CHAPTER II

LITERATURE SURVEY
CHAPTER II

LITERATURE SURVEY

Mining Association Rules is one of the most important field of application of Data Mining. A set of customer transactions on items is provided and the main purpose is to determine the correlations within the sales of items. Mining association rules, is known as Market Basket Analysis, which is also an application field of Data Mining.

It is essential to examine the customer’s purchase behavior and assist in increasing the sales and conserve inventory by focusing on the point of sale transaction data. This works as a wide area for the researchers to increase a better data mining algorithm. This chapter discusses a survey about the existing Association Rule Mining, Apriori Algorithm, Market Basket Analysis system, and agriculture and data mining techniques.

2.1. SURVEY ON ASSOCIATION RULE MINING ALGORITHM

Association rule mining is a specific application of Market Basket Analysis, where retail transaction baskets are analyzed to find the products which are likely to be purchased together. Consider a supermarket setting where the database records items purchased by a customer at a single time as a transaction. The planning department may be interested in finding “associations” between sets of items with some minimum specified confidence. Such associations might be helpful in designing promotions and discounts or shelf organization and store layout. The output of the analysis forms the input for recommendation or marketing strategies.

The most common step in all association rule mining algorithms is to segment the job into two sub-tasks.

- Frequent itemset generation: This sub-task is to discover all the itemsets that satisfy the minsup threshold. The itemsets which satisfy minsup threshold are called frequent itemsets.

- Rule generation: The sub-task extracts all the high confidence rules from the frequent itemsets obtained in above step. These rules are called as strong rules.
Zhixin et al., [135] recommended an improved classification technique based on Predictive Association Rules. Classification Dependent Predictive Association Rules (CPAR) is one of the types of association classification method which integrates the benefits of associative classification and conventional rule-based classification. For generation of the rule, CPAR is more efficient than the conventional rule-based classification, since most of the replicate calculation is ignored and multiple literals can be selected to create multiple rules at the same time. Although the benefit mentioned above avoids the replicate calculation in rule generation, the prediction processes have the disadvantage in class rule distribution inconsistency and interruption of inaccurate class rules. Further, it is ineffective in instances that satisfy no rules. To avoid these difficulties, the author recommends Class Weighting Adjustment, Center Vector-based Pre-classification and Post-processing with Support Vector Machine (SVM).

Wang et al., [124] suggested a novel rule weighting approach in Classification Association Rule Mining. Classification Association Rule Mining (CARM) is the newest classification rule mining technique that built an association rule mining based classifier by using Classification Association Rules (CARs). The specific CARM algorithm which is used is not regarded, a similar set of CARs is continually produced from the data, and a classifier is commonly presented as a structured CAR list, depending on a selected rule ordering approach. Several number of rule ordering approaches have been recognized in the recent past, which can be categorized as rule weighting, support-confidence and hybrid. In this approach, an alternative rule-weighting method, called CISRW (Class-Item Score based Rule Weighting) and a rule-weighting based rule which orders mechanism depending on CISRW. Later on, two hybrid techniques are added and developed by merging support-confidence and CISRW.

Bartik [112] presented association based classification for relational data and its use in web mining. Classification according to the mining association rules is a better and human understandable classification scheme. The intention of the author is to force an alteration of the fundamental association based classification technique that can be used in gathering data from the Web pages. The alteration of the technique and necessary discretization of numeric characteristics are given.
Mining Interesting Rules by Association and Classification Algorithm is put forth by Yanthy et al., [114]. The main purpose in data mining is to reveal hidden knowledge from the data and numerous techniques have been suggested so far. However, the disadvantage is that only a few portions of the created rules would be of interest to some provided user. Thus, several measures like confidence, support, lift, information and gain, have been suggested to discover the best or highly interesting rules. However, some techniques are great at creating rules high in one interestingness measure but not great in other interestingness measures. The connection between the techniques and interestingness measures of the created rules is not obvious right now. The author studied the connection between the techniques and the interesting measures. The author used synthetic data so that the result is not limited to particular situations.

Zongyao et al., [136] proposed a mining local association pattern from spatial dataset. The author offers a model and an algorithm to mine local association rules from the existing spatial dataset, while totally considering the reality that spatial heterogeneity may extensively be presented in realism. The significant element of the model is the computation Localized Measure of Association Strength (LMAS) which is used to measure the local association patterns. Spatial association relations are completely described as spatial relations that are modeled by DE-9IM model. The author offers a mining technique for deciding local association patterns taken from the spatial dataset. The mines technique reference and target objects have probable association patterns and processes LMAS for every object in the reference objects for some interested spatial relation. Therefore, the effect of the algorithm is a LMAS distribution map that repeats association potential variations inside the examination area. Spatial interpolation for LMAS is suggested to produce a continuous LMAS distribution that can be used to examine hot spots that show strong association patterns. This technique was applied in an ecological system research.

Yong et al., [129] proposed a mining association rule with a new measure criteria. In these days, association rules mining from large databases is an active field of research of data mining followed by many application areas. On the other hand, there are some difficulties in the strong association rules mining, depending on the support-confidence framework. Firstly, there are a large number of redundant association rules generated, it
is then complicated for the user to find the interesting ones. The correlation along with the features of the specified application areas is then avoided. Thus, innovative measure criteria called Chi-Square test and cover must be initiated to association rules mining, and the additional main aspect is the use of Chi-Square test to decrease the amount of rules. The Chi-Square test and cover of measures are used by the author for association rules mining for the use of removing the itemsets that are statistic free, as frequent itemsets or rules are created. Thus, the number of patterns of itemsets are reduced and it is effortless for the user to meet the highly noticeable association rules. Finally, the Chi-Square test is effective on diminishing the quantity of patterns through merging support and cover constrain. According to Chi-Square test, little irrelevant attributes can be eliminated and the efficiency and reality of mining association rules enhanced.

Mining traditional association rules using frequent itemsets lattice is given by Vo et al., [11]. Numerous methods have been calculated for the improvement of time in mining frequent itemsets. But, the methods which treat with the time of mining association rules are not placed in high research. Authentically, under the database which contains many frequent itemsets (from ten thousands up to millions), the time of mining association rules is much greater than that required for mining frequent itemsets. An application is generated for lattice in mining conventional association rules which will considerably decrease the time for the mining rules. This technique consists of two stages (1) construction of frequent itemsets lattice and (2) mining association rules from lattice. For the fast determination of association rules, the parent-child relationships in lattice is utilized.

Rastogi et al., [88] presents mining optimized association rules with categorical and numeric attributes. Of late, mining association rules on huge data sets have achieved significant attention. Association rules are supportive for expecting relationship along with the features of a relation and include applications in marketing and many retail sectors. Besides, optimized association rules are a good approach to give attention on the most interesting features connecting certain attributes. Optimized association rules are allowable to contain un instantiated attributes and it is complex to discover instantiations where each of the support or confidence of the rule is maximized. In this approach, the complexity of the optimized association rules is simplified in three ways namely,
(a) association rules are allowed to include disjunctions over un instantiated features
(b) association rules are allowed to include a random number of uninstantiated features, and
(c) uninstantiated features can be either categorical or numeric. The general association rules let to mine supportive information on seasonal and local patterns connecting multiple features. It also propose a good method for pruning the search space when formulating optimized association rules for both categorical and numeric features.

Wang et al., [115] performs an investigation on Association Rules Mining based-on ontology in e-commerce. Commercial actions carried out with the use of Internet are very popular. Plenty of transaction logs are produced, which helps to collect useful information through data mining. Hence, Association Rule Mining is important in e-commerce. However, there are several problems that arise in the existing association rules mining systems. The existing conventional techniques cannot solve these problems. The target of answering these difficulties better, it suggests association rules mining depending on ontology. There are three parts during data mining:

1. methods of ontology creation and principles of commodity categorization;
2. simplifying R-interesting based on actual situations;
3. implementing association rules mining depending on ontology by improved Apriori. In addition, it tests the enhanced algorithm using FoodMart2000, Java as the development language and Jena as the ontology engine, completes the whole process of mining, and verifies the validity of the algorithm by the example of the database.

Dong Liyan et al., [29] proposed a novel method of mining frequent item sets. The target of mining association rules is to decide the association relationship along with the item sets from mass data. In a number of practical applications, its responsibility is mostly to support decision-making. The author proposed an association rule algorithm of mining frequent item sets, which introduces a new data structure and takes compressed storage tree to increase the run presentation of this algorithm. In comparison with the existing algorithms, the proposed algorithm has a lot benefits in load balance and run time.
Lei Wen et al., [65] developed an efficient algorithm for mining frequent closed itemset. Association Rule Mining is a prominent field of data mining analysis. Identifying the useful and significant frequent itemset is a key step. The existed frequent itemset discovery algorithms could discover all the frequent itemset or maximal frequent itemset. Pasquier et al., [78] proposed a novel method of mining frequent closed itemset. The size of frequent closed itemset was much lesser than all the frequent itemsets and no information was lost. A new frequent closed itemset method is proposed using the directed specified itemset graph. This method can identify all the frequent closed itemset significantly through depth first search technique.

Mining frequent itemsets from secondary memory was put forth by Grahne et al., [32]. For the main memory databases It is the main work of mining association rules (i.e.) mining frequent itemsets for the main memory databases. The author reveals techniques for mining frequent itemsets when the database or the data structures used in the mining are bulky to apply in the main memory. Therefore, this technique decreases the required disk used by order of magnitude, and lets actual scalable data mining.

Xuegang Hu et al., [122] suggested mining frequent itemsets using a pruned concept lattice. A critical step in association rule mining is removing frequent itemsets. However, most of the approaches which extract frequent itemsets examine databases numerous times, which reduces the effectiveness. In this method, the association among the concept lattice and frequent itemsets is used, and Pruned Concept Lattice (PCL) method is found to illustrate frequent itemsets in a specified database, and the scale of frequent itemsets is compacted successfully. A technique for removing frequent itemsets based on the PCL is executed, which prunes rare concepts suitably and dynamically throughout the PCL's construction based on the Apriori property.

Mining intratransaction associations are the preceding studies on mining association rules, that is, the associations between items inside the same transaction. It extends the scope to include multidimensional and inter transaction associations. For intertransaction association "if (company) A’s stock goes up on day one, B’s stock will go down on day two but go up on day four:" , where the company or day is treated as unit of transactions, the items belong to different transactions in the stock price
information database. Furthermore, the intertransaction association can be expanded to associate multiple properties in the similar rule, so that multidimensional intertransaction associations can be defined and discovered. Mining intertransaction associations facade more challenges on capable processing than mining intratransaction associations, since the number of potential association rules is very large. Tung et al., [6] introduced the notion of intertransaction association rule and developed an efficient algorithm, FITI (First Intra Then Inter), for mining intertransaction associations, which adopts two major ideas:

1) an intertransaction frequent itemset contains only the frequent itemsets of its corresponding intratransaction counterpart; and

2) a special data structure is built among intratransaction frequent itemsets for efficient mining of intertransaction frequent itemsets.

The question of mining association rules among items in a great database of sales transactions is examined. In mining association rules, a database of sales transactions is provided, to find all the associations among items, that is, the presence of some items in a transaction will involve the presence of other items in the same transaction. The mining of association rules can be drawn into the problem of finding large itemsets where a large itemset is a set of items that emerge in an enough number of transactions. The problem of discovering large itemsets can be answered by constructing a candidate set of itemsets first, and then, identifying, inside the candidate set, the itemsets that gather the large itemset requirement. In general, this is made iteratively for every large k-itemset in mounting order of k, where a large k-itemset is a large itemset with k items. To decide large itemsets from a large number of candidate sets in early iterations is generally the ruling factor for the overall data mining presentation. To address this issue, Jong Soo Park et al., [53] developed an effective algorithm for the candidate set generation. Effective algorithm is a hash-based algorithm which is mainly good for the invention of a candidate set for large 2-itemsets. Explicitly, the number of candidate 2-itemsets invented by the proposed algorithm is, in orders of magnitude, lesser than that by preceding methods, therefore solving the performance bottleneck. The invention of smaller candidate sets enable to trim the transaction database size efficiently at a much earlier stage of the iterations, thereby reducing the computational cost considerably for later
iterations. The benefit of the new algorithm gives the chance of decreasing the amount of disk input/output required.

Association rule mining is a key issue in data mining. On the other hand, the classical models ignore the difference among the transactions, and the weighted association rule mining doesn’t work on databases with only binary attributes. Sun et al., [55] introduced a new measure w-support, which does not require preassigned weights. It gets the quality of transactions into concern using link-based models.

A novel problem of mining common temporal association rules is investigated in publication databases. Essentially, a publication database is a group of transactions where every transaction T is a set of items of which every item includes an individual exhibition period. This model of association rule mining is unable to hold the publication database owing to the following fundamental problems, namely,

1) lack of concern of the exhibition period of every individual item and

2) lack of an equitable support counting basis for every item. To remedy this, Chang-Hung Lee et al., [20] proposed an innovative algorithm Progressive-Partition-Miner (abbreviated as PPM) to discover general temporal association rules in a publication database. The principle idea of PPM is to primarily divide the publication database in the light of exhibition period of items and increasingly build up the incidence count of every candidate 2-itemset, based on the inherent partitioning characteristics. Algorithm PPM is planned to utilize a filtering threshold in every partition to clip out those increasingly infrequent 2-itemsets. The characteristic that the number of candidate 2-itemsets produced by PPM is very near to the number of frequent 2-itemsets permit to employ the scan reduction technique to efficiently decrease the number of database scans. Clearly, the execution time of PPM is, in order of magnitude, lesser than the other required competitive schemes that are straightly expanded from the existing methods. The accuracy of the PPM is verified and some of the theoretical properties are derived. Sensitivity analysis of numerous parameters is conducted to give many insights into Algorithm PPM.

A top-down progressive deepening method is developed by Jiawei Han et al., [51] for efficient mining of multiple-level association rules from large transaction
databases based on the Apriori principle. A collection of alternate algorithms is suggested by sharing intermediate results, and the relative presentation is tested and analyzed. The enforcement of different measures of interest to discover more interesting rules, and the leisure of rule conditions for finding “level-crossing” association rules, are investigated. Thus the effective algorithms can be increased from large databases for the discovery and strong multiple-level association rules.

Association rule mining is a lively data mining research area. But, majority of the ARM algorithms provide central surroundings. In contrast to previous ARM algorithms, Ashrafi et al., [76] have developed a distributed algorithm, called optimized distributed association mining, for geographically distributed data sets. ODAM produces support counts of candidate itemsets which is faster than the other DARM algorithms and decreases the size of average transactions, data sets, and message exchanges.

At different levels of support, the occurrence of interesting patterns is frequent. The classic association mining derived from uniform minimum support, such as Apriori, moreover misses interesting patterns of low support or undergoes the bottleneck of itemset invention caused by a low minimum support. A recovered key lies in exploiting support constraints, which identifies the minimum support that is needed for end itemset, so that only the necessary itemsets are produced. Ke Wang et al., [59] presented a framework of frequent itemset mining in the presence of support constraints. This approach is to "push" support constraints into the Apriori itemset invention so that the "best" minimal support is decided for each itemset at runtime to protect the essence of Apriori. This approach is called Adaptive Apriori.

Various sequential algorithms have been planned for the mining of association rules. However, only a few works can be completed in mining association rules in distributed databases. A straight application of sequential algorithms to distributed databases is not efficient, as it needs a huge amount of communication overhead. An efficient algorithm called DMA (Distributed Mining of Association rules), is proposed by Cheung et al., [27]. It produces a small number of candidate sets and needs only O (n) messages for support-count exchange for every candidate set, where n is the number of
sites in the distributed database. The algorithm has been executed on an experimental testbed, and its performance studied.

This section deals with the association rule, used in mining algorithm and the survey done by various researchers on this algorithm. Association rule mining has been extensively studied in data mining and hence it is sequential pattern mining. Association rule mining assists several businesses to make certain verdicts, like catalogue design, cross marketing and customer shopping behavior analysis. Based on this survey, it is found that association rule algorithms are capable of generating rules with confidence values of less than one, however, the number of possible association rules for a given dataset is normally very large and a high proportion of the rules are generally of little value.

2.2. REVIEW OF APRIORI ALGORITHM

As discussed in chapter 1, apriori algorithm also follows two steps. They are,

- Finding the frequent itemsets: The sets of items that have minimum support.
  - It is to be noted that a subset of a frequent itemset must also be a frequent itemset.
- Use the frequent itemsets to formulate association rules.

Apriori Algorithm is designed to operate on databases containing transactions. Other algorithms are designed for finding association rules in data having no transactions or having no timestamps. The purpose of using the Apriori Algorithm is to find the associations between different sets of data. It is sometimes referred to as "Market Basket Analysis". Each set of data has a number of items and is called a transaction. The output of Apriori is the set of rules that reveal how often items are contained in sets of data. It describes innovative ways to find association rules on a larger scale, allowing implication outcomes that consist of more than one item. The items that occur together frequently can be associated to each other in one combination. These types of items occurring together form a frequent itemset and these frequent itemsets form the association rules. This Apriori Algorithm is used in the data mining process.
Guo Yi-ming et al., [39] presented a vertical format algorithm for mining frequent item sets. Apriori is a traditional algorithm for association rules. For the purpose of giving the support degree of candidate sets, Apriori needs to scan the database numerous times. This author proposed a new algorithm, where mining frequent item sets happen via the vertical format. The proposed technique needs to scan database one at a time. In the follow-up data mining procedure, it is used to acquire new frequent item sets through `and operation' among item sets. This technique takes less storage space, and can improve the effectiveness of data mining.

Sumithra et al., [90] proposed a distributed Apriori association rule and classical Apriori Mining Algorithms for grid based knowledge discovery. The intention is to gain knowledge by using predictive. Apriori and distributed Grid Dependent Apriori algorithms for association rule mining. The author provides the execution of an association rules discovery data mining task by using grid technologies. An effect of execution with a contrast of existing Apriori and Distributed Apriori is also provided by the author. Distributed data mining systems provide a good utilization of multiple processors and databases to fasten the execution of data mining and to facilitate data distribution. For evaluating the efficiency of the described technique, performance investigation of Apriori and Predictive Apriori techniques on a standard database (provides by using weka tool). The key intention of grid computing is to provide the organizations and application builders the ability to develop distributed computing environments in which it uses computing resources on demand. Hence, it can assist to amplify the effectiveness and to decrease the cost of computing networks by reducing the time for data processing and optimizing the resources and distributing workloads, and finally permit the users to gain a much faster outcome on large operations at lesser costs.

Mining association rules based on Apriori Algorithm and application is given by Pei et al., [80]. Mining association rules is a significant subject in data mining. Intended at two difficulties of discovering frequent itemsets in a large database and mining association rules from frequent itemsets, the author carries some analysis on mining frequent itemsets algorithm with the help of Apriori Algorithm and Mining Association Rules Algorithm with the help of enhanced measure system. Mining Association Rules technique is enhanced with the help of support, confidence and interestingness. It aims at
developing interestingness ineffective rules and losing helpful rules. Ineffective rules are
cancelled, which creates many reasonable association rules and includes negative items.
The suggested technique is utilized to mine association rules to the 2002 student score list
of computer dedicated field in Inner Mongolia University of science and technology.

Omari et al., [79] developed a new temporal measure for interesting frequent
itemset mining. Frequent itemset mining helps in searching for powerfully associated
items and transactions in large transaction databases. This measure is based on the fact
that interesting frequent itemsets are typically covered by several recent transactions.
This minimizes the cost of searching for frequent itemsets by minimizing the search
interval. Additionally, this measure can be used to enhance the search approach
implemented by the Apriori Algorithm.

Qiang et al., [82] presented an association classification method based on the
compactness of rules. Associative classification provides maximum classification
correctness and strong flexibility. Simultaneously, this associative classification
undergoes a over fitting because the classification rules satisfied the least support and
lowest confidence are returned as strong association rules return to the classifier. An
innovative association classification technique is based on the presentation of rules, it
extends Apriori Algorithm which considers the interestingness, importance and
overlapping relationship among rules. Experimental observation proves that the proposed
approach has better classification accuracy in comparison with CBA and CMAR.

Rui Chang et al., [93] proposed a new optimization algorithm called APRIORI-
IMPROVE based on the inadequacy of Apriori. APRIORI-IMPROVE algorithm offers
optimizations on 2-items generation, transactions compression and others. APRIORI-
IMPROVE uses hash structure to generate L2, uses an efficient horizontal data representation
and an optimized strategy of storage to save time and space. The performance study
shows that APRIORI-IMPROVE is much faster than Apriori.

Association rule mining is used to recognize association relationships among
large data sets. In association rule mining, mining frequent patterns is an important
feature. An efficient algorithm named Apriori-Growth based on Apriori Algorithm and
the FP-tree structure is presented by Bo Wu et al., [15] to mine frequent patterns.
The benefit of the Apriori-Growth Algorithm is that it need not produce conditional pattern bases and sub-conditional pattern tree recursively. Simulation results reveal that the Apriori-Growth Algorithm executes faster than Apriori Algorithm, which is almost as fast as FP-growth, but it requires only a smaller memory.

Rough Set Theory and the Association Rules Algorithm are mining methods in which it is used to recognize implicit rules from great amounts of data. As the Association Rules Mining Algorithm, Apriori Algorithm has achieved a lot of application owing to its easy use. On the other hand, in practice, it often comes across problems such as low mining efficiency. Hence, many invalid rules are attained and the rules of pattern mining disorder. An algorithm called R_Apriori is created by Chen Chu-xiang et al., [22] for the problems with the decision-making domain. Initially, the conditions of the cores are mined with the Rough Attribute Reduction Algorithm, 1-frequent item sets and the corresponding sample set is then found with use mining cores set by the Apriori Algorithm. After the above mentioned stage, the multi-stage frequent item sets and the corresponding support and confidence can be gained by the sample collection intersection operator. According to the degree of confidence and support, the corresponding strength of the rule is determined. R_Apriori Algorithm resolves the problems of Apriori Algorithm to recover the effectiveness of the algorithm and is in promotion on certain significance.

Retail industry builds up a large number of retail sales data. Apriori algorithm helps to recognize the association rules among commodities and institute cross-selling strategies to enhance the profits of the retail industry. Based on the analysis of the effectiveness of the typical Apriori Algorithm, Changsheng Zhang et al., [21] provides a modified method to enhance the performance of the Apriori Algorithm by deducting the scale of the candidate item set Ck and the spending of input/output. It is also described as the application of the modified Apriori Algorithm in search of the association rules of sales data of commodities by combining with the actual sales data, so that the possibility of the algorithm is proved.

Association rules are the main technique for data mining. Apriori Algorithm is a standard algorithm of association rule mining. Many algorithms for mining association rules and their changes are proposed by Wanjun Yu et al., [116] on the basis
of Apriori Algorithm, but traditional algorithms are not good. For the two bottlenecks of frequent itemsets mining namely, the great multitude of candidate 2-itemsets and the poor effectiveness of increasing their support, it proposes a new algorithm called Reduced Apriori Algorithm with Tag (RAAT), which decreases one redundant pruning operations of $C_2$. If the number of frequent 1-itemsets is $n$, the number of connected candidate 2-itemsets is then $C_n^2$, while the pruning operations is $C_n^2$. The new algorithm reduces the pruning operations of candidate 2-itemsets, which saves time and increases effectiveness. For the problem of poor effectiveness of increasing support, RAAT optimizes the subset operation, during the transaction tag to accelerate support calculations. The experimental results obtained from the tests show that RAAT outperforms only original efficiency.

Efficiency in the research of Association Rule Mining has been disturbed for numerous years. A high-dimension oriented Apriori Algorithm is proposed by Lei Ji et al., [64]. Unlike the existing Apriori improvements, the algorithm takes on a new method to decrease the redundant generation of sub-itemsets during pruning the candidate itemsets, which can acquire higher efficiency of mining than that of the original algorithm when the dimension of data is high. Theoretical proof and analysis are given for the rationality of this algorithm.

It is a difficult job to set rare association rules to hold unpredictable items as approaches, for example, Apriori Algorithm and frequent pattern-growth, and a single minimum support application suffers from low or high minimum support. If the least support is set high to cover the rarely appearing items, it will ignore the frequent patterns involving rare items since rare items fail to satisfy high minimum support. Hence, an effort is made to remove rare association rules with multiple minimum supports. It explores the possibility and proposes a multiple minsup based apriori-like approach called Probability Apriori Multiple Minimum Support (PAMMS) to efficiently discover rare association rules developed by Rawat et al., [100].

An algorithm is established by Gang Fang et al., [37] in mining spatial topology association rules dependent on Apriori, which is used in mining spatial multilayer transverse association rules from spatial database. This algorithm generates candidate
frequent topological itemsets by means of down-top search strategy as Apriori, which is appropriate for mining short spatial topological frequent itemsets. This algorithm compresses a type of spatial topological relation to form a digit. By this method, initially, the algorithm may proficiently reduce some storage space when making mining database. Secondly, the algorithm is simple to compute topological relation among spatial objects, that is, it may compute support of candidate itemsets fast. Finally, the algorithm is fast to attach (k+1) candidate itemsets of k-frequent itemsets as down-top search strategy. The result of the experiment indicates that the algorithm of mining spatial topology association rules based on Apriori is capable of extracting spatial multilayer transverse association rules from spatial database using efficient data store, and it is very good to extract short frequent topology association rules.

Apriori is one of the most important algorithms used in rule association mining given by Dongme Sun et al., [30]. The limitations of the Apriori Algorithm are conversed and then an enhancement planned for improving its effectiveness. The improved algorithm is based on the mixture of forward scan and reverse scan of a given database. If certain conditions are suited, the improved algorithm can highly decrease the scanning times required for the discovery of the candidate itemsets. Theoretical proof and analysis are given for the rationality of the algorithm. A simulation instance is given to compare the advantages of this algorithm with that of Apriori.

Cutting data mining is an important technique to increase efficiency, discover hidden knowledge in cutting database, and give guidance for cutting decisions. It analyzes the Apriori Algorithm for association rules mining, and creates some enhancement for this algorithm based on the features of cutting database given by Guofeng Wang et al., [40]. Apriori Algorithm is enhanced to mine association rules in cutting database. Thus, the Apriori Algorithm can be used well in cutting data mining and the enhanced algorithm can achieve better effect than the traditional algorithm.

For improving the effectiveness of excavation in relational database with multidimensional association, rule is given by Yongge Shi et al., [131]. It analyzed Apriori Algorithm and BUC algorithm based on practice. An enhanced Apriori Algorithm-DGP Algorithm which is based on the multidimensional association rule was then offered,
where the more efficient one will be used in the relational database. Finally, it was applied for analyzing the reasons for users' line which do not reach the standard. It can effectively improve the speed of Data mining and improve ADSL line quality’s analyzing and solving capabilities.

The Apriori Algorithm is the main influential apriori for mining association rules. The basic idea of Apriori Algorithm is to recognize all the frequent sets. Through the frequent sets and derived association rules, these rules should undergo minimum support threshold and minimum confidence threshold. Libing Wu et al., [68] presented improved algorithms, mostly through the introduction of interest items and frequency threshold, to improve the mining effectiveness, dynamic data mining to make it easy for the users.

Based on the customer relationship management system of ShanHua Carpet Company, Peng Gong et al., [81] established an enhanced data mining association rules Apriori Algorithm and it’s useful to Shanhua group Cross-selling analysis. The use of the Apriori Algorithm removes lots of invalid businesses, decreases the records for the following scanning, which increases the effectiveness of data mining. Simultaneously, with the deduction of the business, the scale of database will also decrease. The scanning time will therefore be saved and the effectiveness of processing will be enhanced.

Among the data mining algorithms, Apriori Algorithm is a classical algorithm of association rules. Tang Junfang et al., [110] proposed an improved algorithm based on classical Apriori Algorithm analyzing. Through compressing transaction database, the improved algorithm prove the same number of records by including an attribute named count, and apply count to count the support of itemsets to raise the efficiency and practical experience.

Gang Fang et al., [36] proposed an algorithm of mining spatial topology association rules with constraint condition based on Apriori, for mining spatial multilayer transverse association rules with constraint condition from a big spatial database. This algorithm produces candidate frequent topological items set up using search strategy related to Apriori. This is appropriate for mining short spatial topological frequent item sets with constraint condition. This algorithm reduces the storage structure of spatial topological relation to make an integer. Through this method, the algorithm may initially
decrease some storage space of mining database efficiently. Secondly, the algorithm is natural to differentiate the topological relation of two spatial objects. That is, it might calculate the support of candidate item sets fastly. At last, the algorithm is fast to connect the (k+1) candidate item sets of k-frequent item set by up search strategy. The result of the experiment designate that the algorithm of mining spatial topology association rules with constraint condition based on Apriori is capable to extract spatial multilayer transverse association rules with constraint condition from spatial database by means of efficient data store. It is very good to extract short frequent topology association rules with constraint condition.

Because of the exponential growth in global information, companies have to contract with an ever growing amount of digital information. One of the most important challenges for data mining is quickly and correctly finding the relationship between the data. The Apriori Algorithm is the most well-liked technique in association rules mining. When applying this method, a database has to be scanned many times and many candidate itemsets are generated by Kun-Ming Yu et al., [60]. Parallel computing is an efficient strategy for speeding up the mining process. The Weighted Distributed Parallel Apriori Algorithm (WDPA) is offered as a key to this problem. In the planned method, metadata are stored in TID forms, so only one scan to the database is required. The TID counts are also considered, and better load-balancing as well as minimizing idle time for processors can therefore be achieved.

The Apriori Algorithm is most powerful in excavating association rules. The basic idea of the algorithm is to identify all the frequent itemsets to get association rule. Libing Wu et al., [67] presented the improved Apriori Algorithm based on interested items, which mostly built an ordered interested table and crossed it to excavate frequent itemsets quickly. The paper also achieves the improved algorithm by writing c# code. It has been confirmed through experiments that this algorithm consumes less time than the traditional ones.

Finding frequent itemsets is another main investigated field of data mining. The Apriori Algorithm is the most recognized algorithm for Frequent Itemsets Mining (FIM) given by Yanbin Ye et al., [126]. Numerous implementations of the Apriori
Algorithm have been documented and calculated. One of the implementations of the Apriori Algorithm optimizes the data structure with a trie by Bodon catches the attention. The outcomes of the Bodon's implementation for discovering frequent itemsets show to be faster than the others by Borgelt and Goethals. Bodon's implementation reworks into a parallel one where the input transactions are read by a parallel computer. The parallel computer effected on the modified implementation is presented.

First, evaluation indexes of association rule are provided. Their meaning is analyzed. The Apriori arithmetic about knowledge discovery is introduced by Shihai Zhang et al., [104] and its defects are analyzed. Second, a knowledge discovery method of improved Apriori-based high-rise intelligent form selection is established. Finally, examples of this method are presented. This method provides a new approach to mining knowledge information in engineering cases, guiding structural form selection design and to improve quality, efficiency and intelligence level of structural design.

Sen Guo et al., [103] presented a mechanism called R_Apriori for learning rules from large datasets. The existing rough set based methods are not valid for large data sets owing to its high time and space complexity. Large data sets are separated into numerous parts, in combination with Apriori Algorithm. Implicated rules are obtained in liner relation to the size of the data set. The experimental result shows that this method is better than the existing ones.

Apriori Algorithm is one of the classic and best algorithm for learning association rule and its process. Mining association rules is based on Apriori Algorithm and application. In the data mining research, mining association rules is significant and it can be used effectively. This section deals with Apriori Mining Algorithms and the survey done by various researchers on this algorithm. It is clear from the review that the major shortcoming of association rules data-mining is that the support-confidence framework often generates too many rules. Although apriori algorithm can identify meaningful itemsets and construct association rules, it suffers from the disadvantage of generating numerous candidate itemsets that must be repeatedly compared with the entire database.
2.3. IMPLEMENTATION OF MARKET BASKET ANALYSIS SYSTEM BY VARIOUS AUTHORS

Data mining has become a widely accepted process for organizations to enhance their organizational performance and gain a competitive advantage. As the data mining process is a relatively new concept, it has been defined in various ways by different authors. Data mining allows managers to make more knowledgeable decisions by predicting the future trends and behaviors. One of the most widely used areas of data mining for the retail industry is marketing. Market basket analysis is a marketing method used by many retailers to determine the optimal locations to promote products. The market basket is defined as an item set bought together by a customer on a single visit to a store. The market basket analysis is a powerful tool for the implementation of cross-selling strategies.

Market Basket Analysis is a modeling technique based upon the theory that if any customer buys a certain group of items, they are then more (or less) likely to buy another group of items. For example, if any customer is in an English pub and they buy a pint of beer and don't buy a bar meal, they are more likely to buy crisps at the same time than somebody who didn't buy beer. The set of items a customer buys is referred to as an itemset, and Market Basket Analysis seeks to find the relationships between purchases.

Trnka [3] uses Data Mining Methods for Market Basket Analysis. The technique for the implementation of Market Basket Analysis to Six Sigma is provided. Data mining techniques provide more scenarios in the market sector. Basket Market Analysis is one among them. Six Sigma techniques use numerous statistical techniques. With the implementation of Market Basket Analysis to Six Sigma, the results can be improved and Sigma performance rank of the method can also be customized. The author used GRI (General Rule Induction) technique to build association rules along with the products in the market basket. These associations offer variety among the products. A web plot is used to represent the dependence among the products.

Market Basket Analysis Based on Text Segmentation and Association Rule Mining is suggested by Xie et al., [121]. Market Basket Analysis is very helpful in providing scientific decision support for trade market by mining association rules among
the items people buy collectively. The author presents an innovative Market Basket Analysis technique by mining association rules on the items' internal features that are attained with the help of automatic words segmentation technique. This technique has been used for dynamic dishes recommend system.

Market-Basket Analysis is a procedure to examine the habits of buyers to find the relationship among different items in their market basket. The discovery of these relationships can help the merchant to devise a sales strategy by considering the items often purchased mutually by customers. In this research, data mining with market basket analysis method is implemented by Setiabudi et al., [26], where it can analyze the buying habit of the customers. The testing is conducted in Minimarket X. Searching for frequent itemsets executed by Apriori algorithm to acquire the items that often show in the database and the pair of items in single transaction. Pair of items that go beyond the minimum support will be included the frequent itemsets that are chosen. Frequent itemsets that go beyond the minimum support will produce association rules after decoding. Single frequent itemsets can produce association rules and find the assurance, which uses a hybrid-dimension association rules. This information provides more consideration for the owners of Minimarket X to create the further decision.

Market-basket analysis is a familiar business problem, which can be partially solved computationally utilizing association rules, mined from the transaction data to increase cross-selling effects. Here, Chiu et al., [58] models the Market-Basket Analysis as a finite mixture density of human consumption behavior according to social and cultural events. This guides the use of principle component analysis and probably mixture density analysis of transaction data, which was not obvious before. Comparison of PCA and association rules is mined from a set of benchmark transaction data, to investigate similarities and differences among these two data exploration tools.

The capability to make groups has long been one of the features of interest for the users across the social networks. Groups permit users (account holders) to connect them and be a fan-follower of them. Additionally, there have been conditions for the account holder to also be a part of multiple groups as a member. The primary goal of Bharath et al., [18] is to analyze patterns within group membership. A model is based
on Market Basket Analysis in excess of a subset of the social network groups and then derives a set of associative mining rules among these networks. The information mined may give interesting insights for cross-marketing purposes. For example, using this technique an organization $O_x$ could come across that the buyers of their product $x$ tend to be socially fans or members of groups with the subtype $Y$ or even, possibly, more specifically with another organization $O_y$ that sell a product $y$. Using such data, the organizations may try to influence these findings to build a co-operative (Co-operation + Competition), profitable synergy and might also get insights to decide which products of $O_y$ may be impacted if the store discontinues selling the product $x$ or vice-versa. The aim is to take in a new high-level dimension to the social network analysis area by swallowing the concepts of Market Basket and then try to provide a basic possibility criterion for the whole approach by simple simulations.

Cavique et al., [61] addresses the problem of finding the next-item for each customer in large database marketing and can be seen as an extension of the Market Basket Analysis. Majority of the existing software uses Apriori-like algorithms. The outputs of the Apriori Algorithms are simple to know and many novel patterns can be recognized. But, the sheer number of association rules may make the understanding of the results difficult. The aim of the work is to mechanize the cross-selling strategy. To make the work of the marketer simple, it avoids the analysis of thousands of rules in associating customers with their next-item.

“Market Basket Analysis” algorithms are being used worldwide in analyzing consumer purchasing patterns - specially, in detecting products that are often purchased. Cunningham et al., [99] applied the a-priori market basket tool to the task of detecting subject classification categories that co-occur in the transaction records of books borrowed from a university library. This information can be helpful in directing users to extra portions of the collection that may include documents which are important to their information need, and in deciding a library's physical layout.

Through the process of mining frequent item sets, when minimum support is small, the production of candidate sets is a time-consuming and frequent operation in the mining algorithm. The FP growth algorithm does not require the candidate sets to be
produced, the database which presents the frequent item set is compressed to a frequent pattern tree (or FP tree), and frequent item set is mining by using FP tree. For the sake of researching Market Basket Analysis, the frequent-pattern is introduced by Yongmei Liu et al., [132] Visual C++ is applied to design the program to mine the frequent item sets. In the sight of the frequent K-item set, the goods which sell probably at the same time in the supermarket are set in the same place.

Market Basket Analysis is the most important function of knowledge discovery in databases. Real life market basket databases generally contain temporal coherences, which couldn’t capture via standard association rule mining. Therefore, there is a need for developing algorithms that show such temporal coherences inside the data. Schluter et al., [108] gathers several notions of temporal association rules and presents an approach for mining most of these kinds (cyclic, lifespan- and calendar-based) in a market basket database, enhanced by two novel tree structures. These two tree structures are called as EP- and ET-Tree, which are obtained from accessible approaches improving standard association rule mining. They are used as representation of the database and hence creates the discovery of temporal association rules which are very efficient.

Market Basket Analysis is the main data mining application for finding correlation among buying items in transactional databases. Preceding works reveal that considering constraints which users may be concerned with into the mining process, can efficiently decrease the number of patterns and acquire more promising information.

Ya-Han Hu et al., [125] extend the RFM analysis into the mining process to measure the importance of frequent patterns. In RFM analysis, a customer to be known as important if his/her purchases records are new, frequent, and having high amount of money. First, the RFM-patterns is defined by following the RFM analysis concept. This discovered RFM-patterns have occurred frequently, not only often occurred but also newly bought and have a higher percentage of income. A tree structure is then planned, called RFMP-tree, to reduce and store up whole transactional database, and a pattern growth-based algorithm, called RFMP-growth, is developed to find out all RFM-patterns from RFMP-tree.
A market basket is a group of products that outline a single retail transaction. This purchase data of products can shed significant light on how product(s) might persuade sales of other product(s). Leaving from the average approach of frequent itemset mining, it imagines that purchase data can be formed as a social network. One can discover communities of products that are bought mutually, which can guide to expressive exploration and discovery of a larger influence zone of product(s). Raeder et al., [107] developed a novel utility measure for communities of products and show, both financially and intuitively, that community detection provides a useful complement to association rules for Market Basket Analysis. All the conclusions are calculated on real store data.

To make a marketing campaign efficient, the business must examine the spending behavior of the customers and identify the best promotional time period. However, extracting these data physically is time consuming. Thanks to the current internet and wireless technologies, collectively with the existing data mining tools this system is able to mechanize the data mining process efficiently with instant business intelligence constructed into it. Chan Gaik Yee et al., [19] introduced an Enterprise Marketing Campaign Automation (EMCA) system that can provide data for businesses to instantly assemble them for determining effective and accurate marketing campaign strategy. During Market Basket Analysis, the businesses and customers together are good which have a recommended list of "most likely to purchase" products being produced directly. By developing a mailing list targeted to a particular group of buyers with reference to their buying habits, can decrease the marketing cost by just mailing the promotional items to the particular group of buyers. Being able to predict the low and peak sales seasons will allow the business to plan ahead with good marketing campaign strategies for consumers to have a more pleasant shopping experience.

Many commercial companies gather large quantities of data from every day operations. For instance, customer orders or purchase data are gathered every day at the counters of grocery stores. Data mining is applied on such kind of data to extract patterns that can be helpful to learn the purchasing behavior of the customers. Such information is used to maintain a variety of business related tasks. For example, the investment of that kind of information is used in building a Website for a grocery store. Association rule
mining is one of the techniques used to extract databases. Association rule mining is the discovery of association rules showing attribute values that often happen together. Omari et al., [1] have discovered association rules from a grocery store dataset which represents customer transactions in that grocery store. The discovered association rules have been providing inputs to plan a well-structured Website prototype for that grocery store. Promising results, that could affect the process of Website design, have been found.

Many organizations gather information about customers which are applied to maintain various business related tasks. The feature of the customers and their behavior is stored in the database. Thivakaran et al., [109] proposed a method of using incremental updating technique to mine direct association rules Inter & Intra transactions. Cluster analysis is presented to confirm that the associated objects fall in nominal cluster. The result of this technique can be utilized to develop a well structured e-shop which makes it easy to the customer by helping them to find what the customer wants in a particular way and also aids them to decide the associated products by the method of prediction. It therefore reduces the workload of marketing professionals in direct marketing and thus provides customer satisfaction.

Market Basket Analysis provides the approach into the merchandise by telling which products are likely to be purchased together and which are most possible to be purchased. The Market Basket Analysis is a dominant tool, specially in retailing it is necessary to discover large baskets, as it deals with thousands of items. FP-growth algorithm is a good algorithm for mining frequent patterns. It need not produce the candidate sets. Is quite time consuming. It scans database only twice and frequent item set is mining by using of FP tree. Liu Yongmei et al., [69] applied Visual C++ to design the program to mine the frequent item sets using FP-Growth Algorithm. According to the mining result, the stock in the supermarket is arranged collectively in the same place compatible for customer.

This section describes the Market Basket Analysis and various researches on it. It is the tool necessary to discover large basket to deal with large items. It reduces the workload of marketing professionals to give direct marketing and thus provide customer satisfaction. It is revealed from the survey that the market basket analysis is only the
initial level. If efficient techniques are followed in market basket analysis, then it is possible to discover several interesting associations and relationships and also there is huge possibility for eradicating the problem of a potentially high volume of trivial results.

2.4. AGRICULTURE AND DATA MINING TECHNIQUES

Agriculture is the major business in India. It contributes 10-15% GDP to the Indian economy. In Indian agriculture, the volume of data is enormous. When the data becomes information, it is highly useful for many purposes. The conventional and traditional system of data analysis in agriculture is purely dependent on statistics. Data mining is a modern data analysis technique. It has a wide range of applications in the field of agriculture. Data mining is the process of extracting vital and useful information from large sets of data (Abello et al., 2002; Klosgen and Zykow, 2002; Pardalos et al., 2007). There are many applications present in agriculture which can be used for data mining technique, such as Prediction of problematic wine fermentations, Detection of diseases from sounds issued by animals and Optimizing pesticide usage by data mining. This section describes how the agricultural data are useful in data mining. The agricultural data contains large information such as pest management, various crop-weather-pest/disease relations which can be used effectively in data mining. It can improve the productivity of the crop.

The problem in dealing with pest management is, insufficient knowledge about the factors manipulating pest population dynamics. To know about pest dynamics, scientists and researchers gather pest surveillance data and correlate agricultural operations about crops, farming practices and weather parameters. These databases include features of pest incidence, climate, soil, agricultural practices and provide repositories of information. Correlations among some of these factors and pest incidence can be known through the statistical models developed, On the other hand, a functionally feasible model for pest forecast and pesticide use is still needed by farmers for efficient and effective pest management. Tripathy et al., [4] described the frame work of the pest management system using data mining techniques. These concentrate on providing historical data, current and recommended pest and pesticide information and to be simulated pest models up to the farm level. In this work, an effort has been made to reveal how geospatial data mining included in agriculture including pest scouting,
pesticide and climatological parameters are helpful for optimization of pesticide usage and better management. The outcomes will show interesting patterns of the dynamics of farmer practices in pesticide usage both in spatial & non spatial way and help understand the reasons for pest and pesticide utilization.

Data driven precision agriculture aspects, mainly the pest/disease management, need a dynamic crop-weather data. A research was performed in a semi-arid region to know the crop-weather-pest/disease relations using wireless sensory and field-level observation data on directly related and interdependent pest (Thrips) - disease (Bud Necrosis) dynamics of groundnut crop. Data mining techniques were used by Tripathy et al., [5] to turn the data into useful information/knowledge/relations/trends and correlation of crop-weather-pest/disease continuum. These dynamics that were acquired from the data mining techniques and trained through mathematical models were authenticated with the corresponding surveillance data.

Association Rule Mining, formerly proposed for the market basket data, has prospective applications in many areas. Spatial data, such as Remote Sensed Imagery (RSI) data, is a promising area of application. Mining interesting models and rules from spatial data sets, composed of images and associated ground data, can be of significance in precision agriculture, resource discovery, and other areas. However, in some cases, the sizes of the spatial data sets are very huge to be mined in a reasonable amount of time using the existing algorithms. Qin Ding et al., [84] proposed an efficient approach to derive association rules from spatial data using Peano count tree (P-tree) structure. P-tree structure gives a lossless and compressed representation of spatial data. Based on P-trees, a well-organized association rule mining algorithm PARM with fast support calculation and significant pruning techniques is initiated to improve the effectiveness of the rule mining process. The P-tree Based Association Rule Mining (PARM) algorithm is executed and compared with FP-growth and Apriori Algorithms.

Sustainable development and the growth of agriculture sector calls for developing its competitiveness through a better understanding of land, weather and climate, and plantation, especially prediction of events with increased correction, and systematic integration of observations and prediction into decision-making in agriculture management.
A systematic approach based on Integrated Information Systems (IISs) for agricultural ecosystem management is proposed by Xu et al., [62]. The approach engages establishing an IIS called Agricultural Ecosystem Enterprise Information System (AEEIS) that extracts data on terrain, land use, planting, and others, and combines them for the use of agricultural and ecosystem management. The combination helps in producing managerial/policy alternatives in consultation not only with the agricultural and ecological specialists, but also with agriculture and ecosystems management. AEEIS, a stage of enterprise information system, contains operational database, extract transform and load, data warehouse, data mining, simulation modeling, and knowledge management for producing managerial strategies on land use, planting species/variety, and optimal coverage of plants. AEEIS is an attempt on integrated agricultural information services, which is the main application of China's sustainable agricultural development plan.

Zelu Jia et al., [133] introduced spatial data mining technique especially decision tree algorithm applying to agriculture land grading. The idea is to join spatial data mining/decision tree techniques with expert system techniques and combine them to find an intelligent agriculture land grading information system. The author accepts the decision tree c4.5 algorithm and implements with Mo2.0 and VC++6.0 to construct agriculture land grading expert system.

The decision tree is one of the common modeling methods to classify. Firstly, Gao Yi-yang et al., [38] introduced the concept of classification and the method of the decision tree. The data of rural labor is then analyzed, arable land area and the gross output value of agriculture about 30 cities of China based on the decision tree, and assumes clustering analysis method to sort out continuous data through the process of data mining in order to subjectively evaluate the traditional classification methods. Finally, the decision tree of the agriculture is generated, thus gaining the spatial classification rules and then analyzes the rules.

The main tendency of the agriculture development is accuracy agriculture, so more and more agriculture decision support system about accuracy agriculture shows. Zhiqing Zhu et al., [134] studied how to use GIS (Geography Information System), data mining and Web technologies to apply to the agricultural expert decision system.
With the research background of Hebei expert system for farming soil variable rate fertilization, the GIS-based agriculture expert system according to the foreign advanced experience was put onward, which can combine the situation of our country, and the key technologies to execute this system are explained in detail.

The application of information technology in agriculture hastens the digitization of agriculture information. Investigating the mass of agriculture data and showing inherent knowledge are becoming new areas of research. Efficient application of data mining technology can attain this idea. Ji Dan et al., [50] presented a new improved CA algorithm based on traditional decision tree method. It initiates a pretreatment theory about dual dimension reduction which can pact with large and high-scale datasets. By utilizing CA algorithm in maize seed breeding, it can investigate the potential rules and discover useful information from it to direct the growth of maize.

Association Rule Mining (ARM) finds all the association rules in data that match some measures of interest such as support and confidence. In some situations where high support is not of interest, fixed-consequent association-rule mining for confident rules may be preferred over traditional ARM. The requirement for fixed consequent ARM is becoming more obvious in a number of applications such as Market Basket Research (MBR) or precision agriculture. Highly confident rules are desired in all situations but, support thresholds fluctuate with the applications and the data sets are under study. Rahal et al., [49] proposed an approach for mining minimal confident rules in the context of fixed-consequent ARM that relieves the user from the burden of specifying a minimum support threshold. Therefore, the framework suggested in this is well-organized and can be simply expanded by adding new pruning conditions pertaining to specific situations.

Chun Ping Chen et al., [24] proposes a novel approach for crop identification by using wavelet packet transform combined with weighted Bayes distance based on crop texture and leaf features. Repeated processing in agriculture requires correct identification of crops to target plants for treatment according to their needs. Wavelet analysis, featuring spatial/frequency localization, data compression, denoising, data analysis/data mining, is a fine candidate for recognizing crops. However, if the energy of
the wavelet packet coefficients is the sole identifying characteristic, results may differ considerably depending on different factors such as weather, plant density, growth stage, and sunlight. To conquer these variations, the weighted Bayes distance is initiated for an identification criterion, also referred to as the decision distance, where the weighting is based on the statistic of crop texture and leaves' shapes. By using the decision distance under different climates within three consecutive days of photography, the crop identification can reach a correctness of 94.63%.

In this survey, application of the data mining techniques in agriculture and its allied areas are studied. Different techniques of data mining have been used in this field. Market basket analysis helps the sellers by placing the most frequently purchased items together and thereby it helps the customers to take a quick decision. In the field of agriculture, it helps the farmers to select the right crop based on the quality of soil and weather condition. It also assists in taking decisions to determine the fertilizer level. But it is revealed from the survey that there is no potential MBA technique that can be effectively used for agriculture to select the right crop for right soil and to determine the fertilizer level. The survey aims to come out with the techniques being used in the agriculture and its allied areas.

2.5. SUMMARY

This chapter describes the literature Survey on Association Rule Mining Algorithms, Apriori Algorithms, Market Basket Analysis Systems and Agriculture and Data Mining Techniques. It also tells why and for what purpose this algorithm is used in the research. This works as a wide area for the researchers to develop a better data mining algorithm.