of Buenos Aires. *Computers and electronics in agriculture*, *96*, 173-179.

[23] Abu Al-Haija, Q., Krichen, M., & Abu Elhaija, W. (2022). Machine-learning-based darknet traffic detection system for IoT applications. *Electronics*, *11*(4), 556.

[24] Mihoub, A., Snoun, H., Krichen, M., Salah, R. B. H., & Kahia, M. (2020, November). Predicting covid-19 spread level using socio-economic indicators and machine learning techniques. In *2020 first international conference of smart systems and emerging technologies (SMARTTECH)* (pp. 128-133). IEEE.

[25] Srinivasan, S., Ravi, V., Sowmya, V., Krichen, M., Noureddine, D. B., Anivilla, S., & Soman, K. P. (2020, March). Deep convolutional neural network based image spam classification. In *2020 6th conference on data science and machine learning applications (CDMA)* (pp. 112-117). IEEE.

[26] McEldowney, J. F. (2021). Climate change and the law. In *The Impacts of Climate Change* (pp. 503-519). Elsevier.

[27] Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C., & Athanasiadis, I. N. (2021). Machine learning for large-scale crop yield forecasting. *Agricultural Systems*, *187*, 103016.

[28] Sun, J., Lai, Z., Di, L., Sun, Z., Tao, J., & Shen, Y. (2020). Multilevel deep learning network for county-level corn yield estimation in the us corn belt. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 5048-5060.

[29] Shahhosseini, M., Hu, G., Huber, I., & Archontoulis, S. V. (2021). Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. *Scientific reports*, *11*(1), 1-15.

[30] Abbas, F., Afzaal, H., Farooque, A. A., & Tang, S. (2020). Crop yield prediction through proximal sensing and machine learning algorithms. *Agronomy*, *10*(7), 1046.

[31] Adoni, W. Y. H., Tarik, N., Krichen, M., & El Byed, A. (2021, April). Hgraph: Parallel and distributed tool for large-scale graph processing. In *2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)* (pp. 115-120). IEEE.

[32] Adoni, H. W. Y., Nahhal, T., Krichen, M., Aghezzaf, B., & Elbyed, A. (2020). A survey of current challenges in partitioning and processing of graph-structured data in parallel and distributed systems. *Distributed and Parallel Databases*, *38*(2), 495-530.

[33] Adoni, W. Y. H., Nahhal, T., Krichen, M., & Assayad, I. (2020). DHPV: a distributed algorithm for large-scale graph partitioning. *Journal of big Data*, *7*(1), 1-25.

[34] Krichen, M. (2021). Anomalies detection through smartphone sensors: a review. *IEEE Sensors Journal*, *21*(6), 7207-7217.

[35] Bhanumathi B, Vineeth M, Rohit N: Crop Yield Prediction and Efficient use of Fertilizers. IEEE International Conference on Communication and Signal Processing (ICCSP). 2019; pp. 769–773.

[36] Holzapfel, S., & Brüntrup, M. (2017). SDG 2 (Zero Hunger) in the context of the German Sustainable Development Strategy: are we leaving the starving behind? (No. 13/2017). Briefing Paper.

[37] Ramesh, D., & Vardhan, B. V. (2015). Analysis of crop yield prediction using data mining techniques. *International Journal of research in engineering and technology*, *4*(1), 47-473.

[38] Patil D, Shirdhonkar MS: Rice Crop Yield Prediction using Data Mining Techniques: An Overview. Int. J. Adv. Res. Comp. Sci. Softw. Eng. 2017; 7(5): 427–431.

[39] Rananavare, L. B., & Chitnis, S. (2022, July). Crop Yield Prediction Using Temporal Data. In *2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)* (pp. 1-5). IEEE.

[40] Zingade DS, Buchade O, Mehta N, et al.: Crop Prediction System using Machine Learning. Int. J. Adv. Engin. Res. Develop. Special Issue on Recent Trends in Data Eng. (IJAERD). 2017; 4(5): 01–06.

[41] Manjula, E., & Djodiltachoumy, S. (2017). A model for prediction of crop yield. *International Journal of Computational Intelligence and Informatics*, *6*(4), 298-305.

[42] Khaki, S., Wang, L., & Archontoulis, S. V. (2020). A cnn-rnn framework for crop yield prediction. *Frontiers in Plant Science*, *10*, 1750.

[43] Dharmaraja, S., Jain, V., Anjoy, P., & Chandra, H. (2020). Empirical analysis for crop yield forecasting in india. *Agricultural Research*, *9*(1), 132-138.

[44] Sathiyaraj, R., Bharathi, A., Khan, S., Kiren, T., Khan, I. U., & Fayaz, M. (2022). A Genetic Predictive Model Approach for Smart Traffic Prediction and Congestion Avoidance for Urban Transportation. *Wireless Communications and Mobile Computing*, *2022*.

[45] Sathiyaraj, R., Bharathi, A., & Balusamy, B. (2022). *Advanced intelligent predictive models for urban transportation*. Chapman and Hall/CRC.

[46] Abbas, F., Afzaal, H., Farooque, A. A., & Tang, S. (2020). Crop yield prediction through proximal sensing and machine learning algorithms. *Agronomy*, *10*(7), 1046.

[47] Guo, Y., Fu, Y., Hao, F., Zhang, X., Wu, W., Jin, X., ... & Senthilnath, J. (2021). Integrated phenology and climate in rice yields prediction using machine learning methods. *Ecological Indicators*, *120*, 106935.

[48] Brieman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and regression tree analysis.

[49] Cao, J., Zhang, Z., Tao, F., Zhang, L., Luo, Y., Han, J., & Li, Z. (2020). Identifying the contributions of multi-source data for winter wheat yield prediction in China. *Remote Sensing*, *12*(5), 750.

[50] Zaied, B. K., Rashid, M., Nasrullah, M., Bari, B. S., Zularisam, A. W., Singh, L., ... & Krishnan, S. (2020). Prediction and optimization of biogas production from POME co-digestion in solar bioreactor using artificial neural network coupled with particle swarm optimization (ANN-PSO). *Biomass Conversion and Biorefinery*, 1-16.

[51] Rashid, M., Islam, M., Sulaiman, N., Bari, B. S., Saha, R. K., & Hasan, M. J. (2020). Electrocorticography based motor imagery movements classification using long short-term memory (LSTM) based on deep learning approach. *SN Applied Sciences*, *2*(2), 1-7.

[52] Milunovich, G. (2020). Forecasting Australia's real house price index: A comparison of time series and machine learning methods. *Journal of Forecasting*, *39*(7), 1098-1118.

[53] Martínez, F., Charte, F., Frías, M. P., & Martínez-Rodríguez, A. M. (2022). Strategies for time series forecasting with generalized regression neural networks. *Neurocomputing*, *491*, 509-521.