**RESEARCH MATRIX :2321-7073**
*Peer Reviewed & Refereed | International Multidisciplinary journal of applied research*

# CROP PRICE DATA INTERPRETATION:A COMPARISON OF MACHINE LEARNING

**JIGNESH HIRAPARA**
**DEPARTMENT OF COMPUTER SCIENCESHREEMANIBHAIVIRANI&SMT.**
**NAVALBENVIRANI SCIENCE COLLEGE**
**DR.PRATIK VANJARA**
**DEPARTMENT OF COMPUTER SCIENCESHREEMANIBHAIVIRANI&SMT.**
**NAVALBENVIRANI SCIENCE COLLEGE**

**ABSTRACT**
Machine learning and its methodologies are used in agribusiness domains to predict edit costs based on stock availability and generation. On a daily basis, a massive amount of data is generated through the display of farming products. Horticulture has a large amount of data, but unfortunately, much of it isn't able to find out inconspicuous details in information. Edit cost estimates are more beneficial to agriculturists and the agriculture society since they demand proper timing. Information mining procedures that have progressed play a critical role in the discovery of hidden design in data. Following Designs, Cluster Analysis, and visualization methodologies are used to provide a unique representation to predict the horticultural edit cost. Past trim cost, climate, current advertise cost, stock accessibility, and up and coming trim generation in current year or season are all used to compare information mining procedure execution.Recently, the most often used programmer has been designed for cost inquiry rather than cost determination. When compared to individual agriculturists in various countries with stable environments, India's agribusiness generation is exceptionally instable, and without appropriate MSP, it will not benefit agriculturists and farming crew. If ranchers and agribusiness personnel are given the opportunity to appropriate alter costs, destitution in India can be reduced.

***KEYWORDS: DATA MINING, CROP PRICE, MACHINE LEARNING, AGRICULTURAL, AGRIBUSINESS, FARMING FRATERNITY, AGRIBUSINESS FRATERNITY, MSP, MINIMUM SUPPORT PRICE***

## 1 INTRODUCTION

In farming field farmers and farming fraternity apply impotent decisions of crop selling every time. In our country, farmers are not getting the expected price from their farming crops. Agriculturalfield-cropprice mostly depends on the market and the weather of 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. Agriculturalfield-crop price forecasting is a genuine problem that remains to be determinedbased on statistical data. Tracking patterns, visualization and other techniques of data mining is the best option for solving this problem using computer application. This proposal implements a system to predict crop prices based on the last 10-year agriculturalfield-cropdata and weather condition with rain data. This is done through the use of various data mining techniques. This research paper focuses on future crop price

prediction base on every dependent factor. If we have the last 10-yearAgriculturalfield-cropprice data andweather data are available and this recoded price and weather information will help to classify crop price.

Dimensionality diminishment is wide spread preprocessing in high dimensional information examination, visualization and modeling. One of the only ways to diminish dimensionality is by Include Selection; one chooses as it were those input measurements that contain the significant data for tackling the particular problem. Feature Extraction may be a more common strategy in which one tries to create a change of the input space onto the lowdimensional subspace that jam most of the significant information.

## 2. RELATED WORK

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 sell the product. Currently, there is no sure about future crop price and not computerized system is available for the farmers andfarming fraternity to know the agricultural field-crop price at upcoming time market price where they can sell their products for achieving high profits.

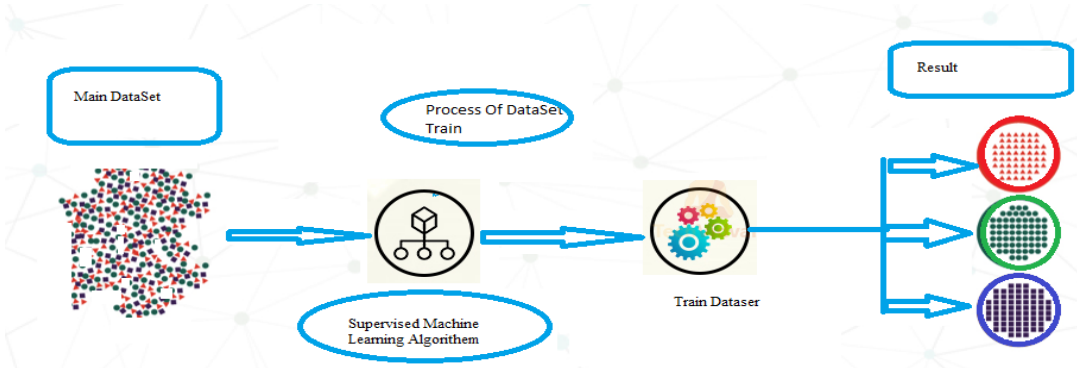
The research aims to predict both the agricultural field-crop price and agricultural field-crop time frame of the crop before crop sowing via Tracking patterns, visualization and other techniques of data mining. Farmers also decide the best time to sell their products. current prediction was performed by thinking about the farmer's and farming fraternitycomprehension of a farming field. Be that as it may, as the conditions change step by step quickly, ranchers are compelled to develop an ever-increasing number of yields. Being this as the present circumstance, a significant number of them need more information about the new harvests and are not totally mindful of the advantages they get while cultivating them. Additionally, the ranch profitability can be expanded by comprehension and estimating crop execution in an assortment of natural conditions.

## 3. MACHINE LEARNING

Machine learning may well be a creating development which engages computers to think about in like manner from once data. Machine education livelihoods distinctive computations for raising numerical models and making vaticinators working out observational data or information. Right presently, it's being utilized for diverse errands comparable as picture affirmation, banter affirmation, correspondence filtering, Facebook transport- trailing, recommender outline, and multitudinous advance. This machine learning instructive exercise gives you a gift to machine learning at the side the wide run of machine education techniques comparable as Coordinated, Unsupervised, and Bastion proficiency. You'll learn generally backslide and bracket models, clustering methodologies, secured up Markov models, and diverse successive models.

### 3.1Supervised Machine Learning

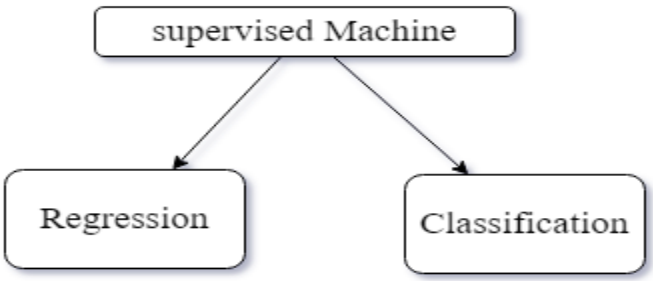
In coordinated education, the planning data given to the machines work as the ace that teaches the machines to prevision the abdicate straightforwardly. It applies the same conception as a understudy learns inside the supervision of the teacher. Directed education may be a plan of giving input data as well as correct abdicate data to the machine proficiency illustrate. The point of a managed proficiency computation is to find a mapping work to diagram the input variable (x) with the abdicate variable (y).




3.1 Supervised Machine Learning Work model.

### 3.2 Types of Supervised Machine Learning

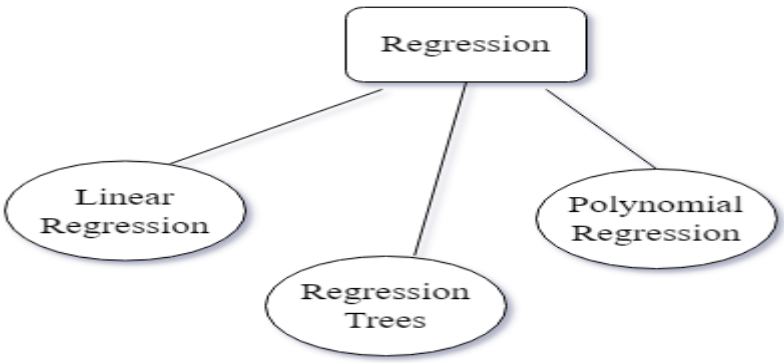
This diversified in two divided into two types of problems




3.2 Supervised Machine Learning Types.

#### 3.2.1 Regression

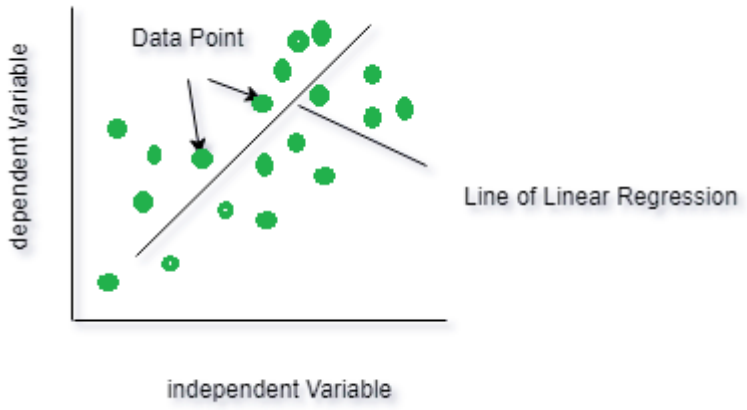
Bigger portion of the machine learning computations drop underneath the managed proficiency arrange. It's the framework where a computation is utilized to expect a result grounded on the once entered values and the comes almost delivered from them. Accept we've an input variable‘x’and a surrender variable‘y’where y may be a work of x (y = f x). Managed education scrutinizes the respect of entered variable‘x’and the coming approximately variable‘y’so that it can utilize those comes approximately to hence expect a significantly exact surrender data of‘y’from the entered respect of‘x’. A backslide issue is when the coming around variable contains an honest to goodness or an understanding respect. It tries to draw the line of in vogue fit from the data collected from a numberof centers.




3.2.1 Supervised Machine Learning Regression Algorithms.

**3.2.1.1 LinearRegression**

Straight relapse calculation appears a straight relationship between a subordinate (y) and one or more free (y) factors, consequently called as direct relapse. Since straight relapse appears the direct relationship, which suggests it finds how the esteem of the subordinate variable is changing agreeing to the esteem of the autonomous variable.




3.2.1.1 Supervised Machine Learning Linear Regression Algorithms.

$y= a_0+a_1x+ \varepsilon$ Linear Regression equation.

Where,

a0= Regression line

a1= slope of the regression line

ε = The error term.

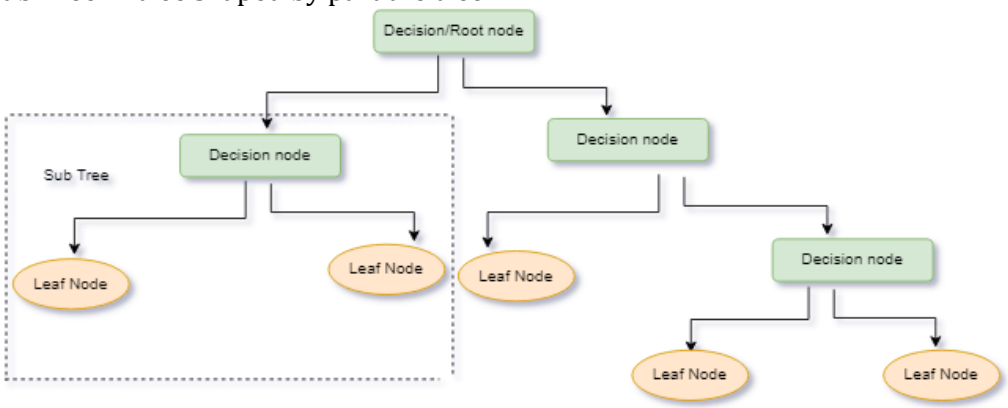
**3.2.1.2**Regression Tree

In a Decision tree, there are two hubs, which are the Choice Hub and Leaf Hub. Choice hubs are utilized to form any choice and have numerous branches, while Leaf hubs are the output of those choices and don't contain any assist branches.

Root Node: Root hub is from where the choice tree begins. It speaks to the complete dataset, which assist gets separated into two or more homogeneous sets.

Leaf Node: Leaf hubs are the ultimate yield hub, and the tree cannot be isolated advance after getting a leaf hub.

Branch/Sub Tree: A tree shaped by part the tree.

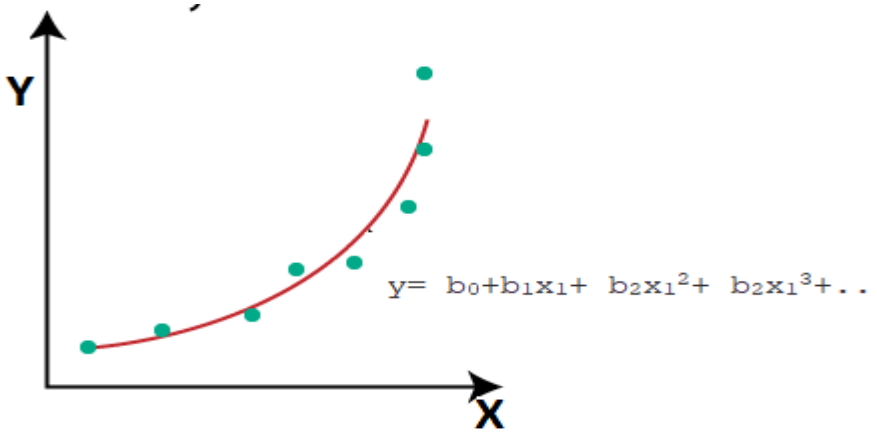



3.2.1.2 Supervised Machine Learning Regression Tree Algorithms.

3.2.1.3 Polynomial Regression

Polynomial Relapse may be a relapse approach that employments an nth degree polynomial to speak to the association between a dependent(y) and free variable(x). The condition for polynomial relapse is as takes after:

$y= b_0+b_1x_1+ b_2x_1^2+ b_2x_1^3+..... b_nx_1^n$

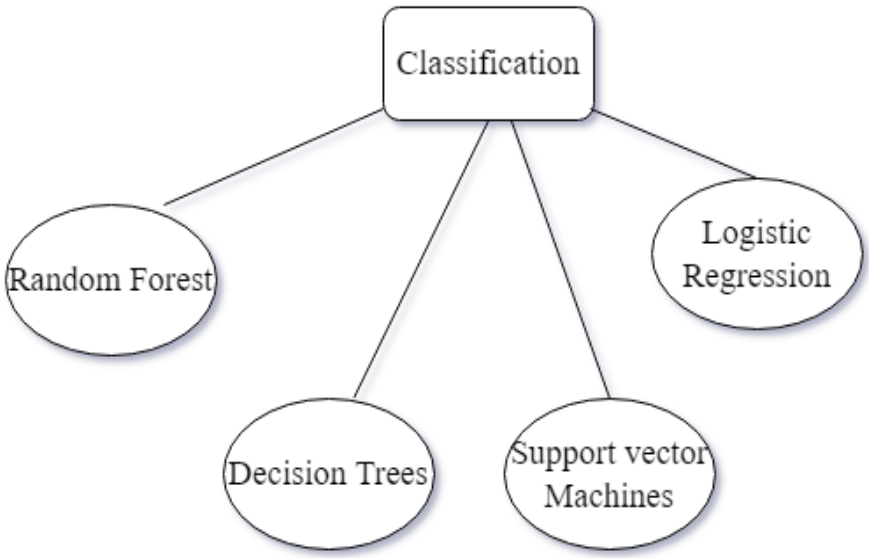



3.2.1.2 Supervised Machine Polynomial Regression Algorithms.

We utilized a dataset that was sorted out non-linearly within the picture over. So, on the off chance that we utilize a direct model to cover it, we will see that it scarcely covers any information focuses. On the other hand, a bend is reasonable to cover most of the information focuses, which is of the Polynomial demonstrate.

**3.2.2** Classification

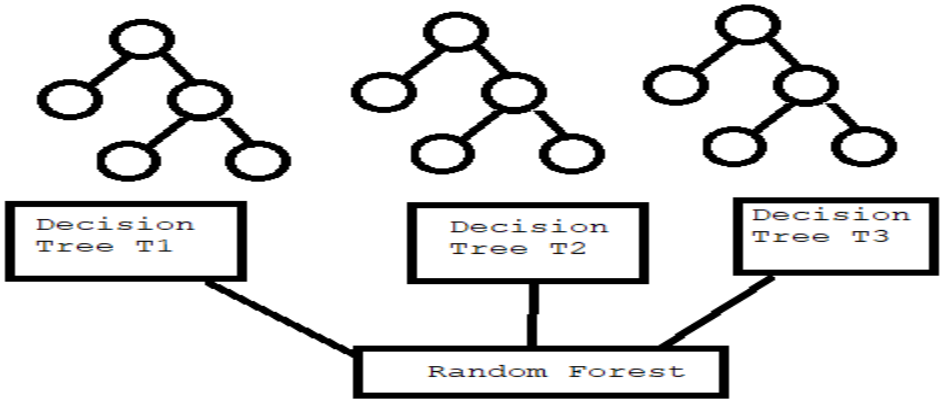
The Classification calculation may be a Directed Learning procedure that's utilized to recognize the category of unused perceptions on the premise of preparing information. In Classification, a program learns from the given dataset or perceptions and after that classifies unused perception into a number of classes or bunches.




3.2.2 Supervised Machine Learning Classification Algorithms.

**3.2.2.1** Random Forest

Random Forest could be a well-known machine learning calculation that has a place to the administered learning strategy. It can be utilized for both Classification and Relapse issues in ML. It is based on the concept of outfit learning, which could be a handle of combining numerous classifiers to unravel a complex issue and to make strides the execution of the show. Since the irregular woodland combines numerous trees to anticipate the lesson of the dataset, it is conceivable that a few choice trees may foresee the right yield, whereas others may not. But together, all the trees foresee the right yield.

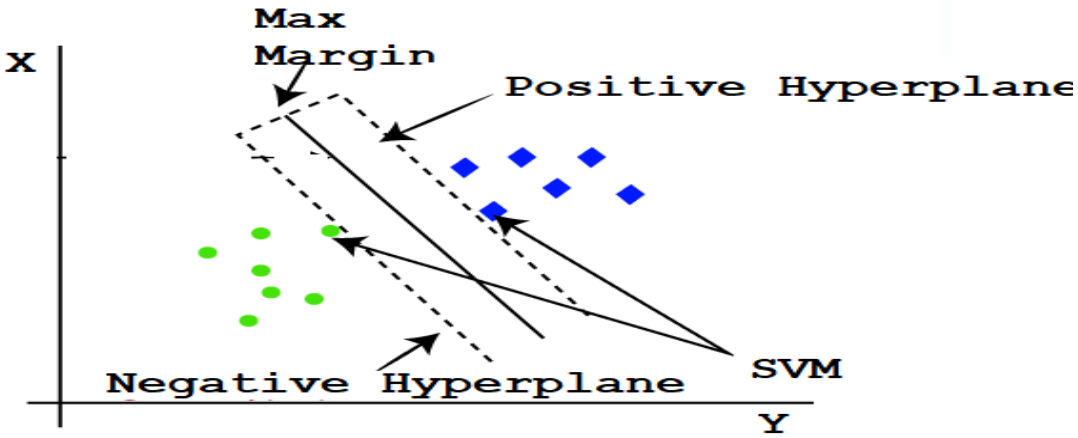



3.2.2.1 Supervised Machine Learning Classification Random Forest Algorithms.

One of the foremost vital highlights of the Irregular Timberland Calculation is that it can handle the information set containing continuous factors as within the case of relapse and categorical factors as within the case of classification. It performs way better comes about for classification issues.

**3.2.2.2**Support Vector Machine(SVM)

The objective of the SVM calculation is to form the finest line or decision boundary that can isolate n-dimensional space into classes so that we are able effectively put the modern information point within the adjust category within the future. This best choice boundary is called a hyperplane.




3.2.2.2 Supervised Machine Learning Classification Support Vector Machine(SVM)Algorithms.

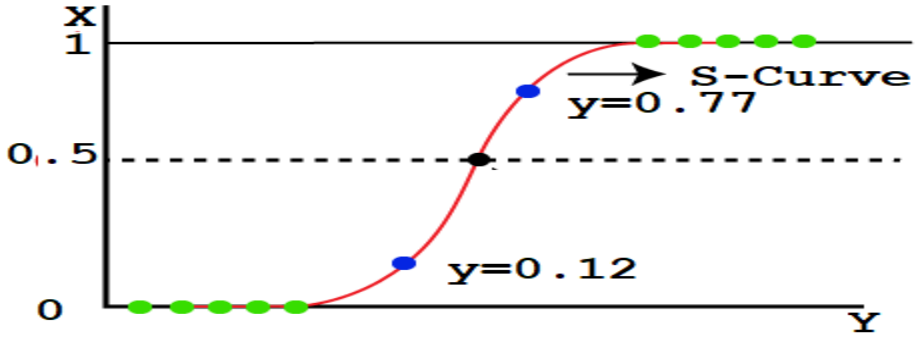
Linear SVM is utilized for straightly distinguishable information, which suggests in the event that a dataset can be classified into two classes by employing a single straight line, at that point such information is named as straightly distinct information, and classifier is utilized called as Direct SVM classifier.

Non-Linear SVM is utilized for non-linearly isolated information, which implies in the event that a dataset cannot be classified by employing a straight line, at that point such information is named as non-linear information and classifier utilized is called as Non-linear SVM classifier.

**3.2.2.3**Logistic Regression

Logistic regression predicts the yield of a categorical subordinate variable. Hence the result must be a categorical or discrete esteem. It can be either Yes or No, or 1, genuine or Untrue, etc. but rather than giving the precise esteem as and 1, it gives the probabilistic values which lie between and 1.

Logistic Regression could be a noteworthy machine learning algorithm since it has the capacity to supply probabilities and classify unused information utilizing nonstop and discrete datasets.




3.2.2.3 Supervised Machine Learning Classification Logistic Regression Algorithms.

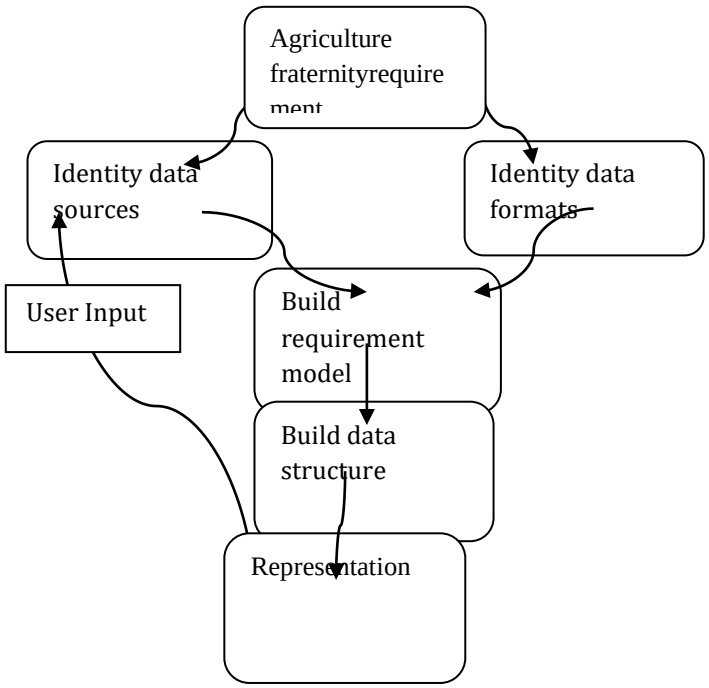
The esteem of the calculated relapse must be between and 1, which cannot go past this restrain, so it shapes a bend just like the "S" shape. The S-form bend is called the Sigmoid work or the calculated work.

The Calculated relapse condition can be gotten from the Straight Relapse condition. The numerical steps to induce Calculated Relapse conditions are given underneath:

$$\log\left[\frac{y}{1-y}\right] = b0 + b1x1 + b2x2 + \cdots . bnxn$$

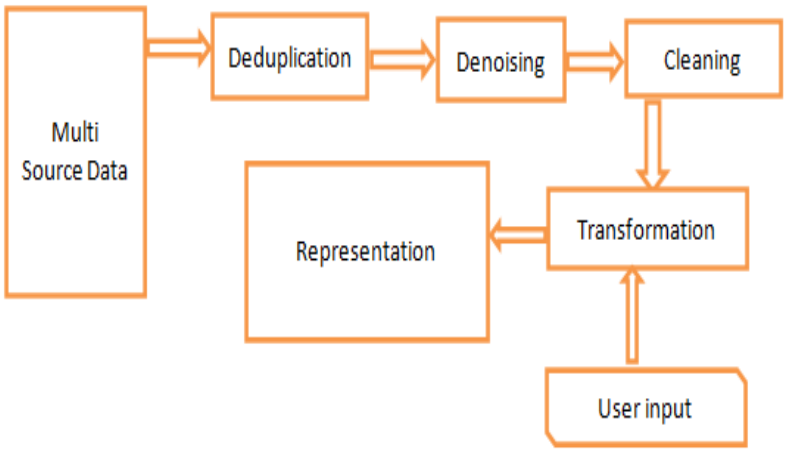
**4. THE DESIGN-METHODS AND PROCEDURES**

The investigates proposed anticipate agricultural field-crop prices and agricultural field-crop time outline utilize of information mining strategies. It's handled to require information from the different information sources and input from clients. Make graphical and printed yield for ranchers and cultivating crew. This design-method is to begin with find Agriculture fraternity requirement utilizing content input or illustrations input and character information sources and designs. Discover duplication of information comes from different information source and removes unnecessary information from information set utilizing denoising strategy of information mining. Construct necessity demonstrate and information structurer and once more take input from agriculturists and cultivating crew for comprehension of a cultivating field and create last result.




**4.1 proposed crop price prediction using Machine Learning.**

This design-method is to begin with discover Agribusiness society necessity utilizing content input or design input and personality information sources and designs. Discover duplication of information comes from different information source and expels superfluous information from information set utilizing denoising strategy of information mining. Construct necessity show and information structurer and once more take input from agriculturists and cultivating society for comprehension of a cultivating field and deliver last result. It is handled to get information from the different information sources as well as input from clients and make graphical and literary yield for the client. The proposed calculation plan based on information mining methods.




4.2 Dataset Formatting using Machine Learning.

## 5. Conclusion

This paper tries to finalized which data mining techniques is use for predict agriculturalfield-crop price and review of data mining techniques involved in price predication as well as importance resolve issue of agriculture field.This research clearly defines which data mining techniques more useful for predication base on existing data of agriculture filed crop price and weather condition of current year or session. This system main aim is to the farmers and farming fraternity getting high amount of their crop and storage time for batter price of crop.Forecasting base on computer application is more growing concept in last few years due to increasing demand and supply of agriculturalfield-crop. This framework primary point is to the agriculturists and cultivating society getting tall sum of their trim and capacity time for hitter cost of trim. Estimating base on computer application is more developing concept in final few a long time due to expanding request and supply.

## 6. References

[1]Sen, J., Datta Chaudhuri, T.: An alternative framework for time series decomposition and forecasting and its relevance for portfolio choice - a comparative study of the Indian consumer durable and small cap sector. J. Econ. Libr. **3**(2), 303–326 (2016)

[2] *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.*

[3] Ayman E. Khedr, S.E.Salama, NagwaYaseen,"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

[4] FangwenZhai, 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.

[5] 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

[6] İşlek and Ş. G. Öğüdücü, "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

[7] 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

[8] 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.

[9] 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.

*[10] 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.*

*[11] 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.*

*[12] S. Lei, "A Feature Selection Method Based on Information Gain and Genetic Algorithm", In: International Conference on Computer Sciences and Electronics Engineering, pp. 355-358, 2012.*

*[13] I. Guyon et al, "Gene Selection for Cancer Classification using Support Vector Machines", Machine Learning, Vol. 46(1-3), pp. 389-422, 2002.*

[14] R.Tibshirani, T.Hastie, B.Narasimhan and G.Chu, "Diagnosis of Multiple Cancer Types by Shrunken Centroids of Gene Expression", In; Proceedings of the National Academy of Sciences of the United States of America, Vol. 99(10), pp. 6567-6572, 2002.

[15] A. Hyyarinen, "Survey on Independent Component Analysis", Neural Computing Surveys, vol. 2, pp.94–128, 1999.

[16] D. Langlois et al, "An Introduction to Independent Component Analysis: InfoMax and FastICA Algorithms", Tutorials in Quantitative Methods for Psychology, vol. 6(1), pp. 31-38, 2010.

[17] Linting et al, "Nonlinear Principle Components Analysis: Introduction and Application", Psychological Methods, 2007.

[18] S. Zhou, "Probabilistic analysis of Kernel Principle Components: Mixture Modeling, and Classification", Draft, pp. 1-26, 2003.

[19] M. Pechenizkiy et al, "Class Noise and Supervised Learning Learning in Medical Domain: The Effect of Feature Extraction" In: Proceeding of the 19th IEEE Symposium on Computer-Based Medical System (CBMS'06), 2006.

[20] C. Ding and H. Ping, "Minimum Redundancy Feature Selection from Microarray Gene Expression Data", Journal of Bioinformatics and Computational Biology, vol. 3(9), pp. 185-205, 2005.

[21] F. Fleuret, "Fast Binary Selection with Conditional Mutual Information", Journal of Machine Learning, pp. 1531-1555, 2004.

[22] I. Guyon, "An Introduction to Variable and Feature Selection", Journal of Machine Learning Research 3, pp. 1157-1182, 2003.

[23] M. Klassen and N. Kim, "Nearest shrunken centroid as feature selection of microarray data," In: Proceedings of Computers and Their Applications, pp. 227-232, 2009.

[24] Ayman E. Khedr, S.E.Salama, NagwaYaseen,"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

[25] FangwenZhai, 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.

[26] 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

[27] İşlek and Ş. G. Öğüdücü, "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

[28] 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

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] "Market Intelligence for Agricultural Commodities Using Forecasting and Deep Learning Techniques" Swapnil Shrivastav, Supriya N. Pal, Ranjeet Walia 12 December 2019, link.springer.com