

A Review Paper on Dengue Disease Forecasting Using Data Mining Techniques

Vandana Rajput¹, Prof. Amit Manjhar²

^{1,2}Department of CSE/IT

^{1,2}Madhav Institute Of Tecnology & Science Gwalior, India

Abstract- Data Mining is a standout amongst the most comprehensively and rousing zones of research with the motivation behind discovering significant information from vast data collections. In Medical ventures, Data Mining gives many points of interest, for example, the location of the extortion in medical coverage, illness forecast, and accessibility of the therapeutic solution for the patients at lower cost, recognition of reasons for sicknesses and distinguishing proof of restorative treatment techniques. It is additionally helpful to foresee the hazardous maladies like-Dengue fever, Cancer, Diabetes and so forth. In this paper, we have learned about data mining that is utilized as a part of the restorative information field. There are different issues in getting information since data are put away in various – diverse arrangement which is profoundly intricate. We depict a dengue fever forecast. Forecasts (or Prediction) are most critical things to foresee the infections on the premise of manifestations, topographical illustrations area, and others. Dengue fever is an influenza-like an infection that happens by the chomp of Aedes mosquitoes which is quick developing as a noteworthy medical issue. In this study, we characterize dengue fever, its infections, and indications. Here, we likewise characterize the association rule mining that has diverse distinctive algorithms which are additionally helpful in prediction.

Keywords- Data Mining, Dengue Fever & its Virus, Association Rule Mining(ARM), ARM Algorithms: Apriori, FP-Growth and so on.

I. INTRODUCTION

Data Mining is a notable of the most prominent and persuading zones of research by the entire of the expectation of discovering moving data from gigantic front page new sets. In Present time, Data Mining is well-suited mainstream in medicinal services exchange on the grounds that there is a favor of down to earth expository system for recognizing long shot and soak data in wellbeing information. In Medical market, Data Mining gives some portion of advantages one as identification of the rupture of trust in wellbeing safeguard, accessibility of restorative union to the patients at corrupt cost, recognition of reasons for illnesses and compassion of

therapeutic help strategies. Data Mining confronting presence in the concentration of 1990's and accomplish an effective apparatus that is sufficient for getting prior obscure example and valuable data from gigantic dataset. Different reviews highlighted that Data Mining methods elevate the information holder to study and catch unsuspected relationship among their divulgence which thus well-off for declaration making[1].

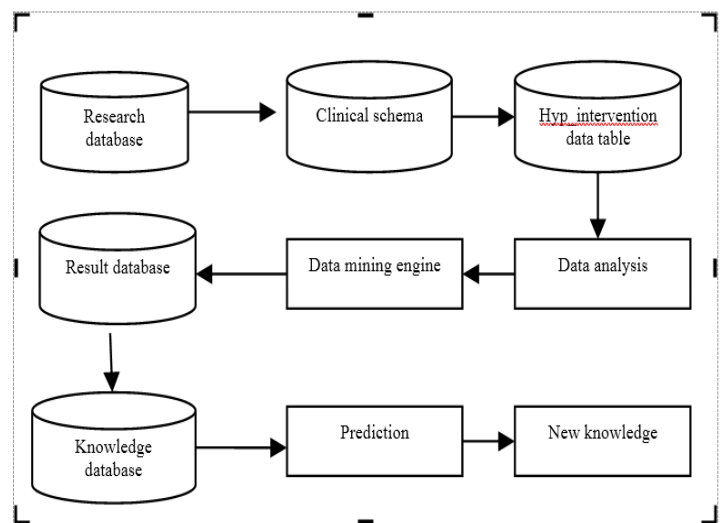


Figure 1 Basic material handling equipment used in the industry

Dengue fever is an influenza-like sickness spread by the chomp of a tainted mosquito which is speedy rising as a noteworthy medical issue. Convenient and cost productive analysis by clinical and observational components would limit the death rates besides giving top grounds to clinical expert and malady surveillance[2]. DENV is a 50 nm adjusted virion by all of three basic proteins to be specific, Capsid (C), premembrane/film (prM/M) and an envelope protein (E) encouraging parasite passage coordinated toward cell[3].

In our paper, we have brief depiction moving toward dengue fever forecast in data mining . Next , we characterize about dengue fever, viruses , side effects and need of expectation of dengue fever in II area. In III entry portrays medicinal data mining and information well known problems.

IV section have literature review, V article incorporate association rule mining algorithms.

II. DENGUE FEVER

Dengue fever (DF) is a mosquito-borne irresistible malady brought on by the infections of the family *Togaviridae*, subgenus *Flavivirus*. The transmission of this sickness is through the nibbles of vectors (*Aedes aegypti*, *Aedes albopictus*) conveying the infections having a place with *Flavi* variety. From its first appearance in the Philippines in 1953, the infection has been recognized as a standout amongst the most critical arthropod-borne viral ailment in people. Dengue infection contamination has been accounted for in more than 100 nations, with 2.5 billion individuals living in regions where dengue is endemic. As far as bleakness and in addition mortality, dengue fever might be viewed as the most imperative arthropod-conceived viral malady. According to the appraisals of the World Health Organization (WHO) roughly 50 million instances of dengue contamination are distinguished each year overall. The yearly event is assessed to associate with 100 million instances of DF and 250 000 instances of dengue hemorrhagic fever (DHF). The conclusion of DF presents incredible difficulties as the indications cover with other febrile ailments. Dengue contamination has been known to be endemic in India for more than two centuries as an amiable and self-constrained illness. Since the mid 1990s, pestilences of dengue have turned out to be more successive in many parts of India[4].

A precise and proficient conclusion of dengue assumes an essential part in the event that affirmation. Precise finding is conceivable simply subsequent to directing conclusive tests, for example, protein connected immunosorbent measures (ELISA) and continuous polymerase-chain response which depend on nucleic and corrosive hybridization[5]. Dengue fever is a sickness brought on by dengue infection and is otherwise called break bone fever is transmitted by *Aedes* mosquito.

Dengue is separated into two sorts, i.e., sort 1 and sort 2, as per world wellbeing association. Initial one is traditional dengue called dengue fever and the other is dengue hemorrhagic fever. DHF1, DHF2, DHF3 and DHF4 are further four sorts of dengue hemorrhagic fever. DHF is uncovered by begin of fever which proceeds for 2 to 7 days with number of signs like spillage of plasma, stun and frail heartbeat. It causes life undermining dengue hemorrhagic fever. In most punctual cases it's difficult to separate dengue fever from dengue hemorrhagic fever.

Dengue Viruses

Dengue fever (DF) is one of the normal most life-threatening illnesses around the world particularly in India, which is prompted by the dengue infections (DENV) group of flavivirus, principally transmitted from *Aedes aegypti* mosquito. As of late, transmission of viral invasion detailed in more than 100 nations with the simultaneous development of endemic territories transcendentally in urban and country regions has turned into a noteworthy general wellbeing concern[6]. Dengue is an existence undermining infection predominantly in a few created and additionally creating nations like India. This is an infection conceived ailment brought about by breeding of *Aedes* mosquito. Data mining is an outstanding procedure utilized by wellbeing associations for an arrangement of maladies, for example- dengue, diabetes, and disease in bioinformatics examine. Dengue fever is a sickness brought around by the Single-stranded RNA flavivirus. It is spread by *Aedes* Mosquitoes and is endemic to the tropical districts. The infection circles as four immunologically unmistakable serotypes. Regardless of this differing qualities dengue fever is ordinarily described by comparative manifestations: fever, joint torment and vascular spillage. Dengue is hard to perceive in light of the fact that the early side effects of fever, migraine, body throbs are not specific. Dengue might be misdiagnosed as different diseases: flu, typhoid, zika, jungle fever. Different arboviruses, for example, Japanese encephalitis infection, West Nile infection, Zika infection, Ross River infection, and chikungunya infection have likewise developed to bring about significant plagues as of late. DENV has been among the most troublesome infections to seclude and spread in light of the fact that the infections don't contaminate and imitate well in most research facility creatures. Contemplates on dengue in the 1940s to 1960s utilized suckling mice and different tissue culture frameworks for confinement and measure of DENV. Numerous unpassed DENV don't create cytopathic impacts (CPE) when developed in these cells. The absence of a delicate confinement and measure framework that could be utilized for unpassed wild-sort infections counteracted fast headway in dengue explore. The advancement of the mosquito immunization strategy in the mid-1970s gave a profoundly delicate technique to the separation, spread, and quantitation of DENV. The utilization of mosquitoes is particularly vital when endeavoring to spread infections with low replication productivity from somewhat symptomatic sickness and to segregate infections from sera, tissues of normally contaminated people, wild creatures, or field-got mosquitoes. Dengue contamination has imperiled 2.5 billion populaces all around the globe. Consistently there are 50 million individuals who experience the ill effects of it all around [7].

In late decades the danger of dengue disease has expanded drastically in tropical, as well as in sub-tropical

areas. There are in the vicinity of 50 and 10 crores dengue diseases consistently, and more than 5 lakh cases are hospitalized. Dengue transmission is affected by an unpredictable arrangement of variables including nature, atmosphere and climate, human conduct and dengue infection serotype-particular crowd invulnerability among the human population[8].

Symptoms of Dengue Fever

symptoms include:

- Bleeding
- Low levels of blood platelets
- Low circulatory strain
- Metallic taste in mouth
- Headache
- Muscle joint torment
- Rashes

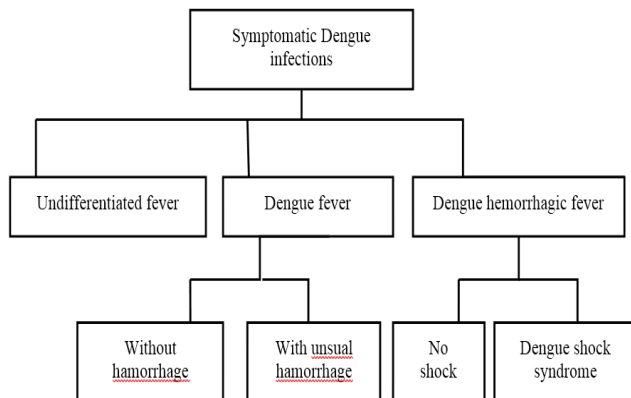


Figure 2. Symptomatic Dengue infections

There is no particular pharmaceutical or antibiotic accessible to treat it. Dengue fever happens in form of cycles and this cycle is available inside the body of a tainted individual for two weeks or under two weeks. It causes stomach torment, discharge (dying), and circulatory fall and Dengue hemorrhagic fever[9].

Need of Dengue Fever Prediction

1. To distinguish an arrangement of clinical side effects that would empower early identification of suspected dengue in kids and adults, which diminishes the danger of transmission of the DF in the group;
2. To recognize the research facility highlights and clinical side effects that would empower better finding and comprehension of the malady in speculated dengue people. This renders ideal usage of the research facility assets required for affirmed determination;

3. To form a prescient model that has a capacity of rendering successful analysis progressively.

III. MEDICAL DATA MINING

1. Issues in Medical Data Mining

Human medicinal information is immediately the most fulfilling and troublesome of every single natural data to separate and analyze. Separating helpful information and giving logical choice - making for the analysis and treatment of illness from the database progressively gets to be distinctly essential. Data mining in solution can manage this issue [3].

1.2 Data Related Problems

The enthusiasm for frameworks for self-governing basic leadership in therapeutic and building applications is developing, as information is turning out to be all the more effectively accessible. In spite of the fact that the two territories solution and designing give off an impression of being remote as far as the hidden procedures, both face numerous normal difficulties. One of the issues important to both zones is a self-sufficient forecast. Since the therapeutic data is normal for excess, multi-attribution, inadequacy, and firmly related with time, medicinal information mining differs from others one. The real zones of heterogeneity of medicinal information are:

- i. Volume and many-sided quality of medicinal information
- ii. Doctor's understanding
- iii. Affectability and specificity investigation
- iv. Poor scientific portrayal
- v. Sanctioned shape

Most related works tend to concentrate on therapeutic information heterogeneity. The synchronous thought of various information hotspots for the motivations behind thinking requires the advancement of exceedingly complex information interpretation and unification strategies – henceforth the endeavors at creating and advancing institutionalization in restorative information stockpiling and handling. The heterogeneity of medicinal information is, be that as it may, likewise communicated in the determination of semantics for individual informational collections. Gaging and profiting by the significance of chose information pieces, for example, X-beam and EMR pictures, ECG readouts and ailment histories for the reasons for expanding the demonstrative procedure demonstrates a troublesome assignment. Every one of these issues called for new strategies for investigating multi-dimensional information. However,

procedures, for example, association rule mining, grouping, outlier detection, supervised classification, social network analysis, time arrangement examination, content mining, and so on., have been utilized with some level of accomplishment. Data mining could be sure to be connected to this information heterogeneous and huge to help on three alternate points of view:

- I. Patient record administration to help clinics;
- II. Decision emotionally supportive networks to help medicinal experts and
- III. Medical research to encourage revelations of episodes, new illnesses, causes or cures.

IV. LITERATURE SURVEY

Valqu'ria Aparecida Rosa Duarte (2016) et al display that the sufficient portrayal of states in the development of wise operators is basic for permitting them to accomplish an attractive execution, essentially for those that incite in a focused domain that has a high state space. One specific sort of portrayal that is extremely fitting for these circumstances is the NetFeatureMap, which depicts by methods for components the pertinent angles that are innate to nature where the specialist actuates. In prestigious shrewd operators, such components are physically chosen, which absolutely prompts to lacking decisions. In this way, researching sufficient methodologies that play out a programming choice of these elements turns into an urgent undertaking. Along these lines, the fundamental commitment of this paper is to propose another approach that consequently chooses proper components in view of the recurrence at which they happen in the states investigated by the specialist throughout its acting over the earth. Such an approach depends on Frequent Pattern Mining. It is intriguing to bring up that there additionally exist Genetic Algorithms based methodologies that effectively adapt to a similar assignment. Not at all like Genetic Algorithms that utilization heuristic capacities to choose the components, the present proposition utilizes genuine information contained in a specific database for playing out this errand. Under the expectation of exploring the adequacy of such a proposition, the creators use the area of Checkers player specialists as their contextual analysis, since they work in an aggressive situation with a wide state space. This examination is performed by methods for competitions in which specialists whose components are chosen by the approach proposed in this face others whose elements are chosen either physically or by Genetic Algorithms. The predominant execution of the Frequent-Pattern-based operators in the competitions demonstrates the adequacy of the present proposition [10].

Neha Goyal (2016) et al display that Most of the information mining calculations were intended to mine the successive example from exact information. Nonetheless, the vulnerability exists in numerous genuine circumstances, for example, sensor system and protection safeguarding applications. To extract significant data from questionable information various frequent pattern mining algorithms have been proposed. While managing indeterminate information U-Apriori, UF-growth, UFP-growth, UH-mine, PUF-growth, TPC-growth algorithms are cases of existing frequent pattern mining algorithms, which use distinctive ways to deal with mine frequent pattern. One vital perception is that algorithms act totally unique in the unverifiable database when contrasted with the exact database because of the consideration of likelihood esteem. In this overview paper, various algorithms have been investigated for finding the frequent pattern from an indeterminate database. The investigation is spoken to as near review on taking after calculation: U-Apriori, UF-growth, UFP-development, UH-mine, and PUF-growth, TPC-growth algorithm on the premise of different parameters, for example, database scan, running time, memory use and capacity structure. The study paper likewise concentrates on the favorable position and constraint of each algorithm[11].

Md. Badi-Uz-Zaman Shajib (2016) et al display that Knowledge discovery in big data is one of most interesting topics in state-of-the-art research , and frequent patterns mining is a noteworthy undertaking. With the fast development of present day innovation, high volumes of data—which are of various veracities (i.e., might be exact or unverifiable)— are streaming at a high speed everywhere throughout the world. Properties of data transiently change with changes in the general population's interests, which make the information dynamic. Because of the vulnerability and element properties of information, finding a suitable and productive way to deal with the guarantee the proficient utilization of accessible assets has turned into an awesome test. In this paper, we outline another memory-efficient information structure, called Uncertain Stream (US)- tree, which stores late meta-information. We likewise build up a probabilistic, sliding window based, proficient algorithm—called Uncertain Stream Frequent Pattern (USFP)-growth—for mining frequent patterns from dubious information streams. Our comprehensive execution assessment demonstrates that USFP-growth is right and productive when contrasted and late related approaches[12].

Fan Jiang (2016) et al introduce that In numerous applications, web surfers might want to get the suggestion on which accumulations of pages that would be interested them or that they ought to take after. Keeping in mind the end goal to find this data and make the suggestion, data mining as a

rule—and frequent pattern mining or web mining in particular—can be relevant. Since its presentation, frequent pattern mining has drawn consideration from numerous specialists. Subsequently, many frequent pattern mining algorithms have been proposed, which incorporate level wise Apriori-based algorithms, tree-based algorithms, hyperlinked cluster structure-based algorithms, and in addition vertical mining algorithms. While these algorithms are well known, they additionally experience the ill effects of a few downsides. To stay away from these disadvantages, we show an option frequent pattern mining (or web mining) algorithm called BW-mine in this paper. An assessment comes about indicate that our recommended algorithm is both space and time-efficient. Besides, to demonstrate the common sense of BWmine, all things considered, applications, we apply BW-mine to find the well known pages on the web, which thus gives the web surfers proposal of pages that may be interested them [13].

Junrui Yang (2016) et al introduce that Frequent itemset mining assumes an imperative part in the mining of different patterns and is in requested in much genuine living. The vast majority of the reviews find visit itemsets from customary exchange databases, in which the substance of every exchange are certainly known and praise. Be that as it may, there are numerous circumstances in which ones are unverifiable, This calls for mining uncertain data. Additionally, surges of dubious information can be created in numerous different circumstances. This prompts to stream mining of unverifiable information. In our paper, we proposed DSUFmine algorithm to manage stream mining of uncertain data[14].

Yi Lin (2016) et al exhibit that It is a noteworthy issue to productively distinguish the much of the frequently occurring patterns in a given dataset, in order to unveil the patterns holed up behind the dataset. This work is motivated from the genuine requests of an elite in-memory frequent-pattern mining technique, with joint optimization over the mining execution and framework strength. While the broadly utilized frequent pattern tree (FP-tree) fills in as a productive approach for frequent-pattern mining, its development method regularly makes it unpleasant for nonvolatile recollections (NVMs). Specifically, the incremental development of FP-tree could create numerous superfluous writes to the NVM and incredibly corrupt the vitality productivity, in light of the fact that NVM composes regularly take additional time and vitality than peruses. To defeat the downsides of FP-tree on NVMs, this paper proposes evergreen FP-tree (EvFP-tree), which incorporates an apathetic counter and a minimum-bit-altered (MBA) encoding plan to make FP-tree agreeable for NVMs. The essential thought of the lethargic counter is to significantly take out the repetitive writes created in FP-tree

development. Then again, the MBA encoding plan is to supplement existing wear-leveling strategies to equally compose every memory cell to expand the NVM lifetime. As checked by experiments, EvFP-tree significantly improves the mining execution and framework lifetime by 28.01% and 82.10% all things considered, individually [15].

G. Lee, U. Yun(2016)propose a correct, proficient calculation for mining questionable continuous examples in view of novel information structures and mining procedures, which can likewise ensure the accuracy of the mining comes about with no false positives. The recently proposed list-based data structures and pruning strategies permit a total arrangement of unverifiable frequent patterns to be mined all the more proficiently without patterns losses. We additionally show that the proposed calculation beats past best in class approaches in both hypothetical and experimental angles. Particularly, we give scientific after effects of execution assessment for different sorts of datasets to shows effectiveness of runtime, memory utilization, and versatility in our method[16].

K.Hemanthakumar, J.Nagamuneiah (2015) Presents handles the issue of deciding Infrequent and weighted Itemsets, i.e. Infrequent weighted item set decide itemsets whose recurrence of events in the analyzed data is not exactly or equivalent to a most extreme edge. To decide Infrequent weighted thing set, two algorithms are stripped infrequent weighted item set (IWI) and Minimal occasional thing set (MIWI). In this review is inspired by the infrequent weighted item sets, as of transactional weighted data indexes to address IWI support amount is characterized as a weighted frequency of events of a data set in the inspected data. Event weights coming about because of the weights identified with items in every transaction and applying a given cost function[17].

Kashish Ara Shakil, Shadma Anis and Mansaf Alam(2015) proposed approach they have utilized WEKA with 10 cross approval to assess information and analyze comes about. Weka has a broad gathering of various machine learning and information mining calculations. In this paper, we have firstly grouped the dengue informational index and after that looked at the different data mining systems in WEKA through Explorer, learning stream and Experimenter interfaces. Besides, keeping in mind the end goal to approve our approach, we have utilized a dengue dataset with 108 instance however WEKA utilized 99 rows and 18 attributes to decide the expectation of disease and their exactness utilizing characterizations of various algorithms to discover the best execution. The fundamental target of this paper is to characterize data and help the clients in separating valuable data from information and effectively recognize a reasonable calculation for a precise prescient model from it. From the

discoveries of this paper it can be reasoned that Naïve Bayes and J48 are the best execution algorithms for characterized exactness since they accomplished most extreme accuracy= 100% with 99 accurately ordered examples, greatest ROC = 1, had least mean absolute error and it required the least time for building this model through Explorer and Knowledge flow results[9].

Durairaj M. what's more, Ranjani V.(2013) focused to analyze comparison of procedures, methodologies and diverse devices and its effect on the health sector. This paper means to make a point by point think about reports of various sorts of data mining applications in the health insurance segment and to lessen the intricacy of the study of the healthcare data transactions. Likewise shows a relative study of various data mining applications, procedures and diverse techniques connected for removing learning from the database created in the health industry. At long last, the current data mining procedures with data mining algorithms and it's application instruments which are more significant for medicinal services[18].

Vadrevu Sree Hari Rao and Mallenahalli Naresh Kumar(2012) proposed another computational intelligence-based technique that forecasts the diagnosis in real time, limiting quantity of false positives and false negatives. Our procedure comprises of three noteworthy parts:

- i. a novel missing worth ascription methodology that can be connected on any dataset comprising of all out (ostensibly) as well as numeric (genuine or number);
- ii. a wrapper-based element choice strategy with genetic search for removing a subset of most persuasive manifestations that can analyze the disease;
- iii. an alternating decision tree strategy that utilizes boosting for producing very precise decision rules.

The predictive models created utilizing our system are observed to be more precise than the cutting edge approaches utilized as a part of the diagnosis of the DF[5].

Deepak Sinwar and Dr. V. S. Dhaka proposed an altered FP-Growth calculation that considers the weights of the characteristics of the datasets. this paper has actualized the idea of Weighted Frequent Patterns by adjusting the notable FP-growth algorithm. The altered algorithm is named as 'Weighted_FP-Growth'. We realize that a few patterns are not frequent by ly huge now and again. Hypothetical examination and experimental work have demonstrated that the adjusted approach can identify the relative significance of properties as far aany stretch of the imagination, but rather they might be sufficients their weights. The characteristics which were less

frequent are currently frequent because of their weights. The execution algorithm as of Weighted_FP-Growth has been contrasted and the FP-growth far as numbers of rules produced in the wake of providing random weights in Weighted_FPGrowth, that Weighted_FPGrowth algorithm beats the conventional FP-Growth as far as number of rules generated[19].

V. ASSOCIATION RULE MINING

Association Rule Mining (ARM) is a settled Data Mining method for finding hidden patterns called Association Rules between various attributes/items of a transaction database (i.e. Market-Basket database). A frequent itemset frequently alludes to an arrangement of things that much of the time seem together in a transaction-based data collection, for example, milk and bread [20]. Numerous algorithms have been proposed to acquire helpful and priceless data from large databases in the exploration literature.

The association rule $A \rightarrow B$ holds in transaction set D with support s, where s is the percentage of transactions in D that contain A B (i.e., the union of sets A and B, or say, both A and B) and with confidence c, where c is the percentage of transactions in D containing A that also contains B. This is taken to be the conditional probability, $P(B|A)$. Decides that fulfill both a minimum support threshold (min_sup) and a minimum confidence threshold (min_conf) are called strong. Association rule mining for the most part a two-stage approaches: to begin with, find all frequent patterns; second, generate association rules through frequent patterns. As a rule, the algorithm which are utilized to produce association rules are either in view of Apriori[21] algorithm or may take after FP-Growth approach [22].

$$\text{Support}(A \rightarrow B) = P(A \cup B)$$

$$\text{Confidence}(A \rightarrow B) = P(B|A) = \frac{P(A \cup B)}{P(A)}$$

Table 1. An example of association rule mining

TID	Items
1	Bread , Milk
2	Beer , Diaper, Bread , Eggs
3	Beer , Coke , Diaper , Milk
4	Beer, Bread , Diaper , Milk
5	Coke , Bread , Diaper, Milk

Association Rule: $X \rightarrow_{s,a} y$

$$\text{Support } s = \frac{\sigma(X \cup y)}{|T|} \quad (s = P(X, y))$$

$$\text{Confidence } \alpha = \frac{\sigma(X \cup Y)}{|\sigma(X)|} \quad (\alpha = P(Y|X))$$

Example:

{Diaper, Milk} =>_{s,α} Beer

$$S = \frac{\sigma(\text{Diaper, Milk, Beer})}{\text{total no. of transactions}} = \frac{2}{5} = 0.4$$

However, these traditional association rule mining (ARM) models accept that things have a similar essentialness without assessing their weight/qualities inside an transaction or inside the entire item space. Be that as it may, this is not generally the situation as a few items might be more essential than others for one or different reasons.

For instance, [wine → salmon, 1%, 80%] might be most essential than [bread → milk, 3%, 80%] despite the fact that the previous holds a lower support. This is on the grounds that items in the rule bring more benefit per unit deal. However, the standard ARM basically disregards this distinction [23]. So as to handle this issue, we have changed the notable association rule mining approach i.e., FP-growth with a specific end goal to incorporate the idea of weights of the attributes. The new algorithm works by considering a weight an incentive for each item (attribute) so that the attribute (s) can mirror their significance[19].

ASSOCIATION RULE MINING ALGORITHMS

A. APRIORI Algorithm

Apriori algorithm is a standout amongst the most robust algorithms to mine the frequent item sets of Boolean association rules. This algorithm is an approach in light of two-phase recurrence, the outline of association rule mining algorithm can be disintegrated into two sub-issues:

- (1) Find all the item sets with the support more noteworthy than the minimum support, which is called frequent item;
- (2) Depending upon the above got frequent set, all the association rules will be created, and for each frequent itemset A_n, all the nonempty subset "a" of A will be found, if the proportion of support (A)/support(a) ≥ min confidence, to produce the association rules A-a. That is, from the association sets acquired in the initial step, to abuse the guidelines with confidence at the very least confidence min confidence the client indicated.

The acknowledgment procedure of the algorithm can be depicted as takes after: First of all, Apriori algorithm explain the frequent sets L with items of 1, then from L to produce the candidate sets C₂ with things of 2, output the

transaction database D, calculate the support to settle L₂, etc., bringing about CK, to scan D and derive LK. Once every now and again set created from the database, strong association rules can be straightforwardly produced. The core algorithm as takes after:

Apriori algorithm called two sub-forms which are Apriori-gen() and the subset(), Apriori-gen() process produces a candidate, then utilize the Apriori property (all non-empty subsets of frequent item sets should likewise be frequent) to remove those candidates of the non-frequent subsets. Once created the majority of the candidates, we will filter the database D, and for every transaction, use the Subset () subprocedure to distinguish all the candidate subsets, and make the cumulative count for each of these candidates. At last, all candidates met the minimum support forms frequent item set L[24].

Pros

It is a simple algorithm to implement.

Cons

It takes more memory, a great deal of space and it will require greater time for the procedure of candidate generation[25].

B. FP-GROWTH Algorithm

Keeping in mind the end goal to defeat the inborn deformities of the Apriori algorithm which deliver countless itemsets and require examining of the database iteratively, Han proposed a frequent pattern growth algorithm (i.e. FP-Growth) in view of Frequent Pattern (FP) tree. The algorithm just needs to examine the database once again. At the primary output of the database, the frequent l-itemset will be created. At the second scan, the l-itemsets are utilized to produce the FP-tree by filtering through infrequent items. The FP-tree contains all the frequent itemsets, so the higher request frequent itemsets can be mined from it. The mining of the FP-tree begins with each frequent pattern of length 1. The conditional pattern base is developed and after that, the (conditional) FP-tree is developed. At that point, the algorithm mines the tree iteratively to create the frequent patterns.

The frequent pattern growth is practiced by interfacing the frequent pattern with the pattern generated by the conditional FP-tree. It utilizes the most repeated item, which gives a superior selectivity. This technique significantly diminishes the cost of seeking. The review on the performance of the FP-tree technique demonstrates that it is proficient and

adaptable to mine both long and short frequent patterns. Coming about frequent patterns are used to form frequent k-itemsets. It is about a request of size quicker than the Apriori algorithm's [26].

Pros

It can be quicker when given the substantial volume of data.

Cons

Regularly more diverse to advance, examination and maintain[25].

a) The updated FP-GROWTH algorithm

By presenting an entropy heuristic and changing the hunt component of the first FP-growth algorithm, we have the accompanying modified FP-growth algorithm. The fundamental contrast between the first and the reconsidered FP-growth algorithms is that the revised one uses the entropy heuristic during the generation of FP-tree and replaces the one-by-one filtering seek system with the bunch search mechanism.

b) WEIGHTED_FPGROWTH

Weighted_FPGrowth is an updated version of FP-growth algorithms. The updated algorithms consider a weight as a parameter to uncover the significance of an attribute. The inspiration driving this alteration is to extract those associations which are infrequent in nature, however some of the time more huge from others. The importance is a general parameter whose value fluctuates from individual to individual. A few characteristics are imperative for some arrangement of people however some are most certainly not. In this altered adaptation we have used arbitrary weight values (from 0.0 to 1.0) using random number generation strategy for Java language.

The working rule of Weighted_FP Growth actualizes an idea of building Frequent-Pattern tree by pruning those attributes/ items from the database whose weights are below the minimum weight threshold and afterward embeddings the rest of the items to the tree and furthermore having more support count than the minimum support threshold. We have reproduced the updated version in Java Language using a few alterations in the already developed version of FPgrowth in WEKA [19].

Algorithm: Weighted_FPGrowth

Input: A transaction dataset D in the form of binary values

Output: A list of weighted association rules

Strategy:

1. Scan the dataset once to discover the support count (m) of every item 'a' of the database.
2. Manufacture the frequent pattern tree (after allocating an arbitrary weight to all attributes) by embeddings just those items of a transaction dataset which fulfills the accompanying conditions:

Condition 1: Weight (Item i) > min_weight
Condition 2: Support_count (Item i) > min_support
3. The FP-Tree is stored by calling FP-growth (FP-tree, invalid).
4. Produce association rules from weighted frequent itemsets generated in step 3.
5. End.

VI. CONCLUSION

In few decades, Data Mining is turning into the most rising field in the healthcare sector in light of the fact that there is a need of proficient and powerful expository technique for searching complicated and valuable data in wellbeing information. In our review, we have talked about data mining in the medical field. Data mining are likewise utilized as a part of dengue fever prediction. Dengue fever is an influenza-like sickness delivered by the chomp of a contaminated mosquito which is quick developing as a noteworthy medical issue. Dengue fever (DF) is a mosquito-endured irresistible infection accompanied around by the infections of the class Togaviridae, subgenus Flavivirus. Additionally, Association rule mining has examined with various – diverse algorithms like - Apriori and FP-Growth.

REFERENCES

- [1] Divya Tomar and Sonali Agarwal, “ A survey on Data Mining approaches for Healthcare”, International Journal of Bio-Science and Bio-Technology Vol.5, No.5 , pp. 241-266, 2013.
- [2] M. Naresh Kumar “Alternating Decision trees for early diagnosis of dengue fever” National Remote Sensing Centre (ISRO), India,, June 6, 2013.
- [3] J. Jain, S.K. Dubey, J. Shrinet, S. Sunil, “Dengue Chikungunya co-infection: A live-in relationship”, Biochemical and Biophysical Research Communications, doi: 10.1016/j.bbrc.2017.02.008, 2017.
- [4] Srinivasa Rao Mutheneni et al. , “Spatial distribution and

- cluster analysis of dengue using self organizing maps in Andhra Pradesh, India, 2011–2013”, *Parasite Epidemiology and Control*, elsevier, 2016
- [5] Vadrevu Sree Hari Rao, Mallenahalli Naresh Kumar, “A New Intelligence-Based Approach for Computer-Aided Diagnosis of Dengue Fever”, *Transactions on Information Technology in Biomedicine*, Vol. 16, No. 1, IEEE , January 2012.
- [6] S.N. Sinha et al. , “Adefovir dipivoxil—A possible regimen for the treatment of dengue virus (DENV) infection”, *Chemometrics and Intelligent Laboratory Systems* 155, Elsevier, pp. 120–127, 2016.
- [7] Farooqi W, Ali S, “A Critical Study of Selected Classification Algorithms for Dengue Fever and Dengue Hemorrhagic Fever” *Frontiers of Information Technology (FIT)*, 11th International Conference on IEEE, 2013.
- [8] Yoon Ling Cheong et al. , “Assessment of land use factors associated with dengue cases in Malaysia using Boosted Regression Trees”, *Spatial and Spatio-temporal Epidemiology* 10, Published by Elsevier, pp. 75–84, 2014.
- [9] Kashish Ara Shakil, Shadma Anis And Mansaf Alam. "Dengue Disease Prediction Using Weka Data Mining Tool".
- [10] Valquíria Aparecida Rosa Duarte, Rita Maria Silva Julia,” Improving NetFeatureMap-based Representation through Frequent Pattern Mining in a Specialized Database”, 28th International Conference on Tools with Artificial Intelligence, IEEE , pp: 954-961, 2016.
- [11] Neha Goyal and S K Jain,” A Comparative Study of Different Frequent Pattern Mining Algorithm For Uncertain Data: A survey”, *International Conference on Computing, Communication and Automation (ICCCA)* IEEE, pp: 183-187, 2016 .
- [12] Md. Badi-Uz-Zaman Shajib Md. Samiullah Chowdhury Farhan Ahmed, Carson K. Leung and Adam G. M. Pazdor,” An Efficient Approach for Mining Frequent Patterns over Uncertain Data Streams”, 28th International Conference on Tools with Artificial Intelligence, IEEE , pp: 979-983, 2016.
- [13] Fan Jiang, Carson K. Leung and Adam G. M. Pazdor,” Web Page Recommendation Based on Bitwise Frequent Pattern Mining”, *International Conference on Web Intelligence, IEEE/WIC/ACM*, pp: 632-635, 2016.
- [14] Junrui Yang¹, Cai Yang², Yanjun Wei³,” Frequent pattern mining algorithm for uncertain data streams based on sliding window”, 8th International Conference on Intelligent Human-Machine Systems and Cybernetics, pp: 265-268, 2016.
- [15] Yi Lin, Po-Chun Huang, Duo Liu, Xiao Zhu, and Liang Liang,” Making In-Memory Frequent Pattern Mining Durable and Energy Efficient”, 45th International Conference on Parallel Processing, pp: 47-56, 2016.
- [16] G. Lee, U. Yun, A new efficient approach for mining uncertain frequent patterns using minimum data structure without false positives, *Future Generation Computer Systems*, <http://dx.doi.org/10.1016/j.future.2016.09.007>, 2016.
- [17] K.Hemanthakumar, J.Nagamuneiah, “Study on Mining Weighted Infrequent Itemsets Using FP Growth”, *International Journal Of Engineering And Computer Science* ISSN:2319-7242 Volume 4 Issue 6 pp.12719-12723, June 2015.
- [18] M. Durairaj, V. Ranjani, “Data Mining Applications In Healthcare Sector: A Study”, *International Journal of Scientific & Technology Research* Volume 2, Issue 10, OCTOBER 2013.
- [19] Deepak Sinwar and Dr. V. S. Dhaka, “Mining Weighted Frequent Patterns using _Weighted_FPGrowth’- A modified FP-Growth”, *Special Conference Issue: National Conference on Cloud Computing & Big Data*.
- [20] W. Jian, L. X. Ming, “An Effective Mining Algorithm for Weighted Association Rules in Communication Networks,” *Journal of Computers*, Vol. 3, pp. 20-27, 2008.
- [21] R. Agrawal, R. Srikant, “Fast Algorithms for Mining Association Rules,” in *Proc. 20th Intl. Conf. on very Large Data Bases (VLDB’94)*, Santiago de Chile, Chile, pp. 487-499, 1994.
- [22] J. Han , J. Pei and Y. Yin ,” Mining frequent patterns without candidate generation”, in *proc. Intl. conference on Management of Data*, pp. 1-12, 2000.
- [23] F. Tao, F. Murtagh and M. Farid, “Weighted Association Rule Mining using weighted support and significance framework,” in *Proc. 9th ACM SIGKDD’ 03*,

Washington, D.C., pp. 661-666, 2003.

- [24] Yanxi Liu, “Study on Application of Apriori Algorithm in Data Mining”, Second International Conference on Computer Modeling and Simulation, 2010.
- [25] S. Muthuselvan ,Dr. K. Soma Sundaram, “ A Survey of Sequence Patterns in Data Mining Techniques”, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, pp. 1807-1815, 2015.
- [26] Lei Wang, Xing-Juan Fan, Xing-Long Liu, Huan Zha, “Mining Data Association Based on a Revised FP-GRowth Algorithm”, Proceedings of the 2012 International Conference on Machine Learning and Cybernetics, IEEE , 15-17 July, 2012.