ISSN (Print): 2476-8316
ISSN (Online): 2635-3490
Dutse Journal of Pure and Applied Sciences (DUJOPAS), *Vol. 5 No. 2b December 2019*

Application of Apriori Associative Data Mining Algorithm in Jigawa State Civil Registration Database

Aminu Musa[1], Aminu Abdullahi Ahmad[2]
[1,2]Department of Computer Science,
Jigawa State College of Education,
Gumel, Jigawa State.
Email: amasayaya@gmail.com

Abstract
The importance of data mining in discovering meaningful information, new correlations, patterns and trends in large repositories is very enormous in a world where data is being collected every passing day. One has to make sense out of it by going through large amounts of data stored in those repositories, using pattern recognition technologies, Clustering techniques as well as statistical and mathematical techniques. Nowadays, data mining is a modern and powerful tool, automating the process of discovering relationships and combinations in raw data and using the results in an automatic decision support. In the research Apriori association data mining algorithm was applied to Jigawa State, Nigeria Civil Registration Database and rules were generated. The algorithm was able to generate ten general rules using minimum support and confidence that provides new in depth knowledge about a given locality such as Number of Single mother's families, Number of families without employment and formal education, literacy level of families in the locality were both identified.

Keywords: Data Mining, Machine learning, Algorithm, Civil registration

INTRODUCTION
Data mining are techniques for finding and describing structural patterns in data as a tool for helping to explain that data and make predictions from it. It is a process of discovering patterns in data. The process must be automatic or (more usually) semi automatic. The patterns discovered must be meaningful in that they lead to some advantage, usually an economic advantage. The data is invariably present in substantial quantities, data mining is about solving problems by analyzing data already present in databases.

According (Kauffman, 1990) Data mining is the process of discovering insightful, interesting, and novel patterns, as well as descriptive, understandable, and predictive models from large scale data. Association rule mining finds interesting association or correlation relationships among a large set of data items. With massive amounts of data continuously being collected and stored in databases, many industries are becoming interested in mining association rules from their databases. This research focuses on using data mining techniques (Association) to study and analyze the huge data of Jigawa state civil registration database and then generate rules that can tell us which local government has a high rate of poverty, illiteracy or unemployment,

*Author for Correspondence

and which local government needs to be developed in a specific field such as education, unemployment, mortality rate to aid in critical decision making and National planning.

This research will tend to apply data mining algorithm on Jigawa State civil registration database and investigate new knowledge with the aim of finding fantastic rules, patterns and relations that will bring out new knowledge from that archive and give an insight on Jigawa states rate of birth, mortality rate, unemployment records and so on, depending on how rich the database is, for example: how many illiterate women per given local government, how many educated people are unemployed, level of education per gender e.t.c.

PREVIOUS WORK

Waikato Environment for Knowledge Analysis (WEKA) is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from Java code. WEKA contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

Machine learning (ML) is a domain of artificial intelligence that involves constructing algorithms that can learn from experience. The way that ML algorithms work is that they detect hidden patterns in the input dataset and build models. Then, they can make accurate predictions for new datasets that are entirely new for the algorithms. This way the machine became more intelligent through learning; so it can identify patterns that are very hard or impossible for humans to detect by themselves. ML algorithms and techniques can operate with large datasets and make decisions and predictions

Dhwani, (2017) on her paper presents an application of data mining in the analysis of census and looks at the noticeable trends. The objective of the paper was to discover the relevant information of gender inequality in all spheres from the primary census of the Raigarh district in Maharashtra with the help of appropriate data mining methods like clustering, visualization using the Weka tool.

Ashokkumar and Solanki, (2014) used WEKA in the areas containing Data Mining techniques for analysis about the disease highly affected tribal zone of Gujarat, which is known as Sickle Cell Disease (SCD) to generate predictive models for classification of blood group.

They further applied two data mining algorithms and compared the two classification techniques J48 and Random tree. It shows that both classification algorithms can classify specific blood group with respect to Age as the dependent variable. Random tree produce depth decision tree respect to J48 and that is helpful to researchers. From their experiments they conclude that those specific blood groups have more chances of SCD.

Maryam *et al,* (2018) in their review provides a survey of the machine learning classification techniques that have been proposed to help healthcare professionals in diagnosing heart disease. They start by over viewing the machine learning and describing brief definitions of the most commonly used classification techniques to diagnose heart disease. Then, reviewed various research works on using machine learning classification techniques in this field. Also, a detailed tabular comparison of the surveyed papers were presented which shows a promising future on how machine learning can be used to aid in heart diseases diagnosis.

ALGORITHMS

In general terms, data mining comprises techniques and algorithms, for determining interesting patterns from large datasets. There are currently hundreds of algorithms that perform tasks such as frequent pattern mining, clustering, and classification, among others. Understanding how these algorithms work and how to use them effectively is a continuous challenge faced by data mining analysts, researchers, and practitioners, because the algorithm behaviour and patterns it provides may change significantly as a function of its parameters:

(a) **C4.5 Algorithm** :
C4.5 constructs a classifier in the form of a decision tree. In order to do this, C4.5 is given a set of data representing things that are already classified. Systems that construct classifiers are one of the commonly used tools in data mining. Such systems take as input a collection of cases, each belonging to one of a small number of classes and described by its values for a fixed set of attributes, and output a classifier that can accurately predict the class to which a new case belongs. Bharti, (2014).

(b) **SVM Algorithm** :
Support vector machine (SVM) learns a hyperplane to classify data into 2 classes. At a high-level, SVM performs a similar task like C4.5 except SVM doesn't use decision trees at all. In today's machine learning applications, support vector machines are considered a must try—it offers one of the most robust and accurate methods among all well-known algorithms. It has a sound theoretical foundation, requires only a dozen examples for training, and is insensitive to the number of dimensions. In addition, efficient methods for training SVM are also being developed at a fast pace. (Madhes, 2016)

(c) **Apriori Algorithm :**
One of the most popular data mining approaches is to find frequent item sets from a transaction dataset and derive association rules. Finding frequent item sets (item sets with frequency larger than or equal to a user specified minimum support) are not trivial because of its combinatorial explosion. Once frequent item sets are obtained, it is straightforward to generate association rules with confidence larger than or equal to a user specified minimum confidence. Apriori is a seminal algorithm for finding frequent item sets using candidate generation. The Apriori algorithm learns association rules and is applied to a database containing a large number of transactions. (Madhes, 2016).

(d) **k-means Algorithm** :
is a simple iterative method to partition a given dataset into a user specified number of clusters, k. k-means creates k groups from a set of objects so that the members of a group are more similar. It's a popular cluster analysis technique for exploring a dataset (Madhes, 2016).

(e) **Naive Bayes Algorithm** :
Is not a single algorithm, but a family of classification algorithms that share one common assumption: Every feature of the data being classified is independent of all other features given the class. Two features are independent when the value of one feature has no effect on the value of another feature (Madhes, 2016).

(f) **PageRank Algorithm :**
Is a search ranking algorithm using hyperlinks on the Web. Page Rank produces a static ranking of Web pages in the sense that a Page Rank value is computed for each page off-line and it does not depend on search queries. It's a type of network analysis looking to explore the associations among objects (Phyu, 2013).

(g) **kNN Algorithm :**
k-Nearest Neighbours, is a classification algorithm. However, it differs from the AdaBoost classifiers because it's a lazy learner. A lazy learner doesn't do much during the training process

other than store the training data. Only when new unlabeled data is input does this type of learner look to classify the items. (Madhes, 2016).

(h) **CART Algorithm :**

CART stands for classification and regression trees. It is a decision tree learning technique that outputs either classification or regression trees. Like C4.5, CART is a classifier. A classification tree is a type of decision tree. (Bhumika, 2017).

DATA MINING TOOLS

Data mining is not all about the tools or database applications or software that are being used we can perform data mining with comparatively modest database systems and simple tools, including creating and writing our own, or using existence software packages.

Data Mining Techniques: Several core techniques were used in data mining depending on the nature of the dataset, work and information intended to extract. Below are most popular techniques of data mining.

(i) Association: an Association technique or relation is probably the better known and most familiar and straightforward data mining technique. Here you make a simple correlation between two or more items, often of the same type to identify patterns.

(ii) Classification: you can use classification to build up an idea of the type object by describing multiple attributes to identify a class. Additionally, you can use classification as a feeder to, or the result of, other techniques. For example, you can use decision trees to determine a classification. Clustering allows you to use common attributes in different classifications to identify clusters.

(iii) Clustering: clustering technique examines one or more attributes or classes, you can group individual pieces of data together to form a structure opinion. At a simple level, clustering is using one or more attributes as your basis for identifying a cluster of correlating results. Clustering is useful to identify different information because it correlates with other examples so you can see where the similarities and ranges agree.

(iv) Prediction: prediction is a wide topic and runs from predicting the failure of components or machinery, to identifying fraud and even the prediction of company profits. Used in combination with the other data mining techniques, prediction involves analyzing trends, classification, pattern matching, and relations by analyzing past events or instances, you can make a prediction about an event.

METHODOLOGY

Cross Industry Standard Process for Data Mining (CRISP-DM) steps were followed. Then WEKA Machine learning tool was used to apply the Apriori association rule algorithm to the dataset.

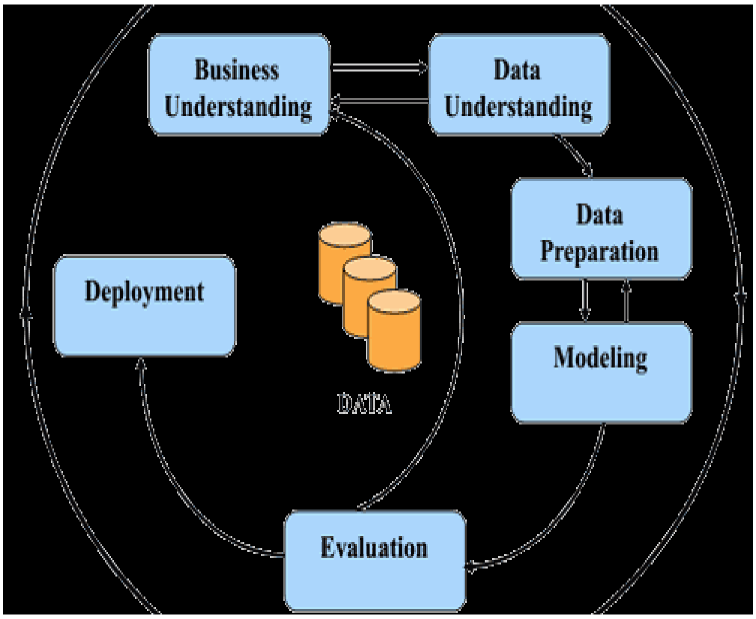


Fig 1: CRISP DM

Data Understanding

The civil registration database is a national database that are made up of birth, death and stillbirth records from continuous and compulsory birth and death registration exercise act of National Population Commission(NPC).

Data Preparation

The data collected from NPC database was in manual format kept in files using paper and pencil, it was collected in that format, converted to digital format, in Microsoft excel, then exported into csv format for cleaning where non relevant information were removed and convert it into a format that WEKA support that is arrf file format.

```
CHILD NAME,CHILD DOB,    ADDRESS,MOTHER NAME,MOTHERS AGE,MOTHERS MARITAL STATUS,MOTHERS EDUCATIONAL LEVEL ,MOTHERS OCCUPAT
MUSBA FATIMA,15/04/2016,ZANGO,UMAR TA'IBA,35,MARRIED,QUR'ANIC,HOUSE WIFE,HABUBAKAR MUSBA,45,QUR'ANIC
YUSUF MAIMUNA,04/04/2009,GALAGAMMA,USMAN AMINA,35,MARRIED,SECONDRY,HOUSE WIFE,MUSA USMAN,40,SECONDRY
TIJJANI ILIHAN,11/08/2012,TSAMIYAR OSI,SA'AD AISHA,28,MARRIED,PRIMARY,TRADER,ABUBAKAR TIJJANI,41,NCE
MADAINU HAFSA,05/05/2011,TANDA,UMAR MAINU,25,MARRIED,PRIMARY,HOUSE WIFE,UMAR MADAINU,41,SECONDRY
TASIU MARYAM,14/05/2015,RIJIYAR MAFAFFA,NURA SHAFA'ATU,21,MARRIED,SECONDRY,HOUSE WIFE,DAHA TASI'U,33,NCE
M.DAHIRU SUMAYYA,04/12/2011,NASARAWA,MAGAJI FATIMA,25,MARRIED,SECONDRY,FARMER,MUHAMMAD DAHIRU,30,SECONDRY
LAUWALI KHADIJA ,04/03/2009,TANDA,MUSA RUKIYYA,24,MARRIED,PRIMARY,FARMER,ABUDULLAHI LAUWALI,50,QUR'ANIC
UMMARU KHADIJA,04/03/2009,TANDA,YAKUBU ZAITU,25,MARRIED,SECONDRY,CIVIL SERVANT,ALASAN UMMARU,40,NCE
M.DAHIRU HAFSAT,02/01/2011,NASARAWA,MAGAJI FATIMA,25,MARRIED,SECONDRY,HOUSE WIFE,MUHAMMAD DAHIRU,30,SECONDRY
AHMAD AISHA,02/12/2016,GUMEL,MUHD RABI,30,MARRIED,SECONDRY,HOUSE WIFE,BAGUDU AHMD,55,NCE
SALISU HAJARA,13/13/2015,KOFAR FADA,AUDULKARIMU MANI,18,MARRIED,JSS,TRADER,ALIYU SALISU,35,NCE
MUHAMMED ZULAIHAT,05/04/2017,TSAMIYAR OSI,DALHAT NAFEESAH,29,MARRIED,ILLITRATE,TRADER,MUHAMMED MUHADDED D,60,NCE
TIJJANI SAKINA,09/01/2016,TSAMIYAR OSI,SA'AD AISHA,28,MARRIED,PRIMARY,TRADER,ABUBAKAR TIJJANI,41,NCE
JAMILU SAFIYYA,12/12/2015,NASARAWA,ZULAIHA UBA,25,MARRIED,SECONDRY,HOUSE WIFE,UBA JAMILU,34,PRIMARY
SAMA'IAL FATIMA,12/12/2015,NASARAWA,SAMA'ILA MARIYA,25,MARRIED,PRIMARY,HOUSE WIFE,SAMA'ILA UMARU,40,PRIMARY
SAMA'IAL HAFSA,04/04/2011,NASARAWA,SAMA'ILA MARIYA,25,MARRIED,PRIMARY,HOUSE WIFE,SAMA'ILA UMARU,40,PRIMARY
MUHAMMAD HAFSA,05/12/2009,ZANGO,YAU HAJARA,35,MARRIED,PRIMARY,HOUSE WIFE,IDRIS MUHAMMAD,71,SECONDRY
MUHAMMAD NASIBA,05/11/2012,ZANGO,YAU HAJARA,47,MARRIED,PRIMARY,HOUSE WIFE,IDRIS MUHAMMAD,71,SECONDRY
YUSIF HADIZA,05/05/2009,GALAGAMMA,IDRIS HALIMA,33,?,?,HOUSE WIFE,IDRIS YUSUF USMAN,40,SECONDRY
SALISU KHADIJA,03/07/2017,TSAMIYAR OSI,SABO AISHA,25,MARRIED,SECONDRY,TRADER,YUSUF SALISU,29,NCE
HASAN ZAINAB ,11/10/20111,TSAMIYAR OSI,SHEHU BILKISU,30,MARRIED,PRIMARY,BUSINESS,ABULLAHI HASAN,60,PRIMARY
TIJJANI MARYAM,06/05/2010,TSAMIYAR OSI,MARYAM,30,MARRIED,PRIMARY,BUSINESS,ABUBAKAR TIJJANI,30,PRIMARY
HARUNA AISHA,07/06/2015,TSAMIYAR OSI,USMAN HADIZA,35,MARRIED,ILLITERATE,TRADER,ISYA HARUNA,50,NCE
AHMED UMMASSALAMA,13/6/2012,TSAMIYAR OSI,ISAH YAMUNA,30,MARRIED,ARABIC,BUSINESS,ADO AHMED,40,PRIMARY
HAMUSU FARIHA,08/11/2015,TSAMIYAR OSI,USMAN HADIZA,35,MARRIED,ILLITERATE,TRADER,ISYA HARUNA,50,NCE
HARUNA HAFSA,02/04/2016,TSAMIYAR OSI,USMAN HADIZA,35,MARRIED,ILLITERATE,TRADER,ISYA HARUNA,50,NCE
LAWAN MARIYA,04/03/2011,?,DAHIRU FAUZIYA,40,WIDOWED,SECONDRY,TRADER,SHA'AIBU LAWAN,70,NCE
HABIBU ZUWAIRA,10/01/2010,TSAMIYAR OSI,MOHD HADIZA,26,MARRIED,PRIMARY,BUSINESS,HABU HABIBU,45,PRIMARY
SALISU KHADIJA,10/12/2014,TSAMIYAR OSI,ABUBAKAR SADIYA ,26,MARRIED,NCE,BUSINESS,MANU SALISU,30,NCE
TIJJANI FATIMA,07/01/2013,TSAMIYAR OSI,SA'AD AISHA,25,MARRIED,PRIMARY,TRADER,ABUBAKAR TIJJANI,38,NCE
AMINU FATIMA,10/12/2014,TSAMIYAR OSI,ADAM NANA,30,MARRIED,?,?,ISIYA AMINU,40,?
MUNTARI HAUWA'U,05/11/2013,TSAMIYAR OSI,ABDULLAHI NASIBA,26,MARRIED,NCE,TRADER,HALILU MUNTARI,40,NCE
AMINU NAFISA ,10/11/2012,TSAMIYAR OSI,ADAM NANA,35,MARRIED,?,?,ISIYA AMINU,40,?
KABIRU KHADIJA,09/10/2015,?,YAHAYA KHADIJA,27,MARRIED,ILLITERATE,TRADER,AHMED KABIRU,50,ILLITERATE
KABIRU SURAYYA,15/10/2013,?,YAHAYA KHADIJA,25,MARRIED,ILLITERATE,TRADER,AHMED KABIRU,50,ILLITERATE
```

Fig.2 Sample dataset

Dataset Description

The dataset collected was only for one local government as the database was in manual format, it consist of 99 instances, with each record having about twenty (20) attributes of many types ranging from numeric and strings, the attributes are sub divided into four major categories, the applicant bio data, father information, mother information and informant information.

Table 1 dataset description

S/N	Month	Record	Total
1	January 2017	33	
2	February 2017	21	
3	March 2017	27	
4	April 2017	18	99

Model Building

After data preparation stage, the next stage in CRISP-DM of Data mining is model building, our clean dataset was brought up to WEKA platform, datasets were visualized, an associative data mining algorithm Apriori algorithm was applied to the dataset to generate rules with minimum support, in order to generate reasonable rules that can be translated into meaningful information that can be used or aid in decision making.

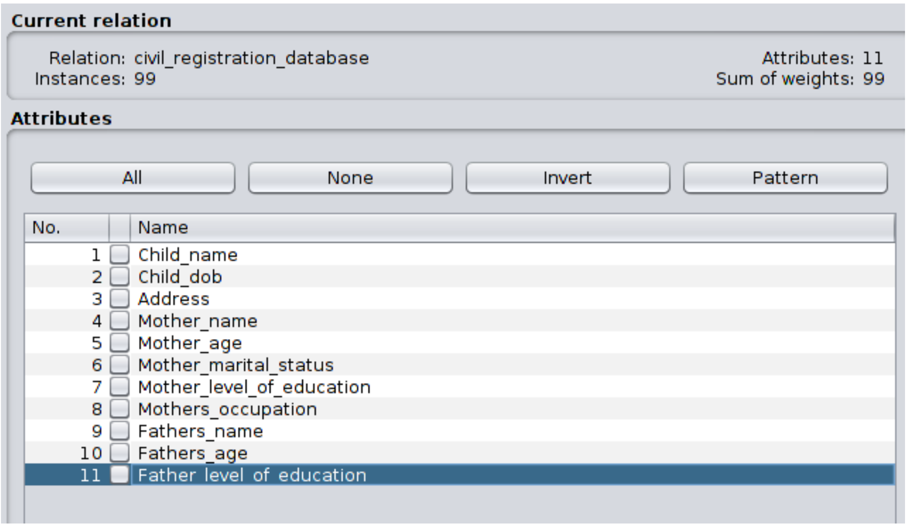


Fig. 3 dataset attributes

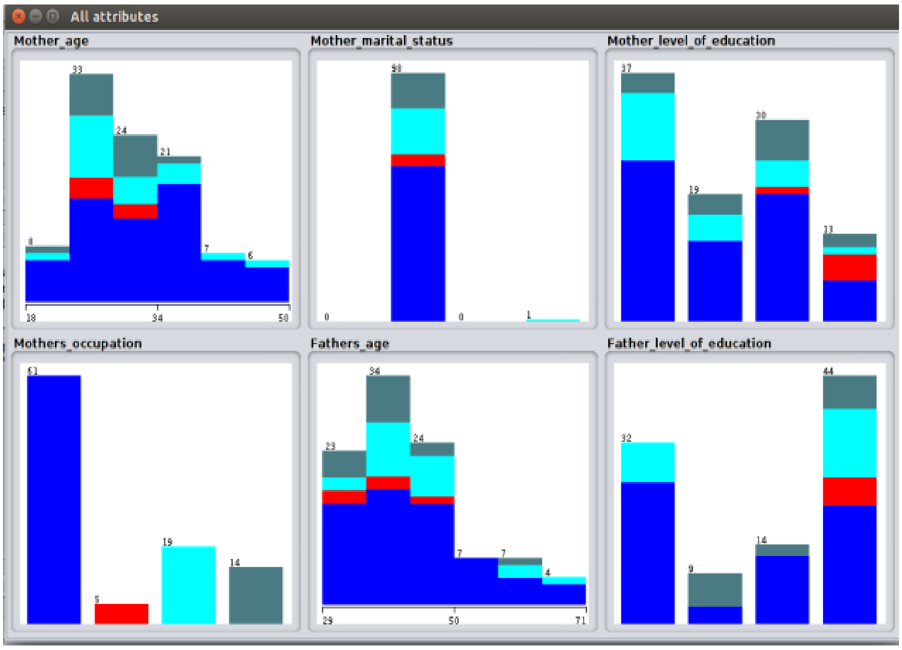


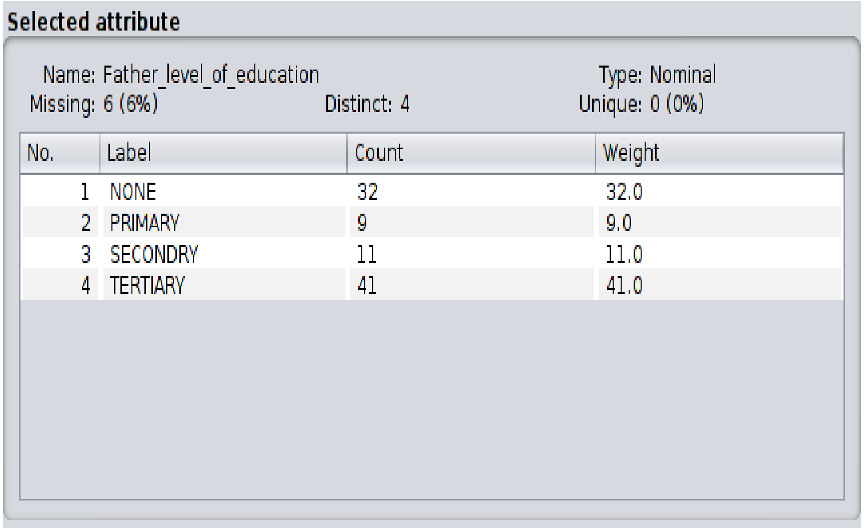
Fig. 4 different attributes visualization

Selected attribute

Name: Mothers_occupation   Type: Nominal
Missing: 0 (0%)   Distinct: 4   Unique: 0 (0%)

No.	Label	Count	Weight
1	HOUSE_WIFE	61	61.0
2	CIVIL_SERVANT	5	5.0
3	TRADER	19	19.0
4	BUSINESS_WOMAN	14	14.0

Fig. 5 Mother occupation attribute



Selected attribute

Name: Father_level_of_education Type: Nominal
Missing: 6 (6%) Distinct: 4 Unique: 0 (0%)

No.	Label	Count	Weight
1	NONE	32	32.0
2	PRIMARY	9	9.0
3	SECONDRY	11	11.0
4	TERTIARY	41	41.0

Fig. 6 Father level of education attribute

APRIORI ASSOCIATIVE ALGORITHM

An association rule is a rule which implies certain association relationships among a set of objects (such as occur together or one implies the other) in a database. Given a set of transactions, where each transaction is a set of literals (called items), an association rule is an expression of the form X Y, where X and Y are sets of items. The intuitive meaning of such a rule is that transactions of the database which contain X tend to contain Y (Rashmi, 2014).
An example of an association rule is: 30% of children registered by thier mother also their father's is employed; 2% of all children are registered by their father's and also works. Here 30% is called the confidence of the rule, and 2% of the support of the rule. The problem is to find all association rules that satisfy user-specified minimum support and minimum confidence constraints.

Support: is an important measure because a rule that has very low support may occur simply by chance. A low support rule is also likely to be uninteresting from a business perspective because it may not be profitable to promote items that customers seldom buy together. For these reasons, support is often used to eliminate uninteresting rules. Support also has the desirable property that can be exploited for efficient discovery of association rules.

Confidence: on the other hand, measures the reliability of the inference made by a rule. For a given rule X −→ Y, the higher the confidence, the more likely it is for Y to be present in transactions that contain X. Confidence also provides an estimate of the conditional probability of Y given X. Association analysis results should be interpreted with caution. The inference made by an association rule does not necessarily imply causality. Instead, it suggests a strong co-occurrence relationship between items in the antecedent and consequent of the rule. Causality, on the other hand, requires knowledge about the causal and effect attributes in the data and typically involves relationships occurring over time. (Kumar, 2005)

DESCRIPTION OF ASSOCIATION RULE MINING PROBLEM:

The association rule mining problem can be formally stated as follows as given a set of transactions T, find all the rules having support ≥ min-sup and confidence ≥ min-conf ,where min-sup and min-conf are the corresponding support and confidence thresholds. A brute-force approach for mining association rules is to compute the support and confidence for every possible rule. This approach is prohibitively expensive because there are exponentially many rules that can be extracted from a data set. More specifically, the total number of possible rules extracted from a data set that contains d items is : (Kumar, 2005).

R = 3d − 2d+1 + 1

Apriori employs an iterative approach known as a level-wise search, where k-itemsets are used to explore. kC 1/-itemsets. First, the set of frequent1-itemsets is found by scanning the database to accumulate the count for each item, and collecting those items that satisfy minimum support. The resulting set is denoted by L1.Next, L1 is used to find L2, the set of frequent 2-itemsets, which is used to find L3, and so on, until no more frequent k-itemsets can be found. The finding of each Lk requires one full scan of the database.

RESULT & DISCUSSION

RESULT

=== Run information ===

Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation: civil_registration_database
Instances: 99
Attributes: 4
Mother_marital_status
Mother_level_of_education
Mothers_occupation
Father_level_of_education
=== Associator model (full training set) ===

Apriori
=======

Minimum support: 0.2 (20 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 16

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6

Size of set of large itemsets L(2): 9

Size of set of large itemsets L(3): 5

Size of set of large itemsets L(4): 1

Best rules found:
1. Mothers_occupation=HOUSE_WIFE 61==> Mother_marital_status=MARRIED 61 <conf:(1)> lift:(1.01) lev:(0.01) [0] conv:(0.62)
2. Mother_level_of_education=NONE 37 ==> ,Mother_marital_status=MARRIED 37 <conf:(1)> lift:(1.01) lev:(0) [0] conv:(0.37)

3. Father_level_of_education=NONE 32 ==> Mother_marital_status=MARRIED 32 <conf:(1)> lift:(1.01) lev:(0) [0] conv:(0.32)
4. Mother_level_of_education=NONE Father_level_of_education=NONE 27 ==> Mother_marital_status=MARRIED 27 <conf:(1)> lift:(1.01) lev:(0) [0] conv:(0.27)
5. Mothers_occupation=HOUSE_WIFE Father_level_of_education=NONE 25 ==> Mother_marital_status=MARRIED 25 <conf:(1)> lift:(1.01) lev:(0) [0] conv:(0.25)
6. Mother_level_of_education=NONE Mothers_occupation=HOUSE_WIFE 24 ==> Mother_marital_status=MARRIED 24 <conf:(1)> lift:(1.01) lev:(0) [0] conv:(0.24)
7. Mothers_occupation=HOUSE_WIFE Father_level_of_education=TERTIARY 21 ==> Mother_marital_status=MARRIED 21 <conf:(1)> lift:(1.01) lev:(0) [0] conv:(0.21)
8. Mother_level_of_education=NONE Mothers_occupation=HOUSE_WIFE Father_level_of_education=NONE 21 ==> Mother_marital_status=MARRIED 21 <conf:(1)> lift:(1.01) lev:(0) [0] conv:(0.21)
9. Father_level_of_education=TERTIARY 44 ==> Mother_level_of_Education=TERTIARY 44 <conf:(0.98)> lift:(0.99) lev:(-0.01) [0] conv:(0.22)
10.Mother_level_of_education=SECONDRY30==>Mother_marital_status=DIVORCED 29 <conf:(0.97)> lift:(0.98) lev:(-0.01) [0] conv:(0.15)

INTERPRETATION OF RESULTS

Rule 1
It was discovered that there were 61 housewives without current or previous employment

Rule 2
There were 37 Housewives who has no prior education at least primary level

Rule 3
There were 32 married Men with at least 1 Child but have no prior Education

Rule 4
There were 27 families where both the husband and the wife had no educational background

Rule 5
There were 25 instances of housewives with no employment, a husband without educational background

Rule 6
There were 25 instances of married men without a job and wives without educational background

Rule 7
It shows that there were 21 families in which the husbands have formal education up to the tertiary level married to fulltime housewives.

Rule 8
It discovered 21 families where both parents are uneducated and jobless at the same time.

Rule 9
It highlights 44 families in which both the husband and the wife have tertiary level of formal Education

Rule 10
It found out 29 classes of women who are divorced with at least one child and has formal education up to tertiary level.

DISCUSSION

The program generated the sets of large item sets found for each support size considered. In this case eight hundred eighty-two item sets of two items were found to have the required Minimum support. By default Apriori tries to generate ten rules. It begins with a minimum support of 100% of the data items and decrease this in steps of 4% until there are at least ten rules with the required minimum confidence, or until the support has reached a lower bound of 10% whichever occur first . the minimum confidence is set 0.4 (40%). the minimum support decreased to 0.2 (20%) in our dataset the father level of education and occupation. Generation of the required number of rules involved a total of 16 iterations. The last part gives the association rules that are found. The number preceding ==> symbol indicates the rules support that is the number of items covered by its premise. Following the rule is the number of those items for which the rules consequent holds as well, In the parentheses there is a confidences of the rule.

CONCLUSION

In this study, data mining technique of Association rule using Apriori algorithm was used to generate rules which presented new knowledge hidden in the dataset that can be used by decision makers to solve specific problem. This study provides an insight to what is going on in Civil registration database in many fields such as the Educational , Employment tailored to Gender and age group, it also went further to highlight the death rate per gender, per age group, single mothers in a given society and many more. The study will enable decision makers do the right thing for a given group of rules. It provides rules that shows decision makers where the problems are and how they could solved it by making the right decision.

REFERENCES

Bharti, T. and Manish M., (2014). Data Mining With Big Data Using C4.5 and Bayesian Classifier: ijarcssee. 4(8), 2014.

Bhumika, G., (2017). Analysis of Various Decision Tree Algorithms for Classification in Data Mining. IJCA. 163(8), 263-284.

Gary, M. and Brian, D., (2010). Data Mining To appear in the Handbook of Technology Management. John Wiley and Sons.

Jiawei, H. and Kamber *et al* (2012). Data Mining Concepts and Techniques Third Edition, Morgan Kaufmann Publishers.

Kumar S. and Pang *at el* (2005). Introduction to Data Mining, Addison-Wesley.

Kavitha, P. and Sasipraba, T., (2011). Performance Evaluation of Algorithms using a Distributed Data Mining Framework based on Association Rule Mining. International Journal on Computer Science & Engineering (IJCSE), 2011.

KNIME (2017). KNIME Official Website, On-line available to https://www.knime.com/, 2017, up to date August 2017.

LIBSVM (2017) LIBSVM Official Website, On-line available at: https://www.csie.ntu.edu.tw/~cjlin/libsvm/, last visitied August 2019.

Moawia, E. (2010), A New Approach for Evaluation of Data Mining Techniques, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 5, September 2010.

Manoj, B. (2012) Manoj Bala, Study of Application of Data Mining Technique in Education. International Journal of Research in Science and Technology, Vol. No. 1, Issue No. IV, Jan-March, 2012.

Martin B. (2012). Data mining techniques 2012 On-line available to https://www.ibm.com/developerworks/library/ba-data-mining-techniques/ last visited August 2019.

Mohammed, J. and Wagner, M., (2014). Data Mining and Analysis: Fundamental Concepts and Algorithms, Cambridge University Press, 2014.

Madhes, D., (2016). Data Mining Tools & Algorithms, 2016 On-line available: https://dwbicastle.com/2016/12/15/data-mining-tools-algorithms/ up to date August 2019.

Phyu, T., (2013)., Proposed Approach For Web Page Access Prediction Using Popularity And Similarity Based Page Rank Algorithm, IJSTR 3(2).

Rupali, G., (2013). Data Mining: Techniques, Applications and Issues, IJARCSEE. 2(2), 2013.

Rashmi, J., (2014). Association Rule Mining Algorithm In Parallel And Distributed Data Mining, International Multidisciplinary e-Journal. 2277 – 4262, 2014, On-line available at http://www.shreeprakashan.com/Documents/20140707132413535.1.Rashmi%20Jha..pdf, up to date August 2019.

Rattle, (2018) Rattle Official Website, On-line available to https://rattle.togaware.com/, 2018, up to date August 2019.

Witten, I. and Eibe, F., (1999). Data Mining, Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann, 1999.

Weka, (2017) Official Website, On-line available to http://www.cs.waikato.ac.nz/ml/weka/, 2017, up to date August 2019

Yihao, L., (2010). Data Mining: Concepts, Background, and Methods of Integrating, Uncertainty in Data Mining. On-line available at : http://www.ccsc.org/southcentral/E-Journal/2010 last visited August 2019.