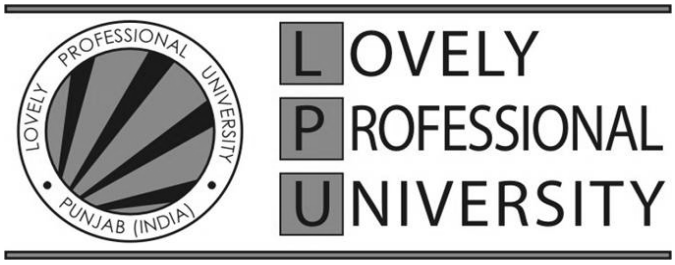

Mining of agriculture data and to do provide recommendation on the basis of same

A Dissertation Proposal

submitted

By

Gulab Singh

to

Department of Computer Science & Technology

In partial fulfilment of the Requirement for

the

Award of the Degree of

Master of Technology

in Computer Science &

Technology

Under the guidance of

Mr. Barjinder Singh

(November 2017)

TOPIC APPROVAL PERFORMA

School of Computer Science and Engineering

Program : P172::M.Tech. (Computer Science and Engineering) [Full Time]

COURSE CODE : CSE548 **REGULAR/BACKLOG :** Regular **GROUP NUMBER :** CSERGD0052

Supervisor Name : Barjinder Singh **UID :** 16469 **Designation :** Assistant Professor
Qualification : _____ **Research Experience :**
_____

SR.NO.	NAME OF STUDENT	REGISTRATION NO	BATCH	SECTION	CONTACT NUMBER
1	Gulab Singh	11607947	2016	K1637	9780561015

SPECIALIZATION AREA : Database Systems **Supervisor Signature:** _____

PROPOSED TOPIC : Mining of agriculture data and to do provide recommendation on the basis of same

Qualitative Assessment of Proposed Topic by PAC		
Sr.No.	**Parameter**	**Rating (out of 10)**
1	Project Novelty: Potential of the project to create new knowledge	6.67
2	Project Feasibility: Project can be timely carried out in-house with low-cost and available resources in the University by the students.	6.33
3	Project Academic Inputs: Project topic is relevant and makes extensive use of academic inputs in UG program and serves as a culminating effort for core study area of the degree program.	7.00
4	Project Supervision: Project supervisor's is technically competent to guide students, resolve any issues, and impart necessary skills.	7.00
5	Social Applicability: Project work intends to solve a practical problem.	7.33
6	Future Scope: Project has potential to become basis of future research work, publication or patent.	7.00

PAC Co ıмittee Members		
PAC Member 1 Name: Kewal Krishan	UID: 11179	Recommended (Y/N): Yes
PAC Member 2 Name: Raj Karan Singh	UID: 14307	Recommended (Y/N): NA
PAC Member 3 Name: Sawal Tandon	UID: 14770	Recommended (Y/N): NA
PAC Member 4 Name: Dr. Pooja Gupta	UID: 19580	Recommended (Y/N): Yes
PAC Member 5 Name: Kamlesh Lakhwani	UID: 20980	Recommended (Y/N): NA
PAC Member 6 Name: Dr.Priyanka Chawla	UID: 22046	Recommended (Y/N): NO
DAA Nominee Name: Kuldeep Kumar Kushwaha	UID: 17118	Recommended (Y/N): NA

Final Topic Approved by PAC: **Mining of agriculture data and to do provide recommendation on the basis of same**

Overall Remarks: Approved

PAC CHAIRPERSON Name: 11024::Amandeep Nagpal **Approval Date:** 04 Nov 2017

Abstract

In the modern era, competition is in every field, it might be agricultural, educational, financial, marketing, etc. For persons having occupations related to agriculture, they need maximum profit with less invested input. The profit may be in terms of money, production, and pest and disease management. The profit has a different aspect with different kind of persons involved in this occupation. For farmers, it has a different aspect, for market vendors it has a different aspect. To achieve these requirements, a lot of researchers has been involved in this particular area to maximize the results by analyzing the trends of previous 50 years and identify the parameters affecting the productivity. This paper is an attempt to achieve those desired outputs with help of various data mining techniques.

CERTIFICATE

This is to certify that Gulab Singh has completed dissertation proposal titled Mining of agriculture data and to do provide recommendation on the basis of same under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original investigation and study. No part of the dissertation has ever been submitted for any other degree or diploma.

Date: **Signature of Advisor**

Name: Barjinder Singh

ACKNOWLEDGEMENT

I take this opportunity to present my votes of thanks to all those guidepost who really acted as lightening pillars to enlighten our way throughout this project that has led to successful and satisfactory completion of this study.

I am grateful to our **Lovely Professional Universit**y for providing me with an opportunity to undertake this project in this university and providing us with all the facilities.

I am really thankful from my heart to **Mr. Barjinder Singh**, who allowed me to do this project under his guidance. I am highly thankful to my family and friends for their active support, valuable time and advice, whole hearted guidance, sincere cooperation and pains-taking involvement during the study.

Lastly, I thankful to all those, particularly the various friends, who have been instrumental in creating proper, healthy and conductive environment and including new and fresh innovative ideas during the project, without their help, it would have been extremely difficult for me to prepare the project in a time bound framework.

DECLARATION

I hereby declare that the dissertation proposal entitled, Mining of agriculture data and to do provide recommendation on the basis of same submitted for the M.Tech Degree is entirely my original work and all ideas and references have been duly acknowledged. It does not contain any work for the award of any other degree or diploma.

Table of Contents

List of Figures

Chapter 1
INTRODUCTION

Data mining in agriculture is a very modern research topic. It contains in the application of data mining techniques to agriculture. Recent technologies are currently able to provide a lot of statistics on agricultural-related activities, which can then be studied in order to find significant knowledge. A related, but not equal term is precision agriculture. Mining and agriculture are straightforwardly interrelated through farming's reliance on mined data sources, land and water assets, and specialists. They are additionally by implication associated where mining firms have improved foundation in a way that arrangements agrarian advancement. The results of this cooperation seem assorted. There is flag that agribusiness is expanding in a few zones as a result of mining and falling in others, subject on neighborhood conditions. AGRICULTURE MINING- The meaning of agriculture mining remains same as data mining but due to agriculture data, we give it a different name-agriculture mining. Now a days there are various research departments and institutes are working in this field. They perform a various experiment and analyze the output. Then the output of various experiments is compared and if they are successful in their experiments, then the similar activity chart is launched to market so that various farmers can get benefit out of it. In agriculture mining, the researcher analyzes the agriculture data of previous 50 years and get various trends. Further, these trends are compared and effective predictions are drawn out. In this paper, a technique of information mining and learning revelation process in expansive, appropriated and heterogeneous databases will be characterized. After data collection, we will be performing some another process, which will produce some trends that

can be used for various predictions required for maximizing profits for persons involved in agriculture occupation.

I. PROCESS OF KNOWLEDGE DISCOVERY

- Selection- Data is selected from various databases.
- Preprocessing- Incomplete, Inconsistent data is completed.
- Transformation- Transforming raw data into the required form.
- Data Mining- After analyzing data patterns are drawn.
- Interpretation- Patterns are interpreted to get knowledge.

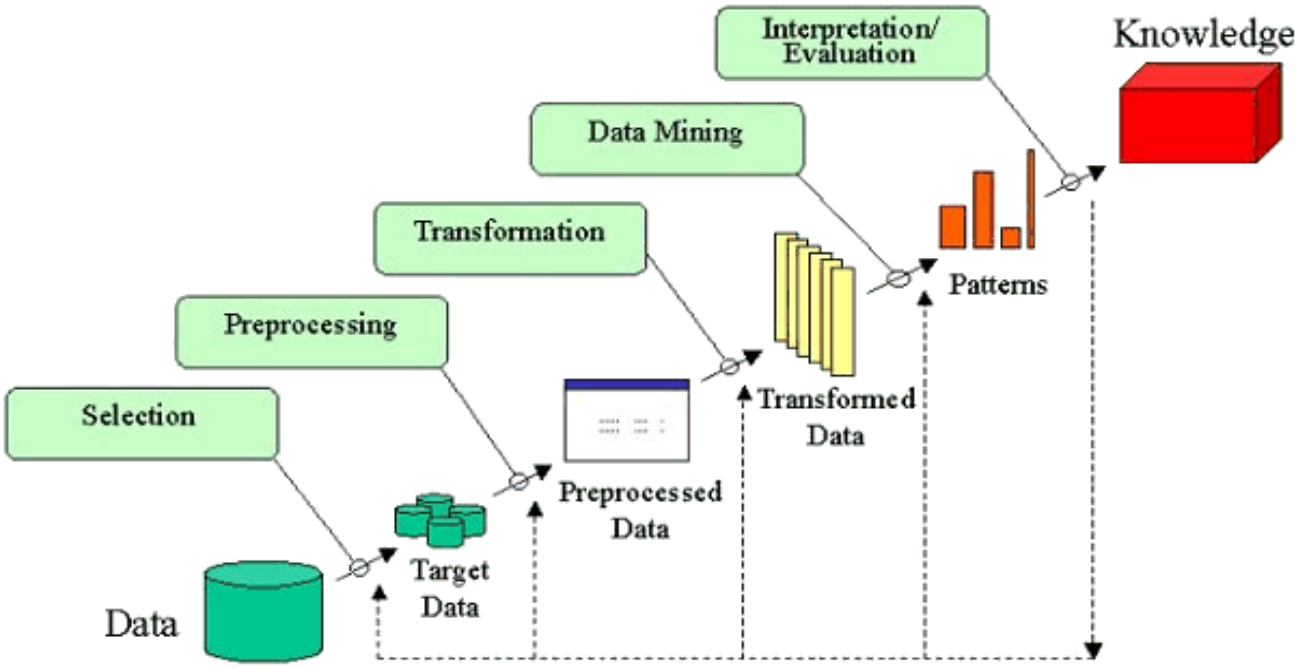

Figure 1.1 Process of KDD

II. ALGORITHMS USED

- Association: In this relation is identified between different dimensions.
- Clustering: In this items of similar attributes are grouped together to form a cluster.
- Regression: In this value of the dependent variable is obtained as a function of

independent variables and set of parameters.

- Classification: In this different items are classified into the class using classifier.
- K-means: It is similar to clustering with a difference that items will be clustered into k groups.

Regression- Regression is an information mining capacity that predicts a number. A Regression work begins with an informational index in which the objective esteems are known. In the model form (preparing) process, a Regression algorithm surmises the estimation of the objective as a component of the indicators for each case in the fabricate information. These relations amongst indicators and target are précised in a model, which would then be able to be connected to an alternate informational index in which the target esteems are obscure. Regression models are tried by registering diverse insights that measure the fluctuation between the anticipated esteems and the normal esteems. The past information for a Regression venture is normally isolated into two informational collections: one for developing the display, the other for investigation the model.

- Linear Regression

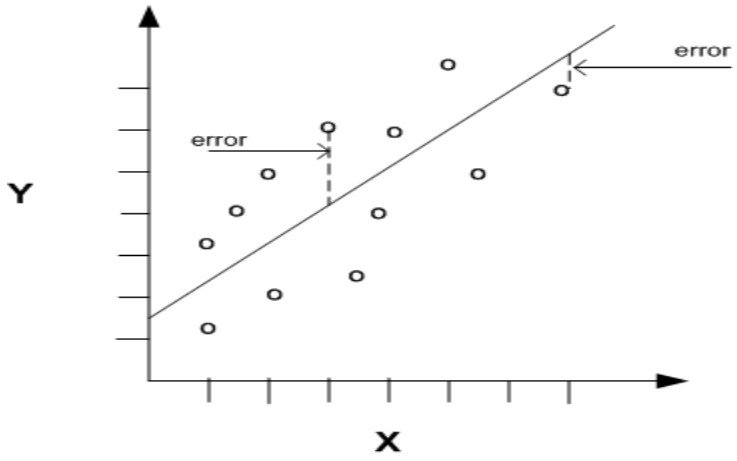

Figure 1.2 Linear Regression with single predictor

- Non Linear Regression

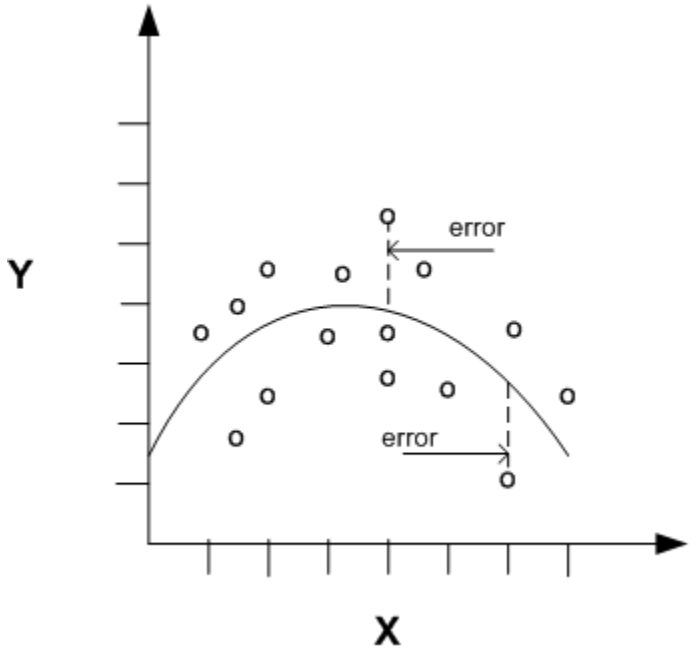

Figure 1.3 Non Linear Regression with single predictor

Classification- Classification is an information mining capacity that distributes things in a gathering to target classifications or classes. The goal of arrangement is to correctly expect the objective class for each case in the information. A grouping work starts with an informational collection in which the class assignments are known. Arrangements are particular and don't suggest arrange. Consistent, drifting point esteems would demonstrate a numerical, as opposed to an all-out, target. A prescient model with a numerical point utilizes a relapse algorithm, not a characterization algorithm. In the model develop (preparing) process, a characterization algorithm finds relations between the estimations of the indicators and the estimations of the objective. Characterization models are tried by contrasting the anticipated esteems with known target esteems in an arrangement of test information. The old information for an order venture is normally isolated into two informational collections: one for development the model; the other for examination the

model.

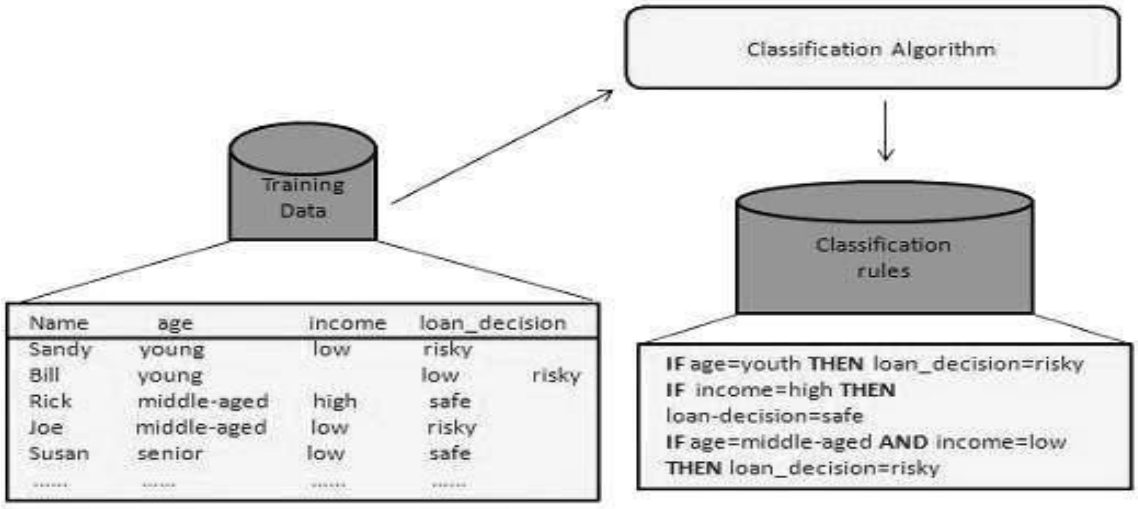

Figure 1.4 Classification Model

Clustering- Clustering examination finds bunches of information questions that are similar in some sense to each other. The members of a bunch are more similar to each other than they resemble members of different bunches. The point of bunching investigation is to find high caliber bunches with the end goal that the between group similarity is low and the intra-group resemblance is high. Bunching, similar to grouping, is utilized to part the information. Not at all like arrangement, has grouping models piece information into bunched that were not beforehand characterized. Order models piece information by dispensing it to already characterized classes, which are expressed in a target. Bunching models don't utilize an objective. Bunching is useful for finding information. In the event that there are various cases and no undeniable organizations together, grouping algorithms can be utilized to find characteristic groupings. Bunching can likewise help as a valuable information preprocessing advance to perceive homogeneous bunches on which to fabricate regulated models. Bunching can likewise be utilized for peculiarity recognition. Once the information has been divided into groups, you may find that a few cases don't reasonable well into any groups. These cases

are irregularities or anomalies.

Association- Association is an information mining errand that decides the likelihood of the co- event of things in an accumulation. The relations between co-happening things are expressed as Association rules. Association rules are regularly used to analyze deals exchanges.

III. PROCESS OF FUZZY REASONING

- Fuzzification- It is the initial phase in the fuzzy inferencing process. This includes an area change where fresh information sources are changed into fuzzy data sources.
- Rule evaluation- It comprises of a progression of IF-THEN guidelines. A decision structure to decide the standards require recognition with the framework and its coveted operation.
- Defuzzification- It includes the way toward transposing the fuzzy yields to fresh yields.

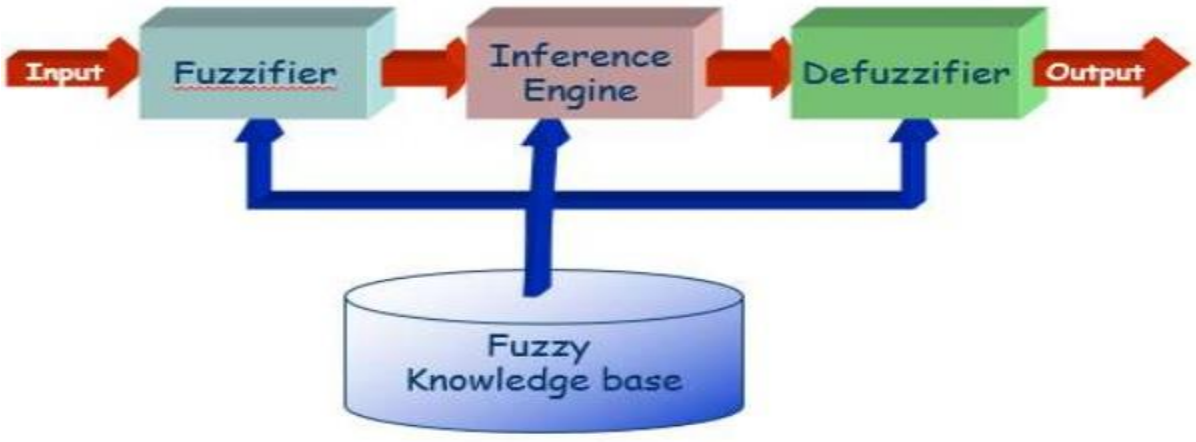

Figure 1.5 Process of fuzzy reasoning

Chapter 2

REVIEW OF LITERATURE

1) A review of the application of data mining techniques for decision making in agriculture

This paper delivers a review of research on the application of data mining methods for decision making in agriculture. The paper accounts the application of a number of data mining methods including artificial neural networks, Bayesian networks and support vector machines. The review has charted a number of capable techniques that have been used to recognize the relationships of various climate and other factors on crop production. This review suggests that further examinations are needed to understand how these techniques can be used with complex agricultural datasets for crop yield prediction integrating seasonal and spatial factors by using GIS technologies [1].

- **Artificial Neural Network (ANN) in Agriculture**

A model can be established using interrelationships of associated variables. These figuratively represent the connected processing neurons or nodes of the human brain that are used. To form model prediction, bulky number of input and output example are used to progress a formula to discover the relationship. With little a priori information of the functional relationship, nonlinear relationships, which are ignored by other prediction methods, can be determined.

- **Bayesian Networks (BN) in Agriculture**

BN is a technique for expressive beliefs and knowledge using probabilities, especially significant for systems that are extremely complex in their structure and functional interactions. BN uses the probabilistic components of a structure, as contrasting to deterministic comparisons to describe the connections among variables.

- **Support Vector Machines (SVM) in Agriculture**

SVMs are one of the latest supervised machine learning systems. A number of studies have stated on the application of SVMs in an agricultural context. The application reported for SVM was the modeling of urban land use conversion. This research derived the association between rural-urban land use change and various factors. SVM was also applied to deliver visions into crop response patterns associated to climate conditions by providing the structures contribution analysis for agricultural yield prediction analyzed different possible deviations of the weather scenarios using SVMs.

- **Association Rule Mining in Agriculture**

The application of Association Rule Mining have been revealed for many circumstances in agriculture for clarifying hidden patterns and connections between different climate and crop production.

2) Data Mining and Wireless Sensor Network for Agriculture Pest/Disease Predictions

Information mining strategies were utilized to attempt the information into valuable learning/relations/patterns and relationship of harvest climate bug/sickness run. These elements found from the information mining strategies and master through numerical models were affirmed with comparing observation information [2].

Materials and Methods

- Sensory Data collection
- Surveillance Data Collection
- Data Mining and Statistical Models

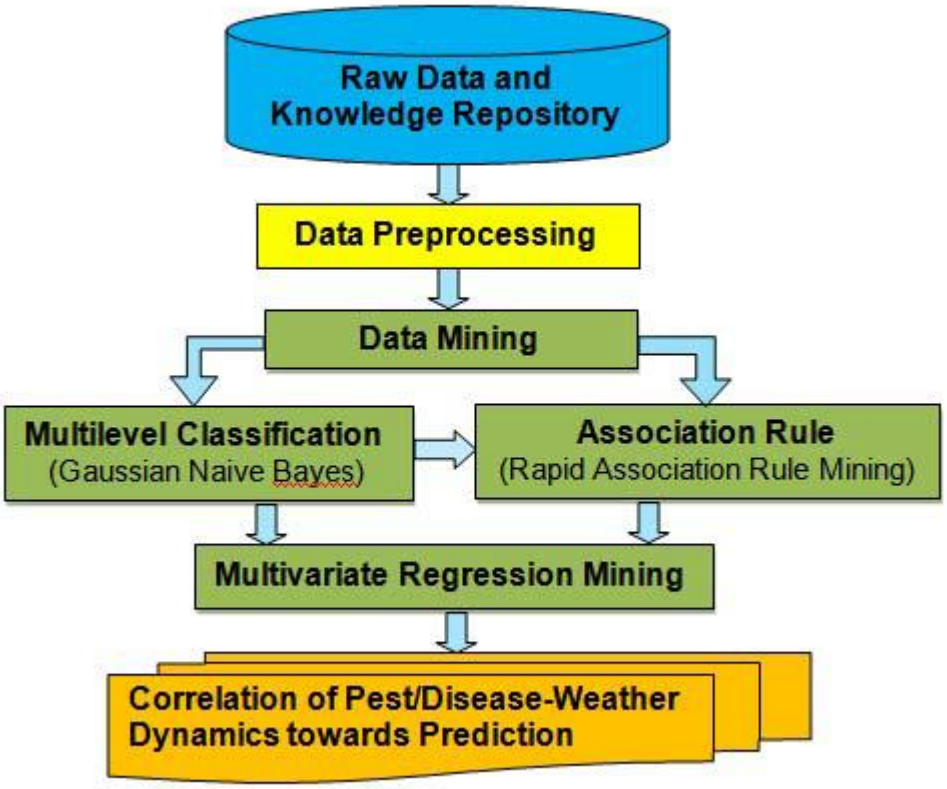

Figure 2.1 DM processing flow for pest/disease dynamics

3) Geospatial Data Mining for Agriculture Pest Management - a Framework

This paper terms the casing work of the pest administration framework utilizing information mining strategies; which concentrate on giving, authentic information, current and prescribed pest also, pesticide information and to be repeated pest models up to cultivate level. In this work, an exertion has been made to indicate how geospatial information mining joined with agriculture counting irritation examination, pesticide and climatological parameters are valuable for advancement of pesticide utilization and better administration. The results will uncover energizing examples of agriculturist rehearses sideways with pesticide use elements both in spatial and non-spatial way and can help to make out the clarifications for pest and pesticide abuse [3].

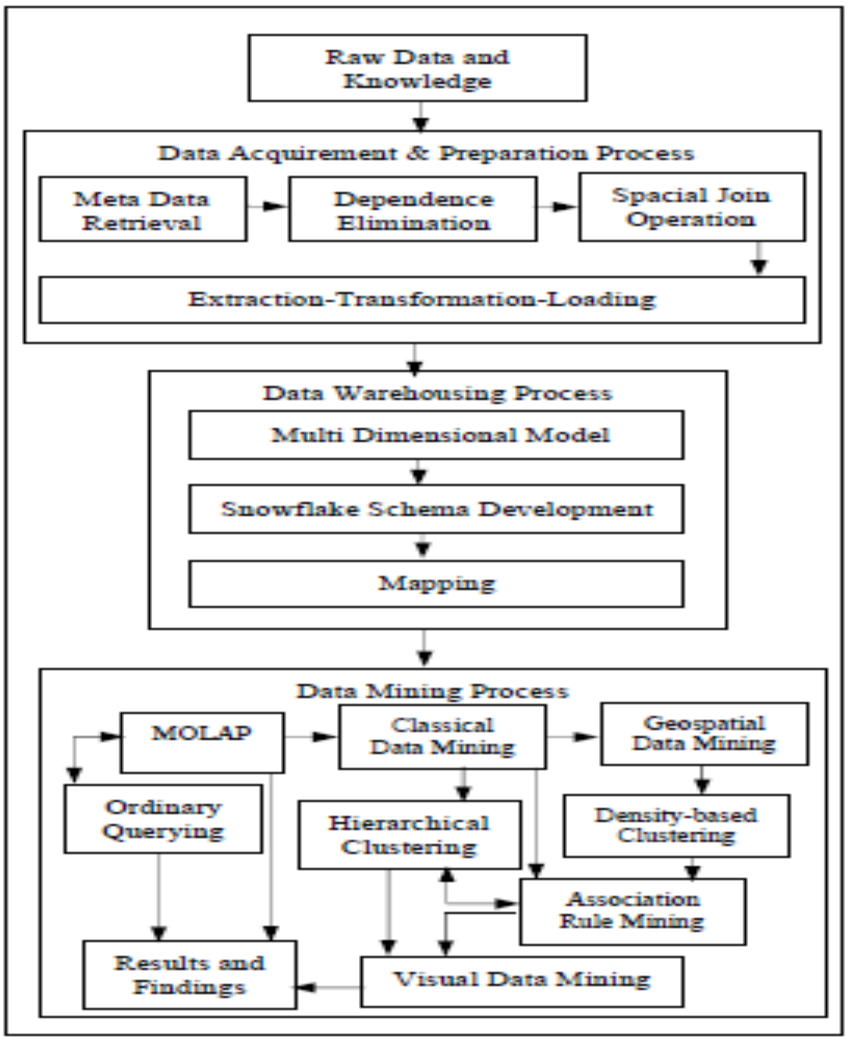

Figure 2.2 A pest management Frame work model

4) An Expert System Based on Spatial Data Mining Used Decision Tree for Agriculture Land Grading

This paper presents spatial information mining technique particularly decision tree algorithm applying to agriculture arrive reviewing. The clue is to pool spatial information mining/decision tree strategies with master framework strategies and apply them to shape a savvy agriculture arrive reviewing data framework. The creator execute decision tree c4.5 algorithm and actualize with Mo2.0 and VC++6.0 to assemble agribusiness arrive evaluating master framework. Likewise, an investigation is offered to demonstrate the specific

advantages of this system in tending to troubles in arrive evaluating, for example, missing area data, and challenges in quantitative investigation of components [4].

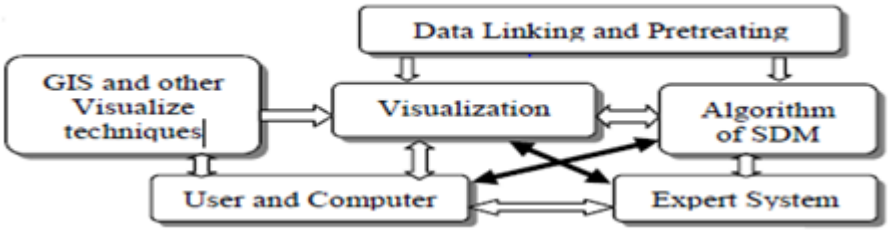

Figure 2.3 The basic function model of system

5) A Methodology and Life Cycle Model for Data Mining and Knowledge Discovery in Precision Agriculture

This paper offers a strategy for information mining and learning disclosure in huge, appropriated and heterogeneous databases. Keeping in mind the end goal to pick up conceivably energizing examples, connections, and standards from such vast and heterogeneous information accumulations, it is important that a procedure be built up to take profit of the suite of existing methodologies and apparatuses available for information mining and learning revelation in databases [5].

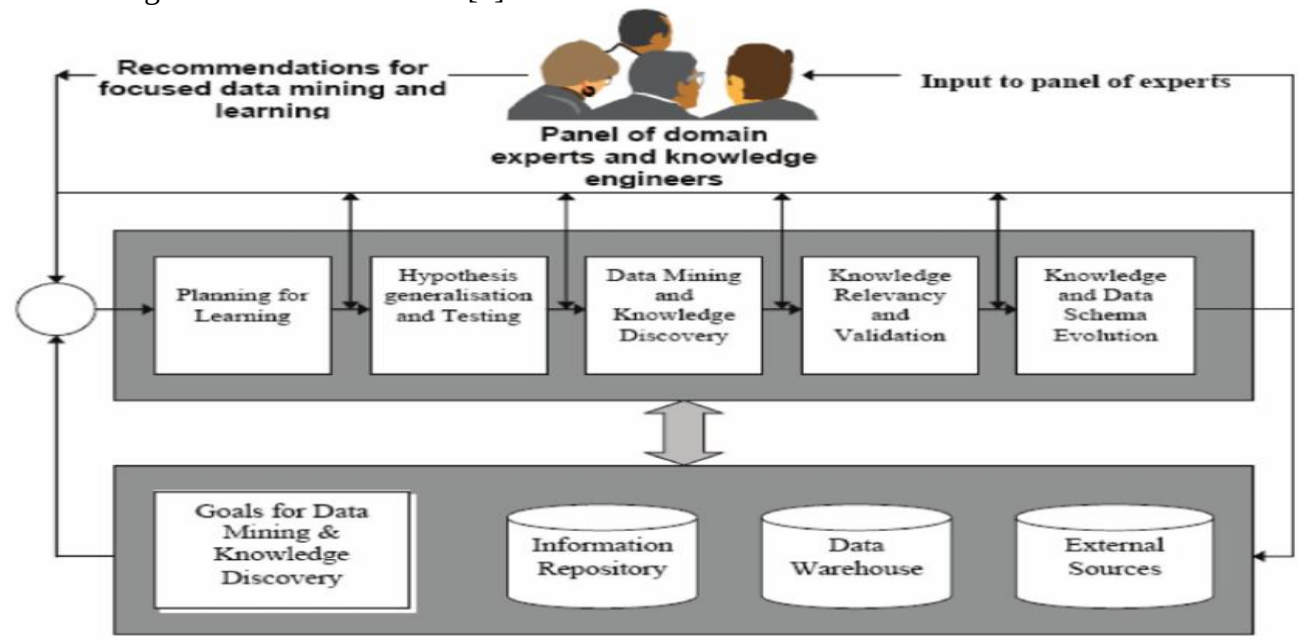

Figure 2.4 The knowledge discovery process model

6) Crop Recommendation System for Precision Agriculture

Data mining is the rehearsal of examining and developing purposeful knowledge from the data. Data mining discovers its application in various fields like finance, retail, medicine, agriculture etc. Data mining in agriculture is used for investigating the various biotic and abiotic factors. Agriculture in India plays a major role in economy and employment. The joint problem existing among the Indian farmers are they don't pick the right crop based on their soil requirements. Due to this they face a serious obstacle in productivity. This problem of the farmers has been talked through precision agriculture. Precision agriculture is a modern farming method that uses research data of soil characteristics, soil types, crop yield data collection and proposes the farmers the right crop based on their site specific parameters. This decreases the wrong decision on a crop and increase in productivity [6].

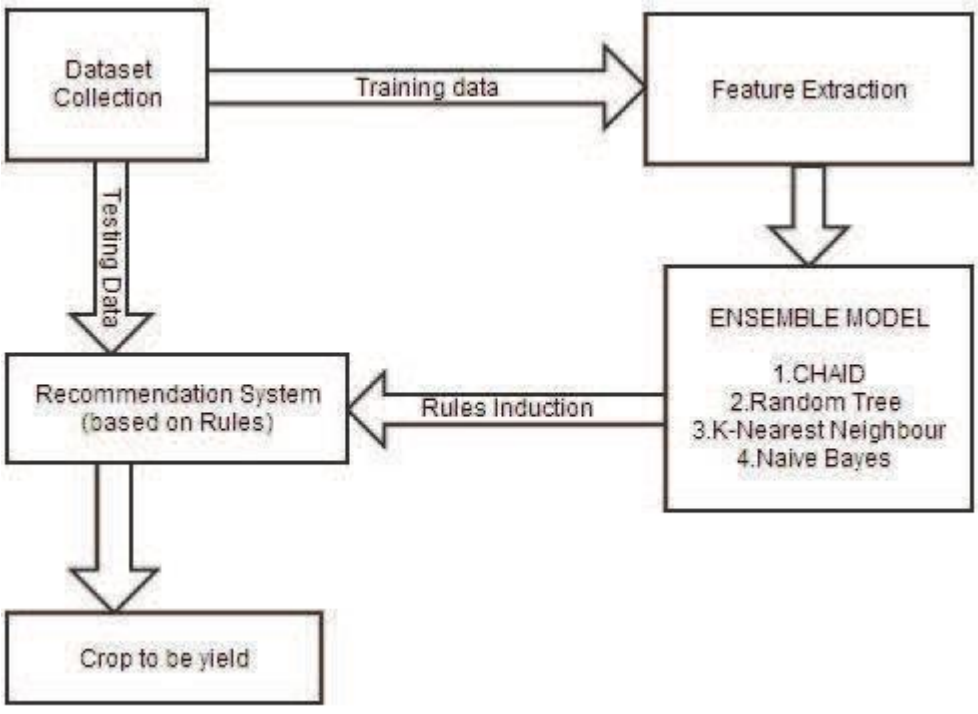

Figure 2.5 Depicts the overall methodology of proposed system

7) Knowledge Based Real Time Monitoring System for Aquaculture Using IoT

Internet of things is one of the quickly growing fields for conveying social and economic profits for evolving and developing economy. The field of IOT is increasing its wings in all the areas like medical industrial, transportation, education, mining etc. Today with the progression in integrated on chip computers like Arduino, Raspberry pi the technology is getting the ground level with its presentation in agriculture and aquaculture. Water quality is serious factor while culturing aquatic organisms. It mainly depends on numerous factors similar dissolved oxygen, ammonia, pH, temperature, salt, nitrates, carbonates etc. The quality of water is observed continuously with the help of sensors to guarantee growth and survival of aquatic life. The sensed data is shifted to the aqua farmer mobile through cloud. As a consequence preventive measures can be taken in time to reduce the losses and increase the productivity [7].

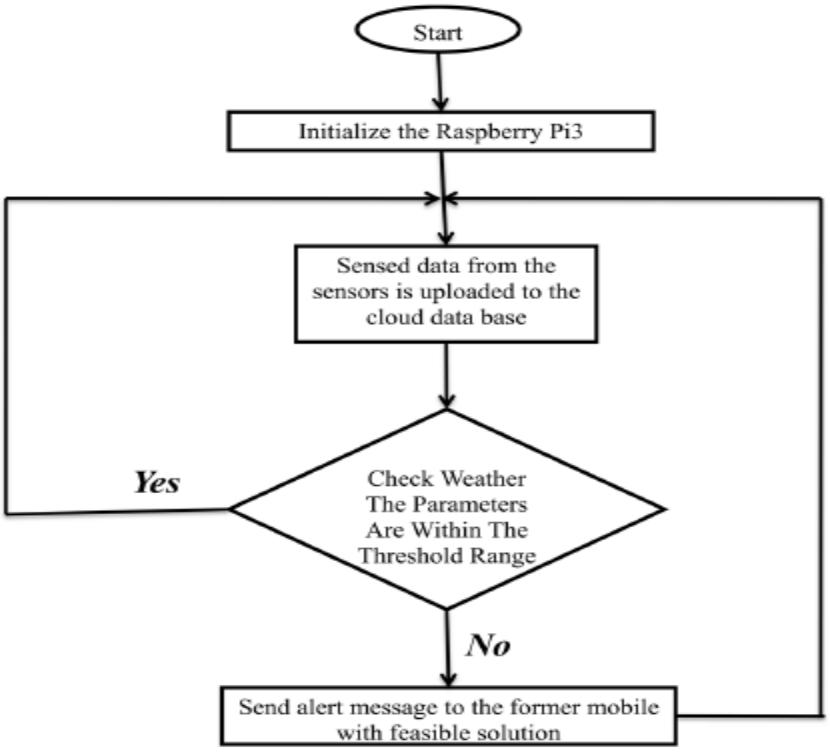

Figure 2.6 Flow chart for the proposed model

8) Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh

Agricultural yield creation relies upon a few perspectives, for example, science, atmosphere, and economy furthermore, geology. A few highlights effect sly affect agriculture, which can be checked utilizing appropriate measurable strategies. Utilizing such approaches and techniques on authentic yield of harvests, it is conceivable to get data or information which can be accommodating to agriculturists and government associations for building more advantageous choices and strategies which prompt enhanced generation. In this paper, our exertion is on use of information mining strategies to extricate information from the agricultural information to figure edit yield for chief cereal crops in areas of Bangladesh [8]. As indicated by this speculation, we described our chose traits for the worry of bunching the areas as takes after:

- Cluster Type-1 depends on the accompanying properties: Rainfall, least temperature, greatest temperature, dampness and daylight. These are the ecological or climatic qualities considered for our examination. The level of similitude of the gathering of these traits ought to demonstrate unmistakable bunches for the chose areas.
- Cluster Type-2 depends on the accompanying traits: soil pH and soil saltiness. As talked about before, these biotic variables contribute to a great extent towards the forecast of the crops. Closeness of the estimations of these traits ought to likewise show independent bunches.
- Cluster Type-3 depends on irrigated zone. Bunching depends on the range characteristics for each region was considered in light of the fact that we can get the different bunches in light of particular scopes of territories that were watered for each region.

- Cluster Type-4 depends on the individual harvest yields of potato and wheat. This sort of grouping was considered so as to order the areas into isolate groups with comparable harvest yields-and after examination of the outcomes, to see regardless of whether they show an example identified with impacts from the chose properties.

9) Data mining model for early fruit diseases detection

Automatic methods for a timely detection of plant diseases could be vigorous for exact fruit protection. Traditionally the agriculture expert's knowledge is descriptive and experiment based, therefore it is tough to describe it mathematically and consequently build decision system which can replace it. Key factors of decision based fruit protection system could fluctuate for classes of plants and diseases. However, such structures are very rare and very complex, and in many cases planned just for one plant class. For actual diseases protection of fruit, meteorological data and data about the disease presence are the most important. In this paper writers propose one clue for data mining based system for detection of possible fruit infection. For this persistence, different types of data mining methods were evaluated on exclusive data sets [9].

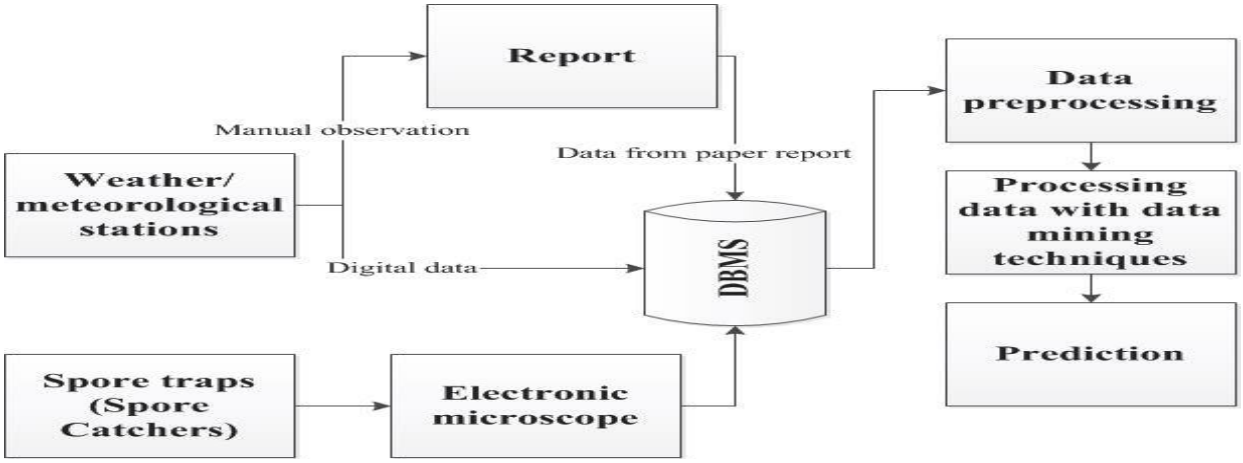

Figure 2.7 Block schema for the proposed system

10) Agriculture Data for All - Integrated Tools for Agriculture Data Integration, Analytics and Sharing

Agribusiness yields a lot of information in both open and private areas. Such information include: national soil databases, long haul information on carbon adjust crosswise over different atmosphere zones and vegetative land covers, computerized rise models, provincial and national inventories, remote detecting information, geophysical information, financial and various other informational indexes. These farming related records are interested not exclusively to the agribusiness part. Biology, condition, business, strategy, different sciences, and so on can utilize this information for their developments. They can investigate the effect of land-administration approaches, for example, preparation, brushing, water system, and that's only the tip of the iceberg. Agriculture information investigation can help perceive the issues and lead approach creators to applying hazard moderation and reclamation procedures [10].

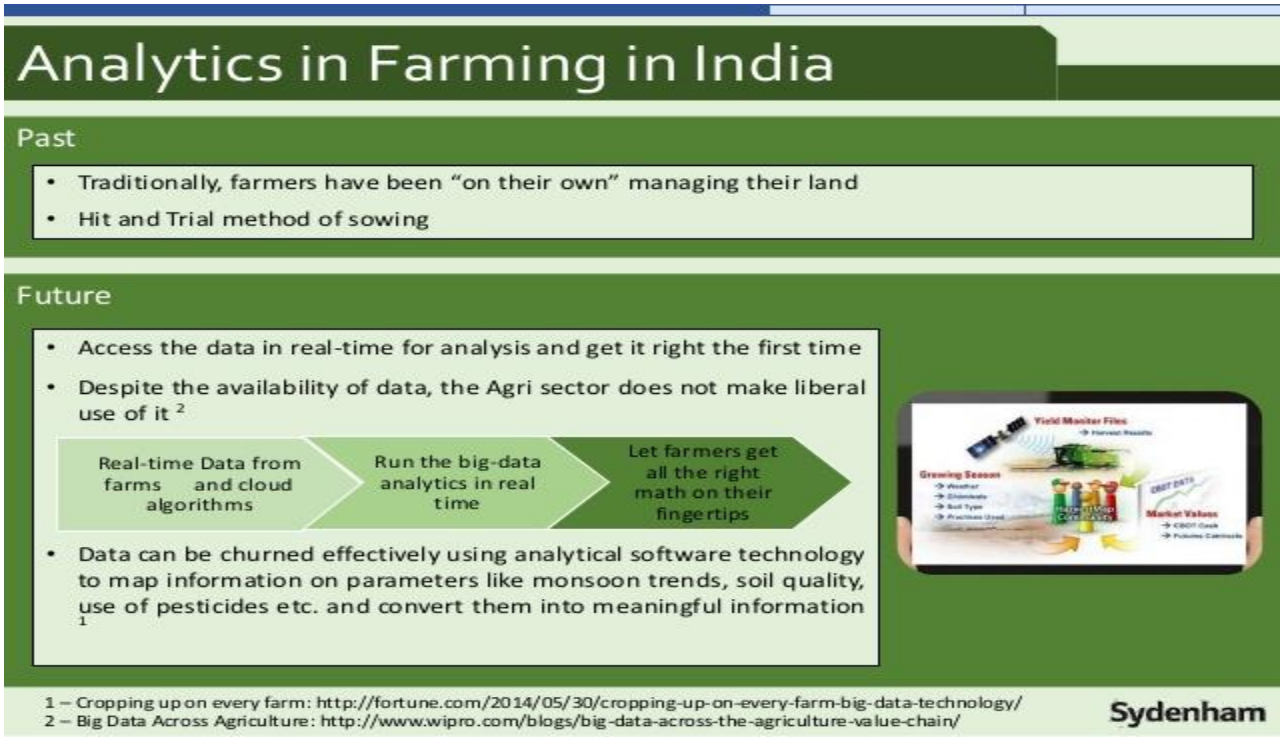

Figure 2.8 Agriculture data for all

Chapter 3
Scope of Study

We will be collecting data from surveys and different official websites. After this, we will be selecting the required data. The data will be extracted from the databases, data files, etc. Now the extracted data might be inconsistent, incomplete data. We have to complete data and make it consistent. After preprocessing of data, data would be transformed into the required format. It will depend on us that what kind of format of data is suitable for us. After transforming data in a required format, the process of data mining will start. In this process, we will try to figure out different kind of patterns using suitable algorithms. It will depend on us that what kind of algorithm we are going to choose. After having different patterns, we will try to analyze those patterns and also try to interpret the hidden knowledge inside those patterns. That knowledge can be used by various scholars involved in the area of agriculture. This will help them to achieve their objectives respectively. Later on a graphical user interface (GUI) will be made with help of MATLAB. The knowledge extracted from those patterns will be used for making fuzzy logics. The GUI can be used by anyone interested in this field. In GUI the inputs can be given and corresponding output will be determined. The output will be generated through inference of fuzzy process. The fuzzy process will generate an output.

Chapter 4
Objectives of the Study

In the modern era, competition is in every field, it might be agricultural, educational, financial, marketing, etc. For persons having occupations related to agriculture, they need maximum profit with less invested input. The profit may be in terms of money, production, and pest and disease management. The profit has a different aspect with different kind of persons involved in this occupation. This paper is an attempt to achieve those desired outputs with help of various data mining techniques.

- To enhance productivity of crops by finding various attributes associated (soil types, weather conditions, etc.)
- To find best insecticide/pesticide for insect/pest management.
- To find various diseases available in crops.

Chapter 5
Research Methodology

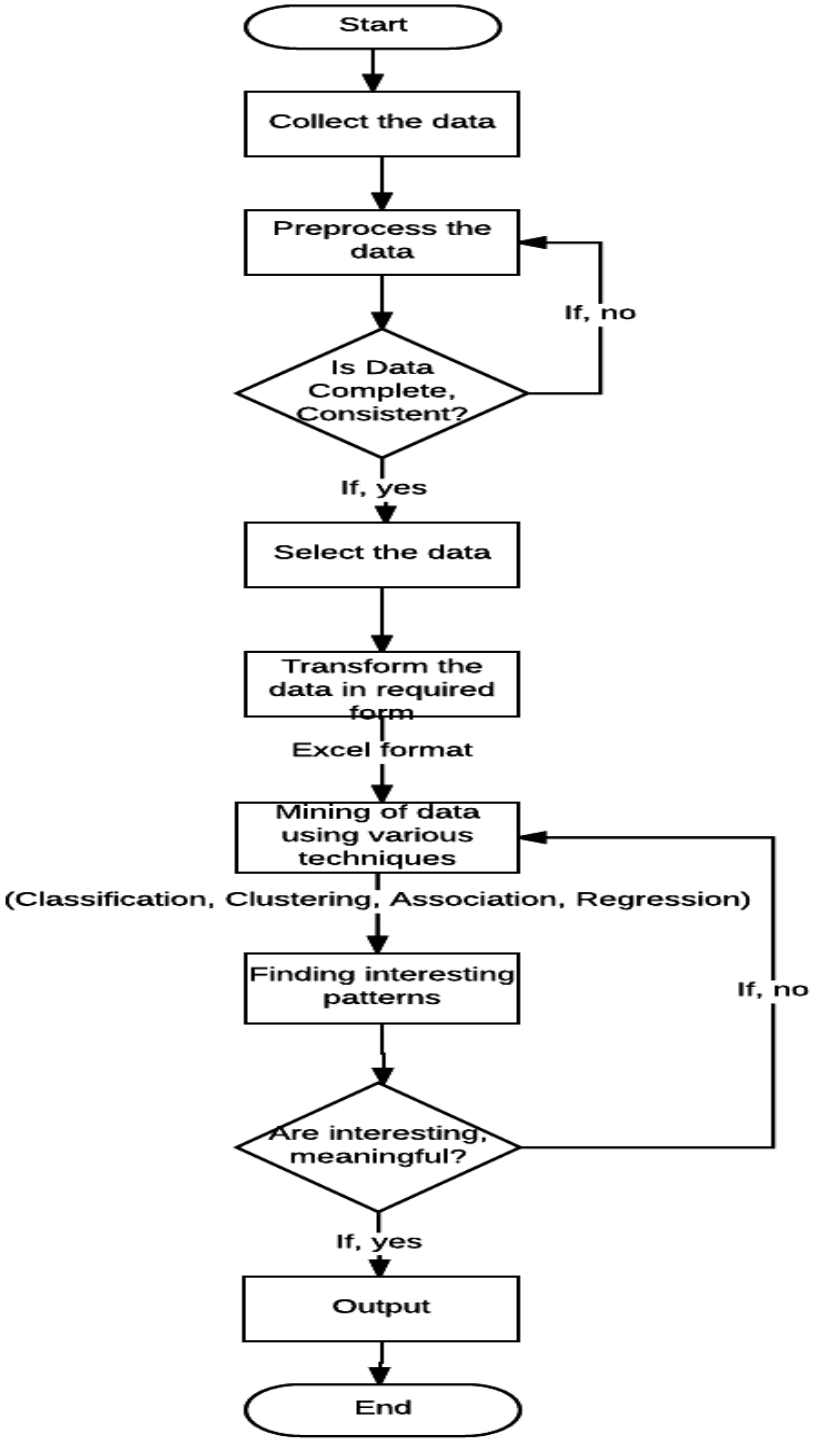

Figure 5.1 Flowchart of extracting knowledge

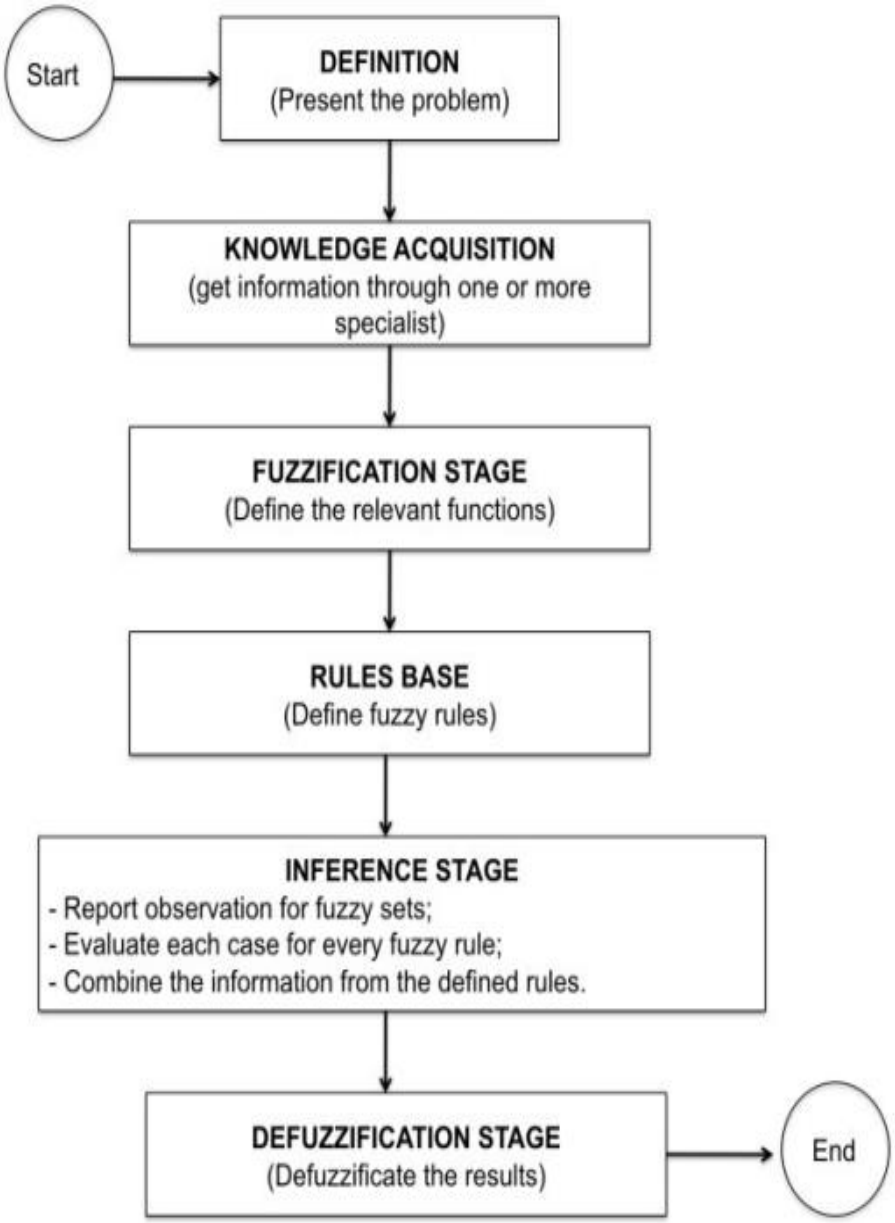

Figure 5.2 Flowchart for inferencing from expert system

Chapter 6
Summary and Conclusions

India is a country in which agriculture plays a major role. In the success of the farmers, grows the nation. Agriculture Mining will help us in the analysis of various data related to agriculture. In this way, farmers can choose the right seed based on soil requirements to increase productivity and obtain profit out of such a technique. Thus the farmers can plant the right crop increasing his yield and also aggregate the overall productivity of the nation. Not only farmers, the other persons involved in other fields related to agriculture such as (market vendors, financier, etc.) can take benefit from this proposed system. We will use various trends which are present in the previous data set and that will help us to provide the efficient and more realistic outcomes regarding various aspects. Further, these outcomes can be used by various scholars who want to involve themselves in the area of agriculture to achieve their respective objectives.

List of References

[1] L. J. Armstrong, “A review of the application of data mining techniques for decision making in agriculture,” pp. 1–6, 2016.

[2] A. K. Tripathy *et al.*, “Data Mining and Wireless Sensor Network for Agriculture Pest / Disease Predictions,” pp. 1229–1234, 2011.

[3] A. K. Tripathy, J. Adinarayana, and D. Sudharsan, “Geospatial Data Mining for Agriculture Pest Management - a Framework.”

[4] Z. Jia, “An Expert System Based on Spatial Data Mining Used Decision Tree for Agriculture Land Grading,” pp. 2–5, 2009.

[5] S. W. Lee and L. Kerschberg, “A ~ e ~ ~ o d o l and ogy Life Cycle Model for and Knowledge Discovery in Precision,” pp. 2882–2887, 1998.

[6] I. Eighth, I. Conference, and A. Computing, “Crop Recommendation System for Precision Agriculture,” pp. 32–36, 2016.

[7] K. R. Sita and R. Raju, “Knowledge Based Real Time Monitoring System for Aquaculture Using IoT,” pp. 7–10, 2017.

[8] A. T. M. S. Ahamed *et al.*, “Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh,” 2015.

[9] M. Ilic, P. Spalevic, M. Veinovic, and A. A. M. Ennaas, “Data mining model for early fruit diseases detection,” pp. 910–913, 2015.

[10] J. Nabrzyski, C. Liu, C. Vardeman, S. Gesing, and M. Budhatoki, “Agriculture Data for All - Integrated Tools for Agriculture Data Integration, Analytics, and Sharing,” 2014.