Determining the correlation between the Hospital Resources and Mortality Rates

[1]Harshitha M.C, [2]Mamatha B.S
[1]PG Scholar, [2]Associate Professor
[1]Department of Computer Science and Engineering,
[1]PESCE, Mandya, India

Abstract : According to the research of public health statistics data the mortality rate is increasing sequentially. So the most important task of health service is to reduce the number of inpatient hospital deaths. The hospitals are rich sources of information of considerable important to the health system, but this role has been neglected. The target of this project is to manifest ab correlation between mortality and hospital resources by implementing Apriori algorithm and Apriori TID algorithm, then the efficiency of the algorithms are compared. This kind of knowledge can be discovered from hospital datasets which is important in order to reduce the death rates in the hospitals. The data set is taken from UCI machine learning repository.

***IndexTerms* - Association Rules, correlation, Data Mining, Mortality.**

I. INTRODUCTION

Association rule mining is an emerging research in data mining. Data mining is a process of extracting and querying undetermined data from enormous amount of data. Extracting frequent patterns, interesting correlations, associations or casual structures among the sets of items in the transaction databases or other data storage is the main aim of association rule. Association rules are extensively used in different areas such as risk management, telecommunication networks and inventory control etc. The dispute is mainly resolved into two sub problem. The first problem is to detect the item sets whose event exceed a predefined threshold in the database, the item sets used in this problem are known as large item sets or frequent item sets. The second problem is to induce association rules from those frequent item sets with the limitations of minimal confidence.

Mortality rate is a measure of the number of deaths in a particular population. Hospital mortality is defined as death occurring during the hospital stay. Hospital death rates are important indicators that can be influenced by the Quality of care. In terms of data analysis, analyze the association between hospital resources and mortality is the mandatory task for public health's policy management. To reducing the mortality of the population the ministry of public health has to provide detail medical knowledge and modern technology. In the yearly budget planning the main task is to allocate the cost control, resource and budget. Thus the main aim of this project is to analyze a correlation between hospital resource information with mortality population. In this project we use data attributes such as number of hospitals, Neurologist, Cardiologist, Gynecologist, Orthopedics, Surgeon, Physician, Beds, ICU departments, Nurses and Mortality rates.

II. LITERATURE SURVEY

"**Mining Medical Data to Identify Frequent Diseases using Apriori Algorithm**" (M Ilayaraja, T Meyyappan 2013) In this paper, creator built up a strategy to recognize recurrence of diseases in specific topographical zone at given time span with the association rule based Apriori data mining technique. To determine the frequency of the disease that are repeating in individuals living in different topological areas during various time spans, WEKA data mining technique is utilized. The examination uncovered the way that 4 unique sicknesses affected the patients as often as possible at different land areas during the year 2012. Existing electronic medical records acquired from clinics are utilized as training data set for the examination. Absolutely 1216 patient records affected by 29 unique diseases during the year 2012 are analysed.

"**A Mortality Study for ICU Patients using Bursty Medical Events**" (Luca Bonomi, Xiaoqian Jiang 2017) In this paper, they exhibited some fundamental outcomes that demonstrated an interesting relation between mortality and the burstiness of intervent time sequences for ICU patients. Specifically, the outcomes demonstrated that the mortality for bursty patients achieves 59.8% contrasted with 52.9% observed the entire patients. Besides, this parameter is symmetrical to the components at presently utilized in critical care mortality studies. The mortality investigation of ICU patients is quite compelling in light of the fact that it gives helpful signs to human services foundations for improving patients experience, internal policies and procedures. Therefore the burstniness parameter gives new information that can be successfully used to improve state of the art prediction models. Future research headings comprise in directing a more inside and out medicinal investigation of specific diseases helped by clinicians and joining worldly highlights for growing new predictive risk models.

"**An Excellent Mortality Prediction Model Based on Support Vector Machine (SVM)-a Pilot Study** " (Chien-Lung Chan, Chia-Li Chen, Hsien-Wei Ting 2010) Intensive therapy is a standout components of the most significant segments of the advanced medicinal framework. Healthcare experts need to use escalated care assets successfully. Mortality prediction method help doctors to decide which patients require concentrated consideration the most and which don't. This pilot consider reflectively gather data on 695 patients admitted to intensive therapy unit and develop a novel mortality forecast model with support vector machine. The precision of new model is great. The exactness rate is 0.899. The review rate is 0.902. The F-Measure is 0.899. The ROC curve is 0.932. This new model can help the doctor's in intensive therapy decision making.

"Can Cluster-Boosted Regression Improve Prediction of Death and Length of Stay in the ICU? Mahsa" (Rouzbahman, Aleksandra Jovicic) In this paper, they look over the extent to which analysis of grouped patient types con coordinate expectations made by breaking down the whole dataset at once. After reviewing all the related literature and clarifying how data is outlined in each group of similar patients, they compare the results of predicting death and length of remain in the ICU1 utilizing regression analysis on unique and grouped data from MIMIC II dataset. Grouping improved regression prediction accuracy for both death and length of remain. We discuss about the results regarding their suggestions for the future utilization of health repository based data analytics to give an enhancement to existing techniques for clinical decision support.

III. METHODOLOGY

This reference system architecture explains the process of converting input data into desired output. We collect large amount of data set from UCI machine learning repository then pre-processing is applied on dataset which removes irrelevant data and extract relevant data from dataset. These relevant data are considered as training data. The correlation prediction model is trained based on the training dataset to get the desired output. The result which is gained through prediction model is represented in a user friendly manner.

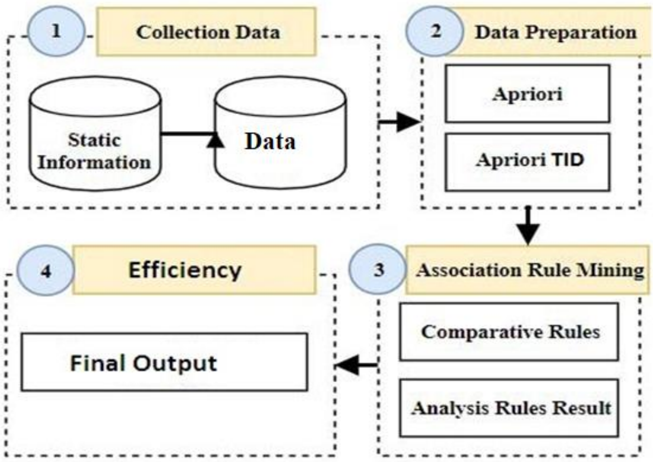

Fig 3.1: System Architecture.

A. Apriori Algorithm

This is the most significant and frequently used algorithm for mining frequent itemsets. Apriori algorithm is utilized to find all the frequent itemsets in a given database. In apriori algorithm it generates less candidate itemsets for testing in each database pass. The finding of association rules id guided by two parameters such as support and confidence. Apriori returns an association rule if its support and confidence esteems are above user defined threshold esteem. The output is ordered by confidence. If some several rules have the same confidence then they are ordered by support. Therefor apriori supports confident rules and describes these rules as additional fascinating. This algorithm utilizes breath first search approach [5].

Steps of Apriori algorithm:

- STEP 1: Scan the data set and determine the support(s) of each item.
- STEP 2: Generate L1 (Frequent one item set).
- STEP 3: Use Lk-1, join Lk-1 to generate the set of candidate k - item set.
- STEP 4: Scan the candidate k item set and generate the support of each candidate k – item set.
- STEP 5: Add to frequent item set, until C=Null Set.
- STEP 6: For each item in the frequent item set generate all non-empty subsets.
- STEP 7: For each non empty subset determine the confidence. If confidence is greater than or equal to this specified confidence .Then add to Strong Association Rule.

B. AprioriTID Algorithm

Just like the Apriori algorithm, AprioriTID algorithm uses the generation function in order to find the candidate item sets. The only difference between the two algorithms is that, in AprioriTID algorithm the database is not referred for counting support after the first pass itself. Here a set of candidate item sets is used for this purpose for k>1. When a transaction doesn't have a candidate k-item set in such a case the set of candidate item sets will not have any entry for that transaction. This will decrease the number of transaction in the set containing the candidate item sets when compared to the database. As value of k increases every entry will become smaller than the corresponding transactions as the number of candidates in the transactions will keep on decreasing. Apriori only performs better than AprioriTID in the initial passes but more passes are given AprioriTID certainly has better performance than Apriori [5].

Steps of AprioriTID algorithm:

- STEP 1: The database is not used for counting support after the first pass.
- STEP 2: Instead information in data structure C_k' is used for counting support in every step.
    - C_k' is generated from C_{k-1}'
    - For small values of K, storage requirements for data structures could be larger than the database.
    - For large values of K, storage requirements can be very small.

The association rules are written as the form of conditions and outcomes, e.g. A -> B where A is conditions and B are outcomes. The formula represents that B appears when A appears. [7]

Association rules are significant if support and confidence values are greater than the defined thresholds and can be obtained as follows.

Support (A) = a / T …………………………………... (1)

When the number of transactions that A appears, T is the number of transactions and confidence is the proportion of the transactions that contains A which also contains B.

Confidence (A->B) = Support (A ∪ B) / Support (A) … (2)

IV. EXPERIMENTAL RESULT

Data set used in this research work contain 1325 medical records. The data set includes number of hospitals, Neurologist, Cardiologist, Gynecologist, Orthopedics, Surgeon, Physician, Beds, ICU departments, Nurses and Mortality rates. The proposed system is implemented using the Visual Studio 2010 data mining tool, the result obtained are promising.

A. Pattern prediction of Apriori with execution time.

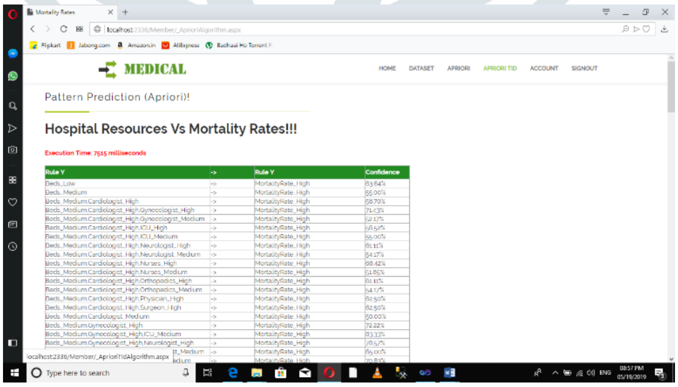

Fig 5.1: Apriori prediction result.

B. Pattern prediction of AprioriTID with execution time.

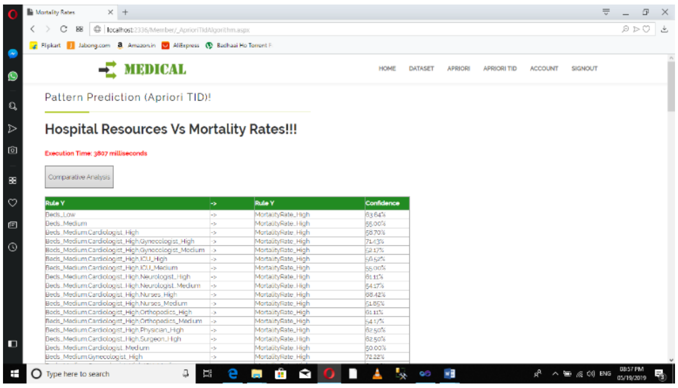

Fig 5.2: AprioriTID prediction result.

C. Comparative analysis with graph representation.

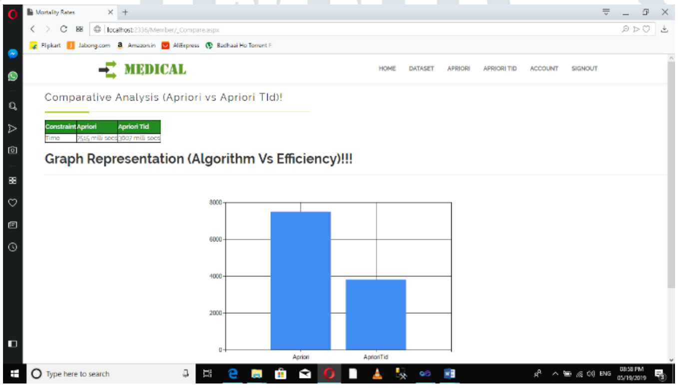

Fig 5.3: result of Comparative analysis.

CONCLUSION

In this proposed system, the study gives a correlation between mortality rate and resources of health care such as number of hospitals, Neurologist, Cardiologist, Gynecologist, Orthopedics, Surgeon, Physician, Beds, ICU departments, Nurses and Mortality rates to manage the health care services, budgeting and for utilizing resources efficiently. Data mining techniques concludes that Apriori and Apriori TID algorithm had similar result and it also concludes Apriori TID algorithm is more efficient than Apriori algorithm.

REFERENCES

[1] Vinod L. Mane, Suja S. Panicker, Vidya B. Patil, "Summarization and sentiment analysis from user health posts", 2015.

[2] M. Haider Abu Yazid, Shukor Talib, Muhammad Haikal Satria, Azmee Abd Ghazi "Neural Network on Mortality Prediction for the Patient Admitted with ADHF (Acute Decompensated Heart Failure)", 2017.

[3] K.M.D.Muthumali Karunarathna, "Predicting ICU Death With Summarized Patient Data", 2015.

[4] Chien-Lung Chan, Chia-Li Chen, Hsien-Wei Ting, "An Excellent Mortality Prediction Model Based on Support Vector Machine (SVM)-a Pilot Study ", 2010.

[5] Manisha Girotra, Kanika Nagpal, Saloni Minocha and Neha Sharma, "Comparative Survey on Association Rule Mining Algorithms ", Volume 84 – No.10, December 2013.

[6] Luca Bonomi, Xiaoqian Jiang, "A Mortality Study for ICU Patients using Bursty Medical Events" , 2017.

[7] Saha Thitithamawat, Prajak Chertchom, Panthep Buraphawichit, and ThanawatThothong, “Association Rules to Analyze Hospital Resources with Mortality Rates”, 2018.
[8] M Ilayaraja, T Meyyappan, “Mining Medical Data to Identify Frequent Diseases using Apriori Algorithm”, 2013.

[9] Rouzbahman, Aleksandra Jovicic, “Can Cluster-Boosted Regression Improve Prediction of Death and Length of Stay in the ICU? Mahsa”.

[10] Chien-Lung Chan, Chia-Li Chen, Hsien-Wei Ting, “An Excellent Mortality Prediction Model Based on Support Vector Machine (SVM)-a Pilot Study”, 2010.