Different Types of Data Mining Techniques Used in Agriculture - A Survey

R. S. Kodeeshwari, K. Tamil Ilakkiya

Department of CSE, Coimbatore Institute of Engineering and Technology, Coimbatore, India

Abstract—The most important domain is Agriculture in broadly cultivating countries like India. The situation of decision making can be amended by using the current technologies, the. So that the farmer's can yield in an improved way. The major role in decision making to agricultural domains is Data mining. In this paper acquaints in connection with some of the most important data mining techniques used in agriculture. Mining in agriculture is a innovative groundwork domain. The problems in the agricultural field can be efficiently solved by using data mining techniques since it anticipate before in hand with the help of raw data's. Previously mentioned, the paper discuses about various data mining techniques such as classification, clustering, association rule and regression.

Keywords— Agriculture, data mining, artificial neural network, k-means, decision tree, classification, clustering, association rule, regression, descriptive, predictive.

I.   INTRODUCTION

Agriculture is the authority of India. Only one-third of cropped part is only inundated in India, in spite of large areas. Since the agriculture data occurred everyday the capacity of data has been enlarged rapidly mostly on last five years. Farmers, researchers, government and agricultural scientists are still searching and extracting for fresh techniques for farming to increase the better production. At present new methods are present in agriculture are used by a very few farmers. For predicting future trends of agriculture processes "data mining" can be used. The process of examining data by summarizing in different perspective and converting it into an beneficial information in large datasets is called Data mining. Data mining has no restriction for analyzing the type of data.

II.   DATA MINING IN AGRICULTURE

In large data sets, data mining is the computational process for discovering new patterns. Data mining provides major advantage in agriculture for disease detection, problem prediction and for optimizing the pesticides. In recent technologies agriculture related activities provide lot of information. Hence this data mining techniques in agriculture are used for pattern reorganization and disease detection. Data's of agriculture in data mining can be presented in form of data marts. Crop production for reliable and timely requirement for various decisions for marketing, pricing, storage distribution and import-export. The yield of agriculture primarily depends on diseases, pests, climatic conditions, planning of different crops for the harvest productivity are the results. So by these predictions are very useful for agriculture domains. Data mining techniques are used for pre-harvest forecasting. For example by applying data mining technique government can fully benefit data about farmers buying patterns and also to gain a superior understanding of their land to protect them in order to gain more profit on farmer's part. Data mining is also called as knowledge discovery database (KDD).

Data mining tasks can be classified into two categories:
  - ➢ Descriptive data mining.
  - ➢ Predictive data mining.

 Descriptive data mining tasks characterize the general properties of the data in the database while predictive data mining is used to predict the direct values based on patterns determined from known results. Prediction involves using some variables or fields in the database to predict unknown or future values of other variables of interest. As far as data mining technique is concern, in the most of cases predictive data mining approach is used. Predictive data mining technique is used to predict future crop, weather forecasting, pesticides and fertilizers to be used, revenue to be generated and so on.[12]

IMPORTANCE OF DATA MINING:

Data mining is the major technique for collection of data's in various forms among the data collected in the process of data mining includes research data, survey data, organization data, competitive data and social media such as whatsapp, Facebook.

Several steps are involved in analyzes on selected set of data where the process involves of filtering, transformation, testing, modelling, visualization and documentation is prepared and the result is outputted (or) the data is stored accordingly in data warehouse or

databases. To propose a smart agriculture we must predict the yield of crop based on the water, texture of soil and climate. It is essential for our country to build a large production of organic crops. So by applying data mining techniques for agriculture we can reduce the cost of food production and improves productivity which encounters in greater decision making process in business world. i.e. agriculture.

FIVE MAJOR ELEMENTS IN DATA MINING:

- ✓ Fetch the data and load the data to transform onto the warehouse system.
- ✓ Store and use the data in the database system.
- ✓ Make available to access data for researchers, IT professionals and for various organizational analytics.
- ✓ Examine the required data using suitable software's.
- ✓ Formulate the data's inform of table or graph to represent data in an useful format.

CHALLENGING PROBLEMS IN DATA MINING:

- ✓ Progress a consolidated theory of data mining.
- ✓ Maximize for large structural data and high speed data streams.
- ✓ Mining time series data and ordered data.
- ✓ Data Mining has composite knowledge from complicated data.
- ✓ Data mining in a certain network environment.
- ✓ Mining multi-agent data and scattered data mining to improve reliability and performance.
- ✓ Data mining for inorganic and atmospheric problems.
- ✓ The processes based on data mining related problems.
- ✓ Privacy, protection and data purity.
- ✓ Dealing with unbalanced, high cost and varied types of data.

DATA MINING TECHNIQUES:
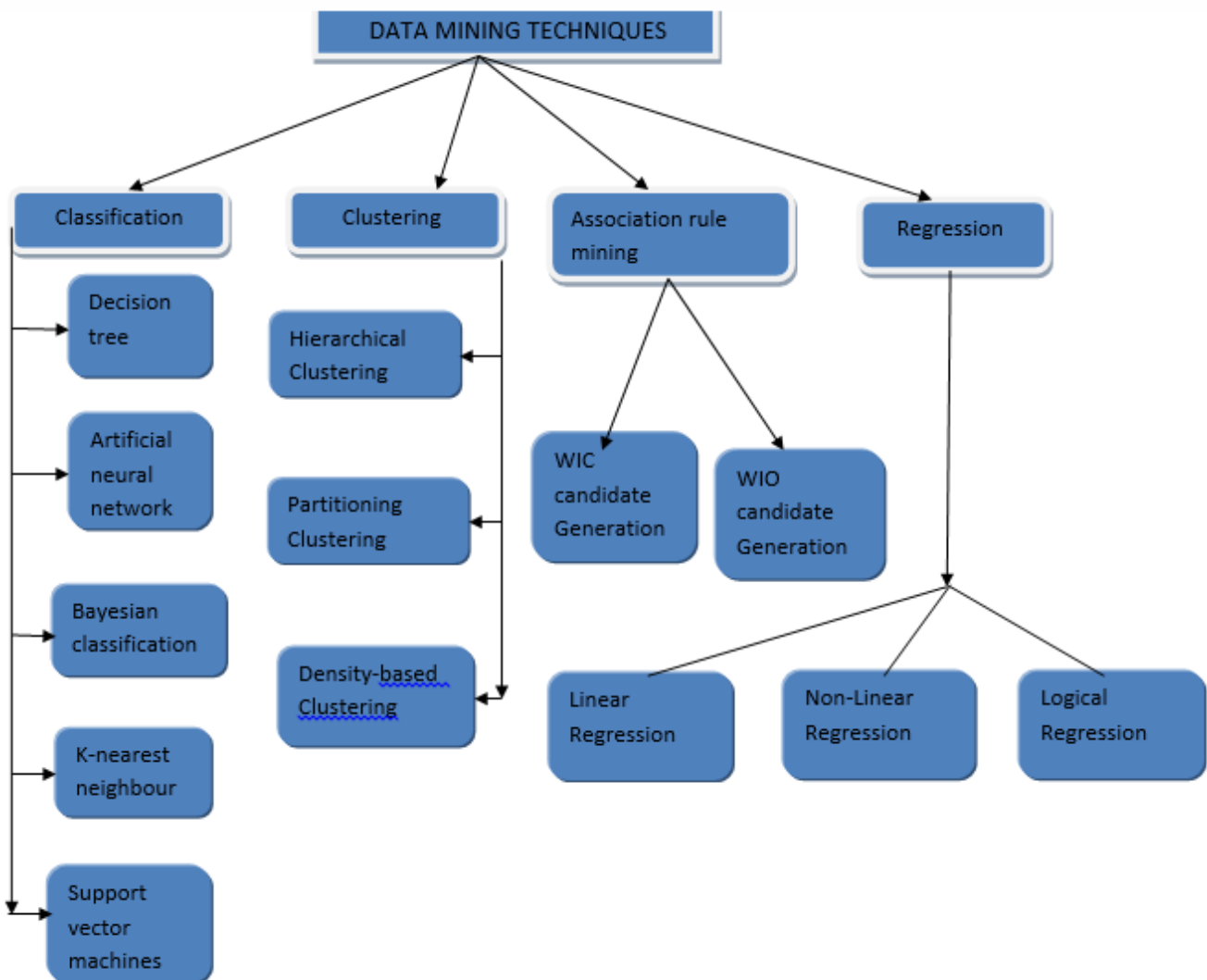
CLASSIFICATION:

Based on machine learning, data mining is a classic technique. one of a predefined set of groups is classified into each time in a set of data. A software is developed in classification that can acquire information about how are data items classified into group a simple example is we can apply classification in the application in that "given all records of stock in a departmental store what all products are sold extensively and products should be paired as combo offer for increased profit in a future period".
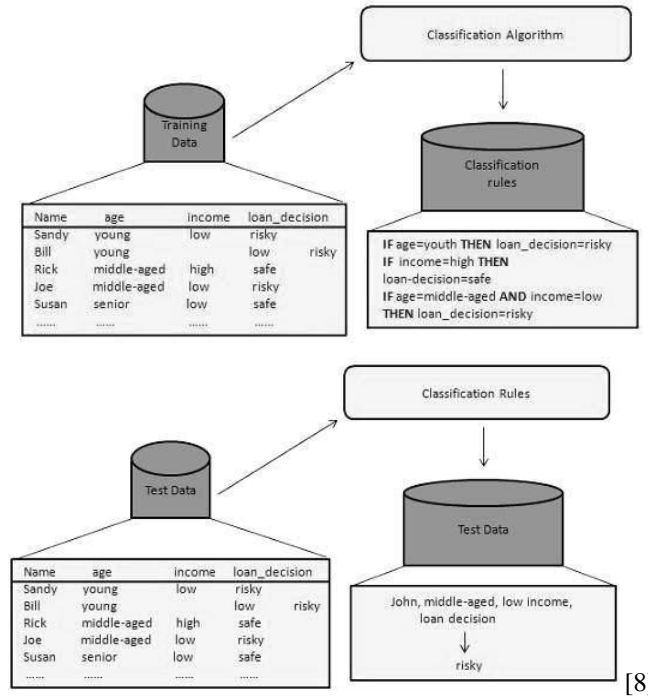
[8]

In this condition, the records of stock products are spitted into number of individuals collectively that names "extensive sale" and "lacking products" and we can classify the stock maintenance into separate groups into data mining software.

In other words certain input is given it predicts the outcome. To predict this outcome, a training set is processed by the predefined algorithm containing the group of attributes and required outcome. Which is called as "prediction attributes". The algorithm in classification helps to analyze the relationship among the attributes and For example we have a medical database so that this database must have the recorded significant patients information earlier for acquiring base knowledge about the patient whether the patient is affected with heart problem previously or not.

Training dataset:
Prediction dataset:

AGE	BP	HEART RATE	PROBLEM IN HEART
20	78	157/70	?
40	98	184/70	?
65	86	167/70	?

makes to predict the possible solutions. A good algorithm can be defined when the prediction is accurate.

The major advantage of classification technique is to give the overall view about the type of customer, object (or) an item to identify a particular class by describing multiple attributes. For example by identifying different attributes (car colour, car shape) we can classify cars into different types. Agriculture uses data mining techniques for knowledge discovering based on the datasets save in past and present yields.

AGE	BLOOD PRESSURE	HEART RATE	PROBLEM IN HEART
35	77	105/70	No
25	78	112/70	No
42	38	160/70	Yes

IF (Blood pressure>140/70) or (heart rate>70) the problem in heart=yes
Else
Problem in heart=no.
A prediction is said to be good when the prediction hit percentage adverse the total count of predictions.

Good prediction= prediction hit percentage/ Total count of prediction

CLUSTERING:

Clustering is a data mining technique which is used to group the set of data objects into multiple clusters[meaning sub-classes] . A sub-set of object which are similar is called a cluster. High similarity occurs when the objects are in same clusters and the objects are dissimilar in other clusters. Similarities and dissimilarities are evaluated by describing the objects based on attribute value. Algorithm of clustering are used in following steps such as for identifying the data, analyze the data, data refinement, model construction, detection of out structure The notation of the cluster is expressed in more number of applications. To understand better about what cluster consists an example is shown below:

a)   **Original points**

b)   **Two clusters**

c)   **Four clusters**

d)   **Six clusters**
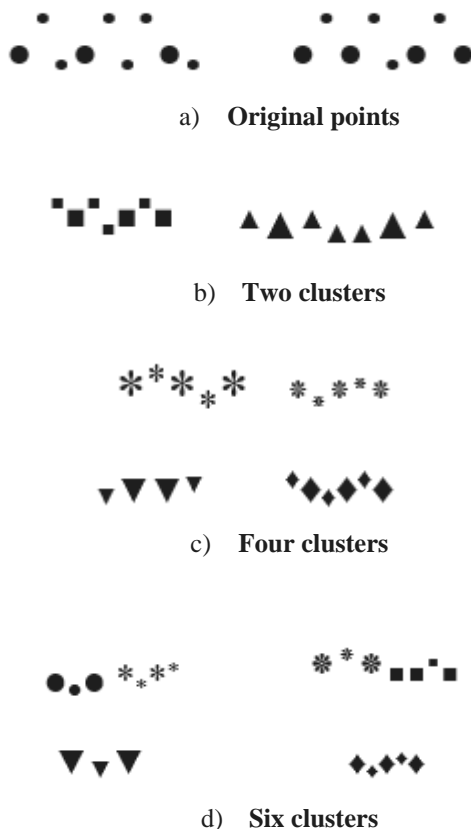
Clustering is a data mining technique which maps the similar instance together, and dissimilar instance together, and dissimilar instance belong to diverse group based on data instance. The data instance are divided into subsets. To identify different information clustering technique is used because it correlates with examples where similarities and ranges agree. In this technique there is no need of prior knowledge about data.

Clustering technique comes under unsupervised learning that takes unlabeled data records and differentiate them

and for processing data. Clustering consists a cluster centre that contains all the clusters. A well defined clustering method will generate a high quality clusters.

- Inter class[similarity low].
- Intra class[similarity high].

A standalone data mining tool in cluster analysis or pre-processing step for various algorithms In order to achieve data distribution. The term clustering is also called as unsupervised learning. "hidden patterns" are used in cluster analysis for machine learning. Clustering is simply defined as more number of attributes with large datasets. Clustering algorithm was brought into life for rapid growth in text mining. Spatial database and information retrieval.

into various clusters. Since on spatial data for optimum clusters there undergoes continuous research in data mining. Because of this clustering is an issue till dated in data mining. One of the first step in data mining analysis is clustering. For example, in an industry with a group of employees may need to know about the various works in their projects in order to check what are all products are completed and to be delivered and which are the project yet to be modified and delivered to the customers.

Clustering is a technique mainly used in agriculture science, monitors the quality of water change, and in precision agriculture to produce high yield. clustering is classified based on various methods such as

- Density based method.
- Partition based method.
- Hierarchical based method.

To make the concept clearer, we can take book management in the library as an example. In a library, there is a wide range of books on various topics available. The challenge is how to keep those books in such a way that readers can take several books on a particular topic without hassle. By using the clustering technique, we can keep books that have some kinds of similarities in one cluster or in one shelf and then label it with a meaningful name. If reader's want to grab books in that topic, they would only have to go to that shelf instead of looking for the entire library.

ASSOCIATION RULE MINING:

Association rule mining is a technique in data mining was developed by agrawal, imielinski and swami in 1993. This is one of the well organized technique of data mining to search the hidden or desired pattern among of data. The main focus in this method is to find relationship between various item in the relational database. Association rules are used to discover rules and to find the elements which occur recursively in a dataset consisting more absolute selections of element.

Association is a data mining technique that determines the possibility of the items which are co-occurred in a collection of data. Association rules are defined as the relationship between the co-occurring items. Hence the "sales transactions" are frequently analyzed using this technique. For processing numerical data association is the best technique. In a given set of transactions, find rules that will indicate the occurrences of an item based on the occurrences of other item. The issue to find all associated rules that satisfy minimum support for which user has specified. Association technique is known best and a straight forward data mining technique.

Strength measures of rules can be defined using two rules

- ➢ Support
- ➢ Confidence

Support:

Rules hold with support in T XUY is the sup percentage of transaction.

Sup=pr (xuy)

Confidence:

Rules holds T with confidence. Confidence percentage of transaction contain X and also contains Y.

Conf=pr (y|x)

For example transactional data:

ID	ITEM
1	Bread, jam
2	Bread, milk, cheese, egg
3	Jam, milk, coke, milkshake
4	Bread, jam, coke, egg
5	Bread, jam, milk, egg

Association rules generated based on the above transactional data.

Assume: minsup=30%

minconf=80%

{cheese, egg}→{milk}[sup=3/5;conf=3/3] for frequent item set

{bread, egg}→{milk}[sup=3/5;conf=3/3]

REGRESSION:

Regression analysis is a predictive modelling technique which gives the relation between the independent variable(y) and dependent variable(x). The variable that is been predicted are dependent variable and the variable which are predicted is used to predict the values of dependent variable is called independent variable. The

important tool for analyzing and modelling data is the regression.

Regression is one of the data mining technique that predicts the number. For example weight, height, income of a man. When the target values are known then the regression task begins with the help of dataset.

Regression analysis seeks to determine the values of parameters for a function that cause to the best fit a set of a data observations that you provide the following equation expresses these relationships in symbols. It shows that regression is the process of estimating the value of continuous target(y) as a function(f) of one (or) more predictions(x1,x2,x3,...xn) a set of parameters(Θ1,Θ2,Θ3,....Θn) and a measure of error(e).

Y=F(X,Θ) + e.

The predictors can be understood as independent variable and the target as the independent variable. The error, also called the residual, is the difference between the expected and the predicted value of the dependent variable. The regression parameters are also known as regression coefficients[**reference**].

For example relationship between the number of road accidents and rash driving by a driver is best analyzed by regression.

y-axis→ Rash drivers(Θ)
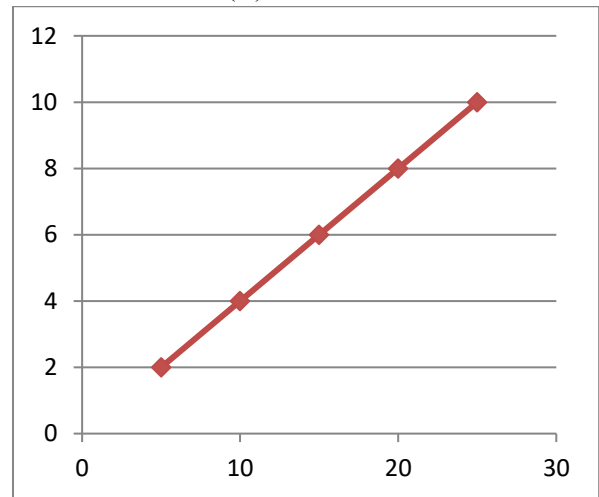
x-axis→ Number of accidents(x)

Multiple industries are using regression technique for financial forecasting, marketing and for trend analysis. For example regression is used in predicting a home's value based on various factors such as square feet, location and prices.

Comparison of data mining techniques:

Differentiator	Classification	Clustering	Association Rules	Regression
Methods	Predictive method.	Descriptive method.	Descriptive method.	Predictive method.
Usage	Used to predict the instance class from a pre- labelled instance.	Used to find the "natural" grouping of instances given un-labelled data.	Used to discover interesting relations between variables in large DB's.	Used to predict a continuous attribute.
Knowledge of classes	Yes.	No.	Yes.	Yes.
Algorithm	• Decision trees. • ANN. • Bayesian Classifier. • K-nearest. • Support vector.	• Hierarchical clustering. • Partition clustering. • Density clustering.	• WIC candidate generation. • WIO candidate generation.	• Linear regression. • Non-Linear regression. • Logical regression.
Data needs	Labelled samples.	Unlabelled samples.	labelled samples.	Labelled samples.
Learning method	Supervised learning.(Class labels of training data is known)	Unsupervised learning.(Class labels of training data is unknown)	Unsupervised learning.	Supervised learning.
Applications	➢ Remote sensing image. ➢ Disaster weather forecasting. ➢ Correlation analysis.	➢ Pattern recognition. ➢ Image analysis. ➢ Machine learning. ➢ Text mining. ➢ Whether report analysis.	➢ Efficient storage management. ➢ Prevention of inconsistencies. ➢ Index structures.	➢ Economic structure. ➢ Air pollution. ➢ Forecasting. ➢ Optimization.

III.   CONCLUSION

Data mining is the most integral component of all the databases for selecting the information from the data. This paper summarizes each and every different types of data mining techniques used in agriculture for decision making. This paper combines the works of many authors and it useful for the current circumstances in the agriculture domain. The main aim of this paper, is to upgrade the procedures of data mining techniques in agriculture. So that the farmer's get high production with supplementary profit.

REFERENCES

[1] https://www.google.co.in/search?q=data+mining&oq=data+mi&aqs=chrome.1.35i39l2j69i57j0j69i61l2.2754j0j4&client=ms-android-lenovo&sourceid=chrome-mobile&ie=UTF-8

[2] https://www.ukessays.com/essays/statistics/important-and-application-of-data-mining.php

[3] https://www.researchgate.net/publication/273261020_An_Implementation_of_Data_Mining_Techniques_and_Applications_in_Agriculture_on_Crop_Price_Prediction

[4] http://metahorizon.com/importance-of-data-and-data-mining/

[5] www.ukessays.com/essays/statistics/important-and-application-of-data-mining.php?essayad=carousel&utm_expid=309629-42.KXZ6CCs5RRCgVDyVYVWeng.1&utm_referrer=https%3A%2F

[6] http://www.comp.dit.ie/btierney/oracle11gdoc/datamine.111/b28129/regress.htm-regression

[7] https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/regress.htm#CIHHFFHB_regression

[8] https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/

[9] https://en.m.wikipedia.org/wiki/Regression_analysis

[10] http://www.tutorialspoint.com/data_mining/dm_classification_prediction.htm

[11] M.C.S.Geetha, Kumaraguru College of Technology "A Survey on Data Mining Techniques in Agriculture", Vol. 3, Issue 2, February 2015.

[12] Developing innovative applications in agriculture using data mining- Sally Jo Cunningham and Geoffrey Holmes.

[13] Different Techniques Used in Data Mining in Agriculture Jyotshna Solanki, Prof. (Dr.) Yusuf Mulge.

[14] A Brief survey of Data Mining Techniques Applied to Agricultural Data Hetal Patel Research Scholar

[15] Hetal Patel, Dharmendra Patel Assistant Professor. Charusat, Changa

[16] "A REVIEW ON DATA MINING ALGORITHMS", International Journal of Computer Applications ,Volume 95– No. 9, June 2014

[17] Petraq Papajorgji ,P. M. Pardalos,"A survey of data mining techniques applied to agriculture " journal of research gate.

[18] Dr. S.Vijayarani, Ms. A.Sakila, Bharathiar University, "MULTIMEDIA MINING RESEARCH – AN OVERVIEW" International Journal of Computer Graphics & Animation (IJCGA) Vol.5, No.1, January 2015.

[19] Raorane A.A, Kulkarni R.V, Vivekanand college, "Review- Role of data mining in agriculture", International Journal of Computer Science and Information Technologies, Vol.4(2),2013,270-272.

[20] G. Nasrin Fathima, R. Geetha, Jamal Mohamed college, "International Journal of Advanced Research in Computer Science and Software Engineering.

[21] Jaiswal, M., & Patel, D. (2015). Data Mining Techniques and Knowledge Discovery Database. *International Journal Of Research And Analytical Reviews (IJRAR)*, *2*(1), 248-259.

[22] Tan,Steinbach, "Kumar Introduction to Data Mining".

[23] FABRICIO VOZNIKA, LEONARDO VIANA, "DATA MINING CLASSIFICATION ".

[24] Aastha Joshi, Rajneet Kaur, Sri Guru Granth Sahib World University, A Review: Comparative Study of Various Clustering Techniques in Data Mining, International Journal of Advanced Research in Computer Science and Software Engineering.

[25] Pavel Berkhin, Accrue Software, Inc, Survey of Clustering Data Mining Techniques.

[26] Periklis Andritsos, University of Toronto, DataClustering Techniques.

[27] Tan, Steinbach, Kumar, Introduction to Data Mining.

[28] Swati Gupta Assistant Professor, Amity University Haryana, A Regression Modeling Technique on Data Mining.

[29] Bob Stine Department of Statistics The Wharton School of the Univ of Pennsylvania, Data Mining with Regression.

[30] Dr L. Arockiam, S.S.Baskar, L. Jeyasimman, Clustering Techniques in Data mining.

[31] https://www.google.co.in/search?q=images+of+difference+between+association+rule+based+and+regression+data+mining+technique&client=ms-android-sonymobile&source=lnms&tbm=isch&sa=X&ved=0ahUKEwiryqil_K3SAhVEGpQKHVBHA1MQ_AUIMigB