Novel efficient approach for frequent data set mining of uncertain databases

[1]Sridevi Malipatil, [2]Dr.T Hanumantha Reddy

[1]Assistant Professor, [2]Principal
Department of computer science & Engineering,
RYMEC, Ballari.

***Abstract*: In sensor monitoring systems, location-based services, integration of data etc., the handling of data is basically imprecise. In this research paper, mining of recurrent item sets from uncertain database which is also indeterminate in nature is concentrated. The efficiency of frequent mining of datasets from uncertain database is accomplished by utilizing Adaboost algorithm. The algorithm is utilized to meet the challenges of exponential number of possible worlds with minimum threshold; the algorithm aims to detect the frequent item sets among uncertain database along with filling the missing values. The proposed algorithm will fill the missing values using Adaboost algorithm and also to find frequent item sets.**

***Index Terms*: Uncertain database, frequent data sets**

I. INTRODUCTION

The significant and innovative applications often involve databases which are uncertain. If we consider the instance of the identification of user location carried out by the methods of RFID/GPS could not assure the accuracy, as there is a chance of observational errors [22][28]. The system which involves sensors for monitoring various data parameters such as rain, humidity, temperature etc., may be noisy [17]. The data's involved in hypermarket are used to predict the needs of a customer which is a statistical information [3] [6]. The correlation of the confidence values with output tuples have carried out by the tools of Record Linkage (RL) and Integration on the basis of their matching quality [16]. The confidence values play a significant role in structured information extraction process [7] by adding certain rules from the unstructured data [31]. These are the few examples which narrates the applications demanding huge data's which are uncertain. Hence there arises many researches recently to contend large number of uncertain databases [5],[10], [16], [19], [20], [27]. Also there exist many algorithms researched by many researchers [40].

Table. 1 Example for uncertain database.

Customer	**Items Purchased**
Ram	{Video:1/2}: {food:1}
Shiva	{Clothing:1}, {Video:1/3}; {book2/3}

The example of uncertain database shown in Table.1 in an application which involves online marketing along with the probabilistic information [8]
The purchase history of Ram and Shiva is stored which represents the probable value that the customer may purchase further. The user browser history may be used to analyze the probability values. In that example video:1/2 indicates that Ram has visited 8 times in last week, out of which he has visited 4 times the market place which indicates 50% likelihood of buying videos by ram.
The references [6][10][20][28] include this uncertainty model based on attributes, in which the confidence values and data attributes have associated together, in order to utilize in the instances of GPS/ RFID system's sensor uncertainty and their model location.
Possible World Semantics (PWS) is a technique employed for interpretation of uncertain databases [16]. In PWS, possible worlds are considered as a set of finite instances consisting of set of tuples. In the above example, there are two tuples food and clothing with probability 1 for Ram and Shiva. Besides, it has necessitated correcting all the query evaluation algorithms under PWS, as to applying it for uncertain database, i.e., the results derived through the algorithm must reflect its equality to the query if it is validated on each possible world [16].
The mining and querying using PWS is spontaneous and quite useful, also the technique is costly due to enormous possible worlds containing uncertain database. The data mining using PWS is yet challenging task technically and economically. Hence many researchers have grabbed the attention for efficient data mining [3]. Various clustering algorithms for uncertain database is given in [23] [37], besides, deliberated the designing of decision tree classifiers for uncertain data [32].
Throughout this research, the identification of frequent item sets (the frequent occurrence of simultaneous sets of attribute values in tuples) have accomplished using two algorithms, particularly in two uncertainty models, i.e., attribute

uncertainty (see. Fig.1), and tuple uncertainty, in which each tuple has allied to a probability, concerning the indication of its existence [15] [16] [19] [27] [34].
The frequent item set is one which replicates the confidence value obtained through uncertain data and sited on the outcomes of mining. It has observed from the Probabilistic Frequent Item set (PFI) is known to be the support probability mass function(s-pmf) for PFI that has represented as the quantity of tuples (the count of support) which includes item set. It gives the collection of semantic worlds defining a (disparate) support count for provided item set. The value of Probability is 1 for the item set which occur twice for all possible worlds.
The Probabilities of the occurrences of these PFIs is stored for mining the frequent patterns from every possible world [41][42]. The complication in storing the PFIs is that it has exponential number of possible worlds. To mitigate these issues, various algorithms have been researched by many researchers, by which PFIs are retrieved without instantiating all possible worlds [6],[30],[35]. Nevertheless, it has noted that the obtainment of empirical outcomes needs long duration to accomplish, for instance, the dynamic programming algorithm [6] requires the duration of 30.1 hrs to identify the entire PFIs among 300k real data set.
This model-based calculation solely needs lesser time to check a PFI, along these lines increasingly appropriate for enormous databases. The proposed calculation has exhibited in what way it needs to be utilized in mine edge-based PFIs, in which its probabilities of being genuine regular thing sets are bigger when compared to a few clients characterized edge [6]. Besides, the entire PFIs have been identified solely in9.2 seconds using our calculation [33], which is four sets of extents quicker compared to the technique in [6] mining advancing databases. Additionally, we scrutinize the significant issue of keeping up digging results for changing, or developing the databases. Here, we stated that the kind of advancing information, concerning the adding/inclusion of a bunch of tuples to the database. We have contemplated the applications, in which the tuple inclusion is common. For instance, a GPS framework might have needed to deal with area esteems because of the enrollment of another client; in an online commercial center application, data about new buy exchanges might have annexed to the database for further examination.
Here analysis is on how the model-based calculation that finds estimated PFIs that deals with developing information. When the difference in the database is little, running our steady mining calculations on data's newer is a lot quicker than discovering PFIs on data's older without any preparation. In an investigation on a genuine informational index, the model-based, gradual mining calculation tends to a fivefold exhibition improvement across its non-incremental partner.
To outline, to build up a model-based calculation that is able to decrease the measure of exertion of checking the database for mining edge based PFIs. In addition, here two steady mining calculations for removing correct as well as rough PFIs. Every one of our calculations can bolster not only the characteristic, but also the tuple vulnerability models. Investigations on genuine, as well as engineered informational indexes, have uncovered that these techniques significantly enhance the presentation of PFI disclosure, accompanying high level of precision.
The remaining part of this paper is sorted out as pursues: Section 2 confers the related works. Section 3 deals with problem definition. Section4 characterizes the algorithm used to be examined. Section 5 portrays results. Section 6 deals with the conclusion of the paper.

II. LITERATURE REVIEW OF RELATEDWORK

Extracting frequent item sets is a significant issue in information mining, and is likewise the initial phase of inferring affiliation rules [4]. Thus, numerous productive itemset mining calculations have been suggested, particularly Apriori [4] and FP-development [18]. Despite these calculations function admirably for databases with exact qualities, it is unclear about the way of utilization in mining probabilistic information. At this point, we create calculations for separating continuous thing sets from dubious databases. In spite of the fact that our calculations are created dependent on the Apriori outline work, they are still able to be taken to assist in different calculations (e.g., FP-development) for dealing with unsure information.
In case of uncertain databases, Aggarwal et al. [2] and Chui et al. [14] created effective continuous example mining calculations dependent on the normal bolster checks of the examples. Be that as it may, Bernecker et al. [6], Sun et al. [30], and Yiu et al. [35] establish that the utilization of projected help might reduce significant examples missing. Consequently, they suggested to register the likelihood that an example is visit, and presented the thought of PFI. Dynamic-programming-based arrangements were created to recover PFIs from quality dubious databases [6],[44]. In any case, their calculations register precise probabilities, and confirm that an item set is a PFI in Oðn2þ time. Our model-based calculations evade the utilization of dynamic programming, and can check a PFI a lot quicker. In [35], estimated calculations for getting edge based PFIs from tuple-dubious information streams have created. While Zhang et al. [35] just thought about the derivation of singletons, (that is to say sets of single items), our answer finds designs with multiple things. As of late, Sun et al. [30] built up a precise limit based PFI mining calculation. In any case, it doesn't bolster quality questionable information, which has considered in this study. Besides, [33] inspects about the model-based methodology for mining PFIs, which is a fundamental adaptation of this work, where we observed the method to stretch out this calculation for helping the mining of developing information.
Different deals with the recovery of continuous designs from uncertain information include: [9], which concentrated inexact successive designs on uproarious information; [24], which inspected mining structure is enlivened by FUP. In [12], the FUP2 calculation was created to deal with both expansion and cancellation of tuples. ZIGZAG [1] likewise inspects the proficient support of highest visit thing sets for databases, those which are always showing signs of change. In [13], an information structure, called CATS Tree, was acquainted with keep up visit thing sets in developing databases. Can Tree [25],[43] is known to be another structure that master minds tree hubs in a request that isn't influenced by the variations of thing recurrence. The information structure has utilized to help mining on an evolving database. As we know, the inspections of keeping up visit thing sets in developing

dubious databases have not carried out previously. Hence, a novel steady digging calculation have proposed for correct, as well as for rough PFI revelation, which is capable of likewise bolster characteristic and tuple, vulnerability models [45].
The execution of significant work in PFI mining have represented in Table 1, where "Static Algorithms" allude to calculations that don't deal with changes in database. Thus, whichever adjustment in the database requires a total implementation of these calculations.

Table.2 PFA mining algorithms

Uncertainity model	Static algorithms	Incremental algorithms
Attribute	Exact [6] Approx. [√]	Exact [√] Approx. [√]
Tuple	Exact [3 0] Approx. (singleton)[3 5] Approx. (multiple items) [√]	Exact [√] Approx. [√]

III. PROBLEM DEFIITION

The issue in Knowledge Discovery is of prompting the association rule [11] with Minimum Support Threshold. The exemplary methodology finds the association rules among the every now and again observed patterns in a lot of transaction records. Experts have generally proposed models to find association administers above a minimum support threshold. While scanning for association rules among frequent patterns is significant, sometimes, detailing the association decides among the items that fall beneath a base help limit are likewise significant.
Our point is to plan a system which finds a solution for filling missing values. With the utilization of Minimum Threshold, adequate and enormous number of rules has been found which produce the correct choices. Additionally, the storage space has been broadened because of all the more examining procedure. The proposed calculations will be intended to locate the successful standard by exploiting the properties of rules and will offer an adjusted strategy for information mining that a defeat the breaking point related with existing techniques and empowers the finding of association rules.
In existing framework apriori calculation gets ended when the regular itemsets can't be broadened further. The favorable position is that numerous outputs are produced for candidate sets. The burden is that the execution time is more as burned through in creating candidates s without fail, it additionally needs more search space and computational expense is excessively high.

IV. PROPOSED SYSTEM

In proposed framework, the AdaBoost performs well with weak learners. Moreover, this model accomplishes great exactness over irregular possibility on a characterization issue. The normal calculations with AdaBoostare utilized in frequent mining.

1. Upload Dataset

1.1 Gathered data about patients under nursing care with features: Name of the Nursing home, Category, Nursing home description, owned by, address etc.
1.2 Upload Medical Residential and Nursing with csv format.
1.3 There are some missing values in columns like provided by and email.
Separate missing value rows from the full dataset.

2. Preprocessing

After uploading the dataset, we have to preprocess the data. It will split the three categories like missing data, empty field row, and original data. We have two methods for preprocessing. One is to remove missing data and other one is to fill the missing data. Here we have proposed a method to fill missing values.
Use Adaboost algorithm on the basis of category and nursing home description to fill missing values.

1) Data point’s weights must be set to a value of (1/number of data points).
2) Decision tree to be trained.
3) Weighted error rate ‘e’ = (number of false predictions/overall predictions). Calculate ‘e’ for the decision tree.
4) Calculate weight of decision tree=learning rate * log(1-e)/e. Decision power is inversely proportional to the weighted error.
5) Weights of erroneously classified points need to be revised. New weight for erroneous point=old weight*np.exp (weight of tree).
6) Until all tree training is completed repeat step one.
7) Final prediction to be done

3. Classification

There are plenty of researches were undergone for frequent item set mining [36]. Here will classify and find the frequent item set based on the Residential and Nursing dataset. Using Apriori Algorithm to find frequent itemset based on minimum support count = 2.

4. Flowchart for Proposed System

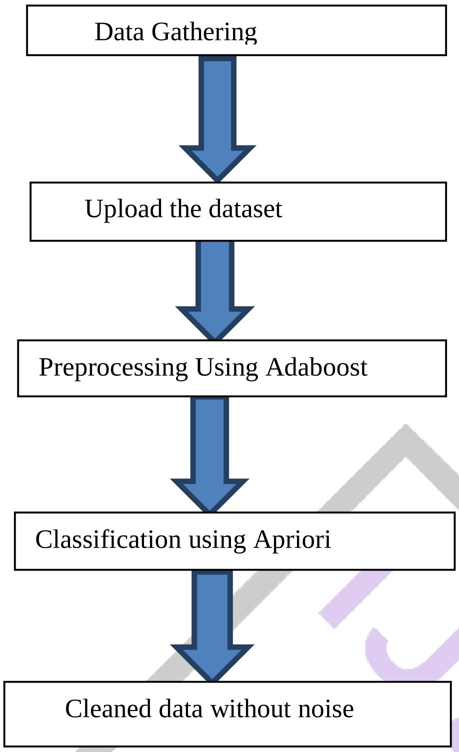


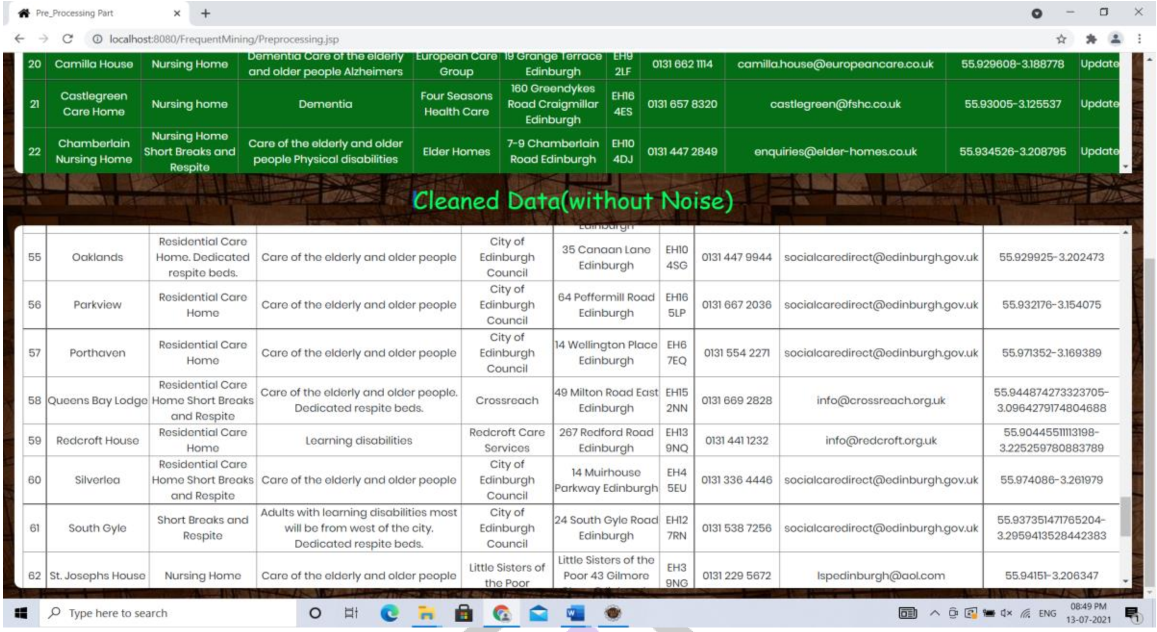
Fig.1 Flowchart for proposed system

V. RESULT

The residential and nursing dataset gathered has missing values, screen shot of the data collected is shown in figure.4. In this figure.4, missing data is highlighted in red colour.

Fig.2 snapshot of missing values in the database

After pre-processing using Adaboost algorithm, screen shot of cleaned data without noise is shown in figure 5. In this figure, cleaned data is highlighted with white colour.



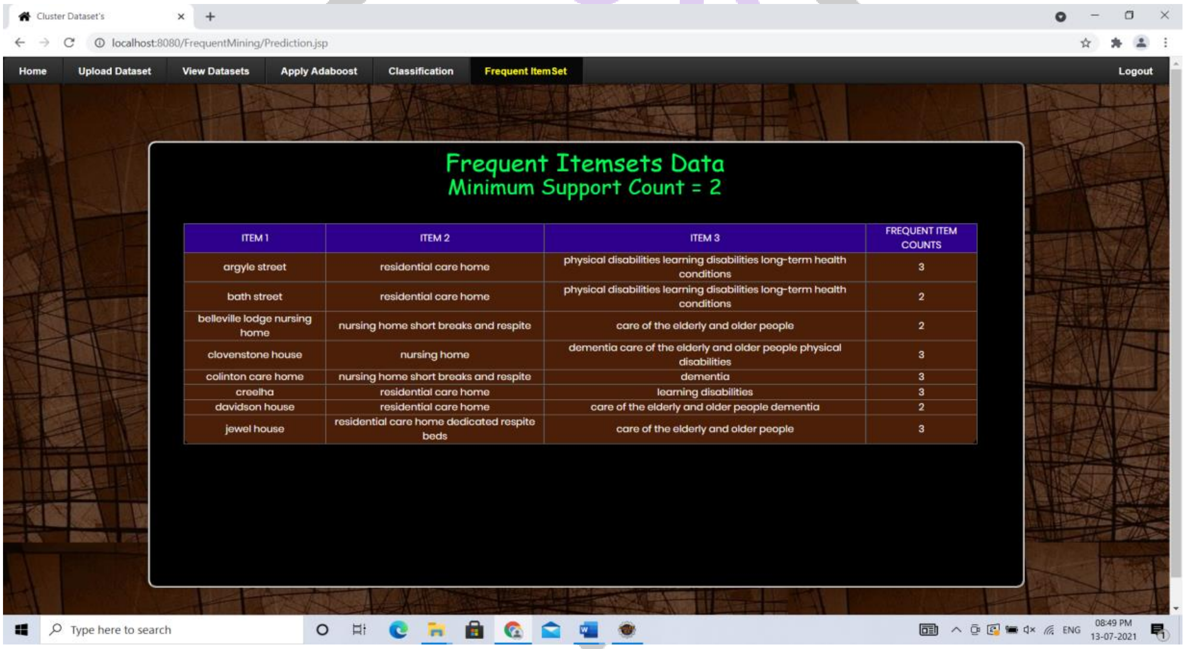
20	Camilla House	Nursing Home	Dementia Care of the elderly and older people Alzheimers	European Care Group	19 Grange Terrace Edinburgh	EH9 2LF	0131 662 1114	camilla.house@europeancare.co.uk	55.929608-3.188778	Update
21	Castlegreen Care Home	Nursing home	Dementia	Four Seasons Health Care	160 Greendykes Road Craigmillar Edinburgh	EH16 4ES	0131 657 8320	castlegreen@fshc.co.uk	55.93005-3.125537	Update
22	Chamberlain Nursing Home	Nursing Home Short Breaks and Respite	Care of the elderly and older people Physical disabilities	Elder Homes	7-9 Chamberlain Road Edinburgh	EH10 4DJ	0131 447 2849	enquiries@elder-homes.co.uk	55.934526-3.208795	Update

Cleaned Data(without Noise)

55	Oaklands	Residential Care Home. Dedicated respite beds.	Care of the elderly and older people	City of Edinburgh Council	35 Canaan Lane Edinburgh	EH10 4SG	0131 447 9944	socialcaredirect@edinburgh.gov.uk	55.929925-3.202473
56	Parkview	Residential Care Home	Care of the elderly and older people	City of Edinburgh Council	64 Peffermill Road Edinburgh	EH16 5LP	0131 667 2036	socialcaredirect@edinburgh.gov.uk	55.932176-3.154075
57	Porthaven	Residential Care Home	Care of the elderly and older people	City of Edinburgh Council	14 Wellington Place Edinburgh	EH6 7EQ	0131 554 2271	socialcaredirect@edinburgh.gov.uk	55.971352-3.169389
58	Queens Bay Lodge	Residential Care Home Short Breaks and Respite	Care of the elderly and older people. Dedicated respite beds.	Crossreach	49 Milton Road East Edinburgh	EH15 2NN	0131 669 2828	info@crossreach.org.uk	55.944874273323705-3.0964279174804688
59	Redcroft House	Residential Care Home	Learning disabilities	Redcroft Care Services	267 Redford Road Edinburgh	EH13 9NQ	0131 441 1232	info@redcroft.org.uk	55.90445511113198-3.225259780883789
60	Silverlea	Residential Care Home Short Breaks and Respite	Care of the elderly and older people	City of Edinburgh Council	14 Muirhouse Parkway Edinburgh	EH4 5EU	0131 336 4446	socialcaredirect@edinburgh.gov.uk	55.974086-3.261979
61	South Gyle	Short Breaks and Respite	Adults with learning disabilities most will be from west of the city. Dedicated respite beds.	City of Edinburgh Council	24 South Gyle Road Edinburgh	EH12 7RN	0131 538 7256	socialcaredirect@edinburgh.gov.uk	55.937351471765204-3.2959413528442383
62	St. Josephs House	Nursing Home	Care of the elderly and older people	Little Sisters of the Poor	Little Sisters of the Poor 43 Gilmore	EH3 9NG	0131 229 5672	lspedinburgh@aol.com	55.94151-3.206347

Fig.3 snapshot of cleaned data without noise

After classification, the below figure. 4 shows the screen shot of frequent item sets in the database using Apriori algorithm.



Frequent Itemsets Data
Minimum Support Count = 2

ITEM 1	ITEM 2	ITEM 3	FREQUENT ITEM COUNTS
argyle street	residential care home	physical disabilities learning disabilities long-term health conditions	3
bath street	residential care home	physical disabilities learning disabilities long-term health conditions	2
belleville lodge nursing home	nursing home short breaks and respite	care of the elderly and older people	2
clovenstone house	nursing home	dementia care of the elderly and older people physical disabilities	3
colinton care home	nursing home short breaks and respite	dementia	3
creelha	residential care home	learning disabilities	3
davidson house	residential care home	care of the elderly and older people dementia	2
jewel house	residential care home dedicated respite beds	care of the elderly and older people	3

Fig.4 snapshot of finding frequent item sets

VI. CONCLUSION

In this paper we propose a System to address the issue in finding missing values in an uncertain database. Specialists have generally proposed models to find association governs over a base help limit. While searching for association rules among the items is significant, now and again, detailing the association decides among the things that fall beneath a base help limit are likewise significant. Our point is to structure a system which finds such sort of decides that are depending on Minimum Support Threshold. With the utilization of Minimum Threshold, fragmented and enormous number of rules has been found which produce the correct choices.

REFERENCES

[1] A. Veloso, W. MeiraJr. M.deCarvalho, B. Possas, S. Parthasarathy, and M.J. Zaki, "Mining Frequent Itemset in Evolving Databases," *Proc. Second SIAM Int'l Conf. Data Mining (SDM)*,2002.

[2] Aggarwal, Y. Li, J. Wang, and J. Wang, "Frequent Pattern Mining with Uncertain Data," *Proc. 15th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD)*,2009.

[3] C Aggarwal and P. Yu, "A Survey of Uncertain Data Algorithms and Applications," IEEE

TransKnowledgeandDataEng., vol.21, no. 5, pp. 609-623, May2009.
[4] R. Agrawal, T. Imieliński and A. Swami, “Mining Association Rules between Sets of Items in Large Databases,” *Proc. ACM SIGMOD Int’l Conf. Management of Data,*1993.
[5] O. Benjelloun, A.D. Sarma, A. Halevy, and J. Widom, “ULDBs: Databases with Uncertainty and Lineage,” *Proc. 32nd Int’l Conf. Very Large Data Bases (VLDB),*2006.
[6] T. Bernecker, H. Kriegel, M. Renz, F. Verhein, and A. Zuefle, “Probabilistic Frequent Itemset Mining in Uncertain Databases,” *Proc.15thACMSIGKDDInt’lConf.KnowledgeDiscoveryandData Mining (KDD),*2009.
[7] T. Bernecker, H. Kriegel, M. Renz, F. Verhein, and A. Zuefle, “Probabilistic Frequent Itemset Mining in Uncertain Databases,” *Proc.15thACMSIGKDDInt’lConf.KnowledgeDiscoveryandData Mining (KDD),*2009.
[8] C.J. van Rijsbergen, *Information Retrieval.* Butterworth,1979
[9] L.L. Cam, “An Approximation Theorem for the Poisson Binomial Distribution,” *Pacific J. Math.*, vol. 10, pp. 1181-1197,1960
[10] H. Cheng, p. Yu, and J. Han, “Approximate Frequent Itemset Mining in the Presence of Random Noise,” Proc. Soft Computing for knowledge Discovery and Data Mining, pp.363-389,2008
[11] R. Cheng, D. Kalashnikov, and S. Prabhakar, “Evaluating Probabilistic Queries over Imprecise Data,” Proc. ACM SIGMOD Int’l Conf. Management of Data,2003.
[12] D. Cheung, J. Han, V. Ng, and C. Wong, “Maintenance of Discovered Association Rules in Large Databases: An Incremental Updating Technique, “Proc.12th Int’l Conf. Data Eng. (ICDE), 1996
[13] D. Cheung, S.D. Lee, and B. Kao, “A General Incremental Technique for Maintaining Discovered Association Rules,” *Proc. Fifth Int’l Conf. Database Systems for Advanced Applications (DASFAA),*1997.
[14] W. Cheung and O.R. Za¨ıane, “Incremental Mining of Frequent Patterns without Candidate Generation or Support Constraint,” Proc.SeventhInt’lDatabaseEng. and Applications Symp (IDEAS),2003
[15] C.K.Chui, B.Kao and E.Hung, “Mining Frequent Itemset from Uncertain Data, “Proc.11thPacific-AsiaConf.AdvancsinKnowledge Discovery and Data Mining
[16] B. Kao, and E. Hung, “Mining Frequent Itemset from UncertainData,”*Proc.11thPacific-AsiaConf.AdvancesinKnowledge Discovery and Data Mining (PAKDD),*2007.
[17] G. Cormode and M. Garofalakis, “Sketching Probabilistic Data Streams,” *Proc. ACM SIGMOD Int’l Conf. Management of Data*, 2007
[18] N. Dalvi and D. Suciu, “Efficient Query Evaluation on Probabilistic Databases,” *Proc. 13th Int’l Conf. Very Large Data Bases (VLDB),*2004.
[19] A. Deshpande, C. Guestrin, S. Madden, J. Hellerstein, and W. Hong, “Model-Driven Data Acquisition in Sensor Networks,” *Proc. 13th Int’l Conf. Very Large Data Bases (VLDB),*2004.
[20] J. Han, J. Pei, and Y. Yin, “Mining Frequent Patterns without Candidate Generation,” *Proc. ACM SIGMOD Int’l Conf. Management of Data,*2000.
[21] J. Huang “MayBMS”: A Probabilistic Database Management System, “Proc.35th ACMSIGMOD Int’l Conf. Management of Data, 2009.
[22] R. Jampani, L. Perez, M. Wu, F. Xu, C. Jermaine, and P. Haas, “MCDB: A Monte Carlo Approach to Managing Uncertain Data,” *Proc. ACM SIGMOD Int’l Conf. Management of Data,*2008.
[23] J. Ren, S.D. Lee, X. Chen, B. Kao, R. Cheng, and D.W. Cheung, “Naive Bayes Classification of Uncertain Data,” *Proc. IEEE Ninth Int’l Conf. Data Mining (ICDM),* 2009.
[24] N. Khoussainova, M. Balazinska, and D. Suciu, “Towards Correcting Input Data Errors Probabilistically using Integrity Constraint,” Proc.FifthACMInt’lWorkshopDataEng. *for Wireless and Mobile Access (MobiDE), 2006*
[25] H. Kriegel and M. Pfeifle, “Density-Based Clustering of Uncertain Data,” *Proc.ACMSIGKDDInt’lConf. Knowledge Discovery in Data Mining (KDD),*2005.
[26] C.Kuok, A. Fu and M. Wong, “Mining Fuzzy Association Rules in Databases, “SIGMOD Record, vol.27, no 1, pp.41-46,1998.
[27] C.K.-S. Leung, Q.I. Khan, and T. Hoque, “Cantree: A Tree Structure for Efficient Incremental Mining of Frequent Patterns,” *Proc. IEEE Fifth Int’l Conf. Data Mining (ICDM),*2005.
[28] A. Lu, Y. Ke, J. Cheng, and W. Ng, “Mining Vague Association Rules,” *Proc. 12th Int’l Conf. Database Systems for Advanced Applications (DASFAA),*2007.
[29] M. Mutsuzaki, “Trio-One: Layering Uncertainty and Lineage on a Conventional DBMS,” *Proc. Third Biennial Conf. Innovative Data Systems Research (CIDR),*2007.
[30] P. Sistla, O. Wolfson, S. Chamberlain, and S. Dao, “Querying the Uncertain Position of Moving Objects,” *Temporal Databases: Research and Practice,* Springer Verlag,1998.
[31] C. Stein, Approximate Computation of Expectations, Lecture Notes Monograph Series, vol. 7, Inst. of Math. Statistics, 1986.
[32] L. Sun, R. Cheng, D.W. Cheung, and J. Cheng, “Mining Uncertain Data with Probabilistic Guarantees,” *Proc. 16th ACM SIGKDD Int’l Conf. Knowledge Discovery and Data Mining,*2010.
[33] T. Jayram et al., “Avatar Information Extraction System,” *IEEE Data Eng. Bull.*, vol. 29, no. 1, pp. 40-48, Mar.2006.
[34] S. Tsang, B. Kao, K.Y. Yip, W.-S. Ho, and S.D. Lee., “Decision Trees for Uncertain Data,” *Proc. IEEE Int’l Conf. Data Eng.(ICDE),* 2009.
[35] L. Wang, R. Cheng, S.D. Lee, and D. Cheung, “Accelerating Probabilistic Frequent Itemset Mining: A Model-Based

Approach," *Proc. 19th ACM Int'l Conf. Information and Knowledge Management (CIKM)*,2010.
[36] M. Yiu, N. Mamoulis, X. Dai, Y. Tao, and M. Vaitis, "Efficient Evaluation of Probabilistic Advanced Spatial Queries on Existentially Uncertain Data," *IEEE Trans Knowledge and Data Eng.*, *vol.*21,no.9,pp.108-122,Jan.2009.
[37] Q. Zhang, F. Li, and K. Yi, "Finding Frequent Items in Probabilistic Data," *Proc. ACM SIGMOD Int'l Conf. Management of Data*,2008.
[38] Daniele Apiletti, Elena Baralis, Tania Cerquitelli, Paolo Garza, Fabio Pulvirenti ∗, Luca Venturini, "Frequent Itemset Mining for Big Data: A Comparative Analysis", Big Data Research, Elsevier Journal, Volume9, pp no 67-83,2017.
[39] Sagar Bhise, Prof. Sweta Kale, "Efficient Algorithms to find Frequent Itemset Using Data Mining" International Research Journal of Engineering and Technology (IRJET), Volume 04, Issue 06, June -2017.
[40] Mohamed A. Gawwad Mona F. Ahmed, Magda B. Fayek, "Frequent Itemset Mining for Big Data Using Greatest Common Divisor Technique" Data Science Journal,16: 25, pp. 1–10, DOI: https://doi.org/10.5334/dsj-2017-025.
[41] MiSamar Wazir M. M. Sufyan Beg Tanvir Ahmadning, "Weighted Frequent Item Set Mining Using OWA on Uncertain Transactional Database" Proceedings of GUCON 2019, Data Communication and Networks Journal, Springer, Part of the Advances in Intelligent Systems and Computing book series (AISC, volume 1049), pp 183-193.
[42] Chee, C., Jaafar, J., Aziz, I.A. et al. Algorithms for frequent itemset mining: a literature review. Artif Intell Rev 52, 2603–2621 (2019). https://doi.org/10.1007/s10462-018-9629-z
[43] Maylawati D S A, Ramdhani M A, Rahman A and Darmalaksana W, Incremental technique with set of frequent word item sets for mining large Indonesian text data in *2017 5th International Conference on Cyber and IT Service Management CITSM 2017*
[44] Md. Abed Hasan, Naima Hassan, Md. Hasibuzzaman, "A Scalable Approach for Improving Implementation of a Frequent Pattern Mining Algorithm using MapReduce Programming" 5th International Conference on Science in Information Technology (ICSITech), 2019.
[45] Xiang li, Jiaxuan li, Philippe fournier-viger, m. Saqib Nawaz, Jie yao, jerry chun-wei lin, "Mining Productive Itemset in Dynamic Databases", IEEE access volume 8, 2020.