

A Comparative study of Data Mining Techniques for Agriculture Crop Price Prediction

Jignesh Hirapara

Department of Computer Science Shree Manibhai Virani & Smt.
Navalben Virani Science College
jignesh.hirapara@atmiyauni.ac.in

Dr.Pratik Vanjara

Department of Computer Science Shree Manibhai Virani & Smt.
Navalben Virani Science College

Abstract- Agriculture crop prices forecasting is a very interesting and high challenging process as it is fully dependent on upcoming production in entire country. Recently most available application is designed for price analysis rather than price forecasting. In India agriculture production, when it is calculated per farmer, it is very instable is there compare to rest of the world, when compared to individual farmer in various countries with stable environment, and without providing sufficient MSP it will not benefit farmers and agriculture fraternity. If the farmers and agriculture fraternity get an access to appropriate crop prices, then poverty can be reduced in India. In advanced agriculture development, a large quantity of data is generated from the agriculture commodity market. Agriculture has a large amount of data, however regrettably, most of this data is not extracted to find out unseen information in data—crop price forecast is more beneficial to the farmers and agriculture fraternity to take proper and timely decisions. According to the output of process, Advanced data mining techniques play a pivotal role in analysis to discover a hidden pattern in data. Performance of data mining techniques is compared with past crop prices, weather, current market prices, stock availability and the upcoming production of the crop in recent years. The data mining that is a regression analysis, Tracking Patterns, Cluster Analysis, and visualization techniques are used to create an inventive representation to predict the agricultural crop prices.

Keywords: Data Mining, Crop Price, Agriculture Commodity, Data Mining techniques, visualization techniques, Tracking Patterns of Data mining, K-Means Clustering, Regression analysis techniques of Data mining, MSP, Minimum support price

I. INTRODUCTION

In farming occupation, farming fraternity apply impotent decisions of crop selling every time. In our country, farmers are not receiving the expected price from their farming crops. Agricultural field-crop price mainly depends on the market and the weather of the current season. Any farmer is enthused about knowing how a great deal of crop price he will foresee. Previously, the yield figure was performed by contemplating a farmer's understanding of the explicit field and collect. Agricultural field-crop price forecasting is a genuine problem that remains to be determined based on statistical data. Tracking patterns, visualization, and other data mining techniques is the best option for solving this problem using computer applications. This model implements a system to predict crop prices based on the last 10-year agricultural field-crops data and weather conditions including the rain data. This research paper focuses on future crop price prediction based on every dependent factor. If we have the last 10-year Agricultural

field-crop price data and weather data, these recorded price and weather information will help classify crop prices.

II. RELATED WORKS

Today there is no fully computerized system for price forecasting and crop planning for high return. Currently, the farmer or agriculture fraternity goes to the nearest market and sells the product. Now, there is no sure about future crop price. The computerized system is available for the farmers and farming fraternity to know the agricultural field-crop price at the upcoming time market price to sell their products to achieve high profits. This research aims to predict both the agricultural field-crop price and crop time frame of the crop before crop sowing. Using Tracking patterns, K-Means Clustering, visualization, and other data mining techniques. Farmers also decide the best time to sell their products in high price.

The ultimate aim of the Data Mining process is to derive the data from a data set and transform it into a specific format for agricultural field-crop prices at upcoming market prices [5]. Agricultural field-crop price analysis used various techniques of data mining such as K-Nearest Neighbor (KNN), regression analysis, Tracking Patterns, Tracking Patterns, and visualization [6].

In Nov- 2019, Guangyu Ding & Liangxi Qin used the techniques called “ Deep recurrent neural network and Associated neural network model ” paper titled “Study on the prediction of stock price based on the associated network model of LSTM ” Their main area focused on How computer based applications are used historical stock market price uses for predication[4]. In 2018, P.Akulwar.S.Pardeshi and Kamble used the techniques called “Association Rules; Classifications; Clusterings; Decision Tree; Neural Networks in their” paper titled “Survey on Different Data Mining Techniques for Prediction” Their main area focused and compare data mining techniques for more useful techniques used for accurate result for forecasting. Classification, Association Rules, Classification Clustering using lowest center values, Neural Networks, and Decision Tree data used for predicting [17].

In 2019, Swapnil Shrivastava, Supriya N. Pal, Ranjeet Walia used the techniques “Recurrent Neural Network and Auto-Regressive Integrated Moving Average” paper titled “Market Intelligence for Agricultural Commodities Using Forecasting and Deep Learning Techniques” Their main area focused on analyzed historical agriculture crop prices, Deep learning techniques are used for intelligent farming and how to overcome challenges for price prediction [2]. G. M. Nasira and

N. HemaGeetha say the Agriculture sector in this digital time is undergoing a spectacular change due to the power of the Digital world. Farming fraternities harvest crops and harvest a massive quantity of data [18]. Artificial Neural Network Data mining techniques are for analyzing price data.

In February 2019, YiranLiu, QinglingDuan, DongjieWang, ZhentaoZhang, Chunhong Liu used the techniques “support vector and regression data mining techniques” paper titled “Prediction for hog prices based on similar sub-series search and support vector regression” Their main area focused on Forecasting hog crop price and farmer need applies impotent decisions of crop selling timing [8]. In Feb. 2019, Aman Vohra, Nitin Pandey, S.K. Khatri used the techniques called “ARIMA model, neural networks, and PLS regression” paper titled “Decision Making Support System for Prediction of Prices in Agricultural Commodity” Their fundamental region centered Price estimating for the horticultural yield has turned into the need of great importance for ranchers and the cultivating crew [9]. In 2016, E.Manjula, S.Djodiltachoumy used the techniques called “Neural networks, biclustering, k-means, k-nearest neighbor Naive Bayes Classifier, and support vector machine are valuable techniques for predicting agricultural commodities” paper titled “Analysis of Data Mining Techniques for Agriculture Data. International Journal of Computer Science and Engineering Communications” Their main area focused Data mining and its techniques extract past agriculture crop price data and valuable information from large data sets [11].

III. DATA MINING AND ITS TECHNIQUES

Data mining is the way toward on the other hand seeing colossal information game plans with the objective that new information can be gotten from that which exists. All things considered, you make and see to expect. Data mining is defined as a process used to find usable information is characterized as a cycle used to observe usable data from a broad informational collection of any fundamental information. It suggests designs in enormous clusters of information utilizing at least one PC applications. Information mining has PC programming in numerous fields, similar to Agriculture, clinical, science, and examination. As computer software of data mining, farmer and agriculture fraternity can learn more about forecasting crop price.

The list of data mining techniques is Data cleaning and preparation, Regression, Classification, Clustering, Sequential patterns, Neural networks, Decision trees, Long-term memory processing, Machine learning, Visualization, artificial intelligence, Tracking patterns, Outlier detection, Association, Prediction, and Statistical techniques. The most suitable methods identified for agricultural field-crop price prediction as bellows.

A. regression analysis

This analysis technique is one of the most valuable techniques for forecasting modeling. Regression analysis is used to forecast a range of numeric or continuous values. This technique used for calculate the impourity calculate using

$(E) = 1 - \sum_{j=1}^c p2j$. Linear regression is used to guess an association between two variables. This technique uses the numerical blueprint of a straight line ($y=mx+b$). when dataset clone creates from internet they may not contain proper cluster format so this technique resolves this problem for crop price predication.

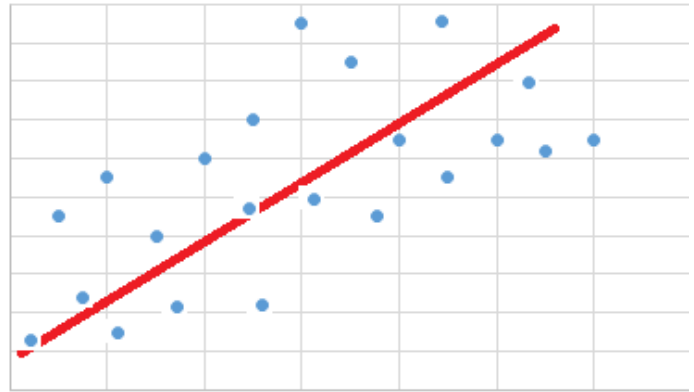


Fig. 1. Regression analysis

Regression analysis is also used for identify noisy data and Data Cleaning values from dataset. When we create dataset clone from internet at that time we got some noisy data it will change output accuracy.

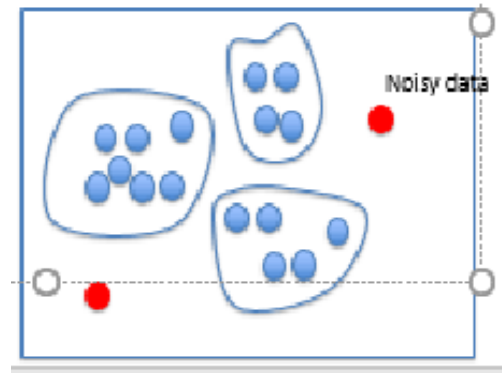


Fig. 2. Regression analysis for identify noisy data

Red indicates noisy data and they not connect with any cluster, if we consider as same value then it may be produce some garbage predication.

B. Tracking Patterns

Tracking Patterns is a primary and more helpful data mining technique for predict agricultural field-crop prices. Once farmer or agricultural fraternity identifies a trend in rural field crop price data, there is a ground for predicting crop price. Tracking Patterns determine that a particular weather condition or crop available stock is more affected selling price. Agriculture fraternity and farmers can use this knowledge to manage crop production or storage time frame for a better price.

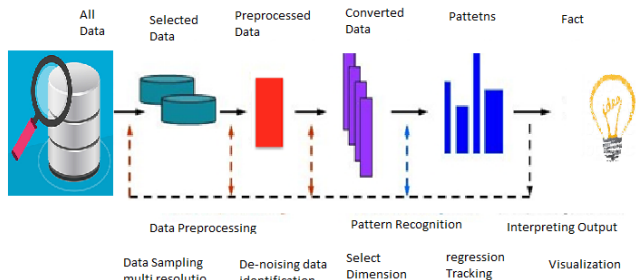


Fig. 3. Tracking Patterns

If we find Tracking Patterns from data source than first, we need to normalization data source. normalization has two part first; Min-Max Normalization calculate using $\frac{Value-min}{max-min} = ?$, second; Z-score Normalization calculate using $\frac{Value-(mean\ value\ of\ the\ eatures)}{standard\ deviation\ of\ the\ features} = ?$. After normalization of dataset we need to train dataset and find pattern and predicate crop prices.

C. visualization techniques

Visualization techniques play an essential role in predicting concept searching inferences by analyzing the past data through Tracking Patterns or regression analysis techniques. This technique is a new hypothesis about the train dataset. Without this technique output produce in any range. Output range requirement: $0.95 < Lie\ Factor < 1.05$ and calculate

$$Lie\ Factor = \frac{size\ of\ effect\ shown\ in\ graphic}{size\ of\ effect\ in\ data} =$$

This technique is the process of production output in a way that the user can fast and easily digest. This technique gives the outcome of your result in a different format like Box plots, graphs, Histograms, digital images, and Charts. If we need to produce any output, then this technique is more beneficial. This technique provides the facility to make the final result to the end-user.

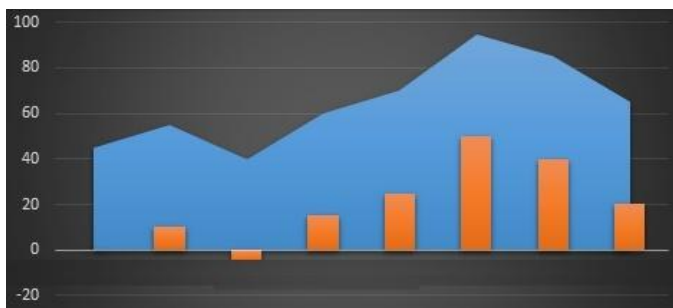


Fig. 4. Visualization techniques

D. K-means Clustering

K-Means Clustering mainly user for find number of cluster and create cluster group in given dataset. This technique points as initial centroids and find distance from one value to other value of same cluster as well other cluster point. When we get dataset clone from internet they contain row data, if we use this dataset then first create data cluster. Without creating cluster,

we not find duplication, repetition and noisy data from dataset. This technique uses for create proper dataset before trained and without train dataset is not user for any forecasting.

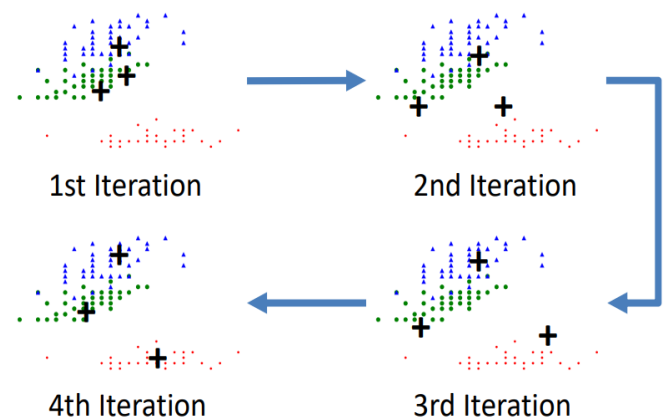


Fig. 5. K-means Clustering

IV. MODEL DESIGN

This design proposed to predict agricultural field-crop prices. In this method, initially finds the Agriculture fraternity requirements using text input or graphics input and identity data sources and formats. Initially data found as unstructured data later on our algorithm works on it and make data format in organized way with structured data. So it will be easy to find the data for such data for any application and such API creation. This structured data will become as an outcome. We found the historical data since many years and that will be structured in Year wise, Month wise even we can set this data from different regional set so it become easy to find specific region data based on given date range. The data has different primitive data types as format and it has stored in collection which contains document.

State	object	Rainfall	float64
District	object	Crop	object
Year	int64	Price	float64
Month	int64	dtype:	object

Fig. 6. Data source format for predication

Here we filtered the data and make clustering based on year, rain flow and district as region. This data is now become information from raw data. We can use specific equation based on need and requirement. This equation will help to train the system we can set such constrains as well for supervised and semi-supervised learning methods. System will collect actual data for further knowledge and it will make perfect prediction system. Clustering has many equations we followed but one of them I had shown below.

$$J = \sum_{i=1}^K \sum_{x \in C_i} dist(c_i, x)^2$$

Fig. 7. Equation of cluster variation

The data contains duplication, unwanted data, remaining fields data, unformatted data so all these things need to filter and rectify using the mathematical formulas and make classification to do cluster.

$$c = \left(\frac{1}{N_i} \right) \sum_{j=1}^{N_i} d_j$$

Fig. 8. Equation of cluster classification

TABLE I. SAMPLE TRAIN DATA SET

	Price	Change	Rainfall	Result
count	1628.000000	1628.000000	1628.000000	1628.000000
mean	205.000000	6.800000	86.570238	0.480000
Std	2.05129	3.322373	144.454880	0.481495
min	203.000000	1.000000	0.000000	0.000000
25%	203.000000	3.430000	0.242000	0.000000
50%	206.000000	6.300000	12.100000	0.500000
75%	207.000000	9.120000	107.900000	1.000000
max	209.000000	12.000000	918.400000	1.000000

Here we have sample trained dataset which will help us to predict the crop prices. Based on classification and we set the frequencies to map actual vs fitted values which is actual desired outcome. Here we can see the graph of proposed system as an outcome crop prices comparison which actual meaningful and helpful to all the different public domain application and stockholders who are belongs to agriculture field and production seller.

When the data-set is ready after removing all errors, then we take some inputs from the users as well as weather and production news sentiments from the web sources. for prediction we use user knowledge based on their requirement and we mining the weather information from current and future weather environmental changes data. Even we can find data from news portal using data mining of specified categories related news and match the sentiments accordingly. That data mining technique we build based on price prediction not for other knowledge so we had query for such specific crop production information as well its demands in market.

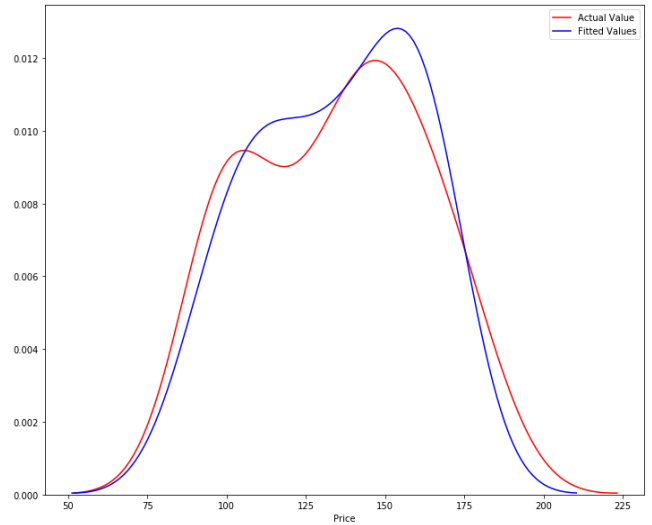


Fig. 9. Actual vs Fitted Values of Crop Price

Now, the whole proposed algorithm become as knowledge outcome will really help a lot to predict crop price for current market sells and future market price for sells to gain more profit.

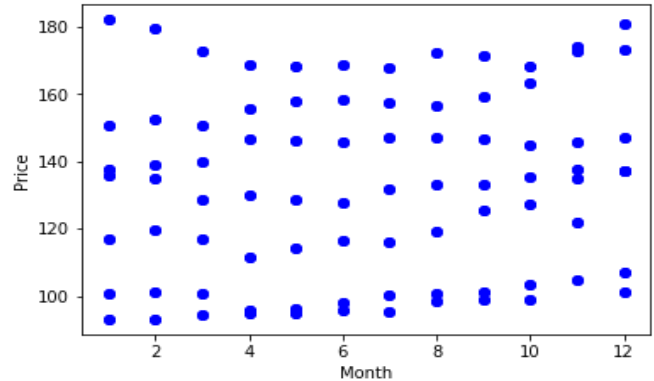


Fig. 10. Price possibility after data mining techniques apply

when result is not acceptable by user then again train dataset and get input from user.

To add further to this, according to the proposed model, there may be a point where things may not get as per expectations. On the other hand, there are chances that the results module may get rejected by one's purview. In such a case if the same model is repeated, there are more than expected chances that the desired results may bump in. All in all, if the results are not derived as per the model proposed here, one should start again, as there are huge chances to the output as per need and expectations.

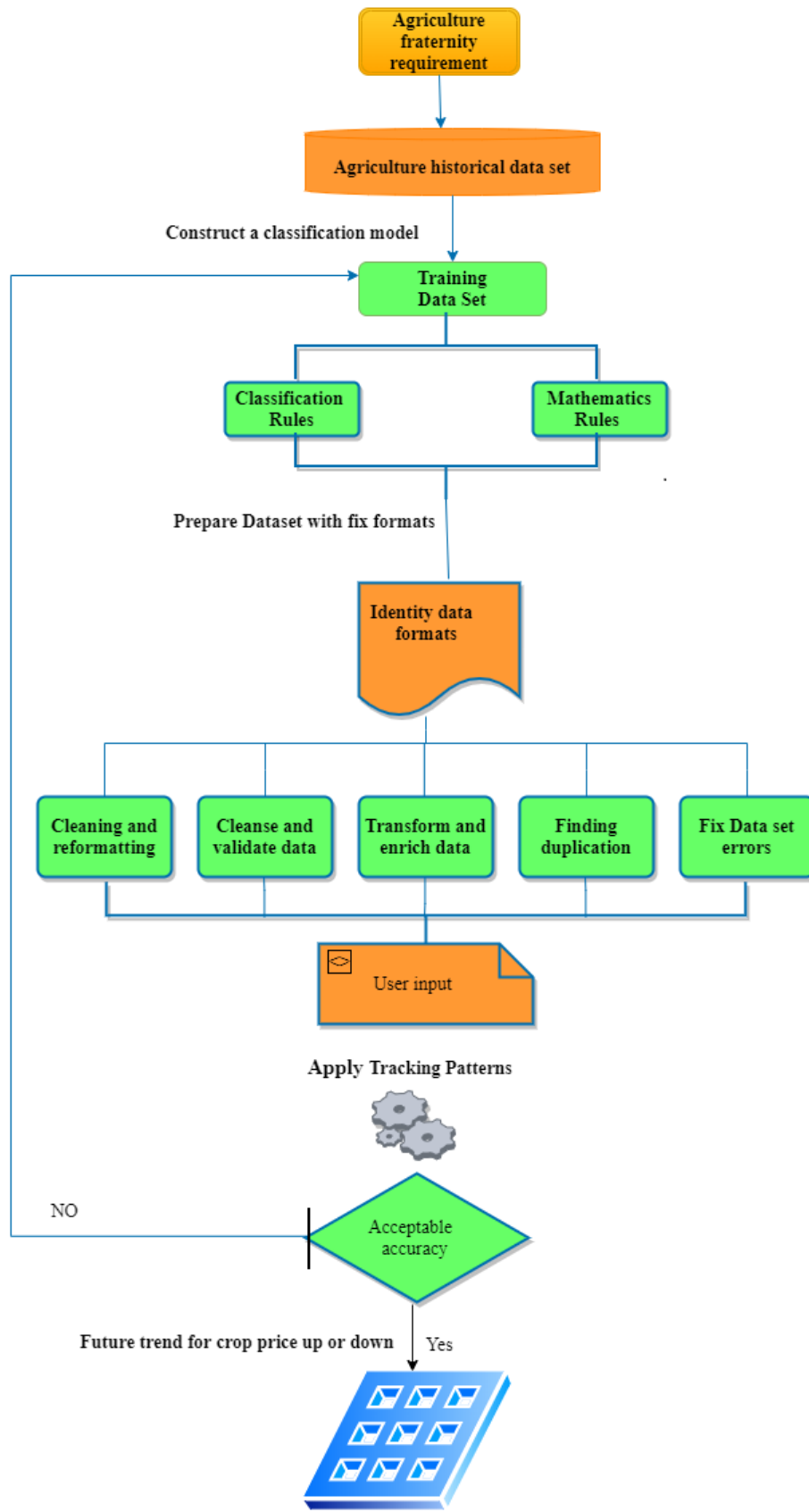


Fig. 11. Proposed crop price prediction using data mining techniques.

V. CONCLUSION

This paper tries to finalize which data mining techniques are used to predict agricultural field-crop prices and reviews data mining techniques involved in price prediction and the importance resolve issue of the agriculture field. This system's major aim is to the farmers and farming fraternity getting a high amount of their crop and storage time for batter price of the products. Forecasting base on computer application is a more growing concepts in the last few years due to the increasing demand and supply of agricultural field-crops.

To have a sneak peak in the proposed model, before concluding, there are a few clarifications which are essential. So far all the models presented by the researchers had to do with "only analysis" of the price. Whereas this model has a lot more accuracy. This model presented, according to the research work will get the price of the crop yield. Again, all the research has some limitations so has this. The research has a limitation to present great amount of accuracy only in the certainty of the environmental condition. Like for instance in case of conditions like draught or excessive rains, this may not provide the data as per expectation. Rest all shall be great to compute in and for. In the agriculture economy weather disasters can determine significant changes in crop prices which make the forecasting with "low accuracy".

REFERENCES

- [1] Guangyu Ding & Liangxi Qin, "Study on the prediction of stock price based on the associated network model of LSTM ", *International Journal of Machine Learning and Cybernetics* volume 11, pages1307–1317(2020)
- [2] Shrivastava S., Pal S.N., Walia R. (2019) Market Intelligence for Agricultural Commodities Using Forecasting and Deep Learning Techniques. *Big Data Analytics. BDA 2019. Lecture Notes in Computer Science*, vol 11932. Springer, Cham. https://doi.org/10.1007/978-3-030-37188-3_12
- [3] J. C. Price, "Estimating leaf area index from satellite data," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 31, no. 3, pp. 727-734, May 1993, doi: 10.1109/36.225538.
- [4] DING, G., QIN, L. STUDY ON THE PREDICTION OF STOCK PRICE BASED ON THE ASSOCIATED NETWORK MODEL OF LSTM. *INT. J. MACH. LEARN. & CYBER.* **11**, 1307–1317 (2020). [HTTPS://DOI.ORG/10.1007/S13042-019-01041-1](https://doi.org/10.1007/s13042-019-01041-1)
- [5] RADADIYA, B., & SHUKLA, P. (2019). "STRENGTH, WEAKNESS, OPPORTUNITY, AND CHALLENGES IN AGRICULTURE MARKETS". *ASIAN JOURNAL OF AGRICULTURAL EXTENSION, ECONOMICS & SOCIOLOGY*, *33*(2), 1-7.
- [6] YiranLiu, QinglingDuan , DongjieWang, ZhentaoZhang and ChunhongLiu,"Prediction for hog prices based on similar sub-series search and support vector regression",*Computers and Electronics in Agriculture Volume 157, February 2019, Pages 581-588*
- [7] A. VOHRA, N. PANDEY, AND S. K. KHATRI, "DECISION MAKING SUPPORT SYSTEM FOR PREDICTION OF PRICES IN AGRICULTURAL COMMODITY," 2019 AMITY INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE (AICAI), DUBAI, UNITED ARAB EMIRATES, 2019, PP. 345-348, DOI: 10.1109/AICAI.2019.8701273.
- [8] LakkanaRuekkasaem and MontaleeSasananan, "FORECASTING AGRICULTURAL PRODUCTS PRICES USING TIME SERIES METHODS FOR CROP PLANNING", *International Journal of Mechanical Engineering and Technology (IJMET) Volume 9, Issue 7, July 2018, pp. 957–971*
- [9] E.MANJULA, S.DJODILTACHOUY. "ANALYSIS OF DATA MINING TECHNIQUES FOR AGRICULTURE DATA." *INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND ENGINEERING COMMUNICATIONS* **4.2** (2016): 1311-1313.
- [10] R, MANJUNATH. (2018). CROP MANDI-DEMAND AND PRICE FORECASTING OF AGRICULTURAL CROPS THROUGH MOBILE APPLICATION. *INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY.* **6.** 4512-4520. 10.22214/IJRASET.2018.4741.
- [11] Manjunath, C. , Devaiah, B. and Govind Yadav. Crop mandi-demand and price forecasting of agricultural crops through a mobile application, *International Journal for Research in Applied Science and Engineering Technology 2018 Vol.6 No.4 pp.4512-4520 ref.18*
- [12] Ponce-Guevara, K. L., Palacios-Echeverria, J. A., Maya-Olalla, E., DominguezLimaico, H. M., Suarez-Zambrano, L. E., Rosero-Montalvo, P.D., Alvarado-Perez, J. C. (2017). GreenFarm-DM: A tool for analyzing vegetable crops data from a greenhouse using data mining techniques (First trial). *2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM)*.
- [13] Suraparaju, Veenadhari. (2016). *Data mining Techniques for Predicting Crop Productivity – A review article.* 10.13140/RG.2.1.1755.0324.
- [14] Majumdar, J., Naraseyappa, S. & Ankalaki, S. Analysis of agriculture data using data mining techniques: application of big data. *J Big Data* **4**, 20 (2017). <https://doi.org/10.1186/s40537-017-0077-4>
- [15] P. Akulwar, S. Pardeshi, and A. Kamble, "Survey on Different Data Mining Techniques for Prediction," *2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud) (I-SMAC)/I-SMAC (IoT in Social, Mobile, Analytics, and Cloud) (I-SMAC), 2018 2nd International Conference on*, Palladam, India, 2018, pp. 513-519, DOI: 10.1109/I-SMAC.2018.8653734.
- [16] G. M. Nasira and N. Hemageetha, "Vegetable price prediction using data mining classification technique," *International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012)*, Salem, Tamilnadu, 2012, pp. 99-102, DOI: 10.1109/ICPRIME.2012.6208294.
- [17] Ayman E. Khedr, S.E.Salama, Nagwa Yaseen,"Predicting Stock Market Behavior using Data Mining Technique and News Sentiment Analysis", *International Journal of Intelligent Systems and Applications(IJISA)*, Vol.9, No.7, pp.22-30, 2017. DOI: 10.5815/ijisa.2017.07.03
- [18] Fangwen Zhai, Qinghua Wen, Zehong Yang and Yixu Song, "Hybrid forecasting model research on stock data mining," 4th International Conference on New Trends in Information Science and Service Science, 2010, pp. 630-633.
- [19] M. Huang, T. Zhang, J. Wang and L. Zhu, "A new air quality forecasting model using data mining and artificial neural network," *2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, 2015, pp. 259-262, doi: 10.1109/ICSESS.2015.7339050
- [20] İşlek and Ş. G. Ögüdücü, "A retail demand forecasting model based on data mining techniques," *2015 IEEE 24th International Symposium on Industrial Electronics (ISIE)*, 2015, pp. 55-60, doi: 10.1109/ISIE.2015.7281443
- [21] Yau-Ren Shiau, Ching-Hsing Tsai, Yung-Hsiang Hung, Yu-Ting Kuo, "The Application of Data Mining Technology to Build a Forecasting Model for Classification of Road Traffic Accidents", *Mathematical Problems in Engineering*, vol. 2015, Article ID 170635, 8 pages, 2015. <https://doi.org/10.1155/2015/170635>
- [22] H. Mori and A. Takahashi, "A data mining method for selecting input variables for forecasting model of global solar radiation," *PES T&D 2012*, 2012, pp. 1-6, doi: 10.1109/TDC.2012.6281569.
- [23] M. Huang, T. Zhang, J. Wang and L. Zhu, "A new air quality forecasting model using data mining and artificial neural network," *2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, 2015, pp. 259-262, doi: 10.1109/ICSESS.2015.7339050.
- [24] Y. Wang and Y. Wang, "Using social media mining technology to assist in price prediction of stock market," *2016 IEEE International Conference on Big Data Analysis (ICBDA)*, 2016, pp. 1-4, doi: 10.1109/ICBDA.2016.7509794.
- [25] J. H. Zhao, Z. Y. Dong, X. Li and K. P. Wong, "A Framework for Electricity Price Spike Analysis With Advanced Data Mining Methods," in *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp. 376-385, Feb. 2007, doi: 10.1109/TPWRS.2006.889139.