



International Journal of Intellectual Advancements and Research in Engineering Computations

AN EFFICIENT DISCOVERY OF HIGH UTILITY ITEMSETS FROM TRANSACTIONAL DATABASE

¹B.Sangameshwari, ²P. Uma

ABSTRACT

There are gathering of methods, methodologies and novel zones of the examination which are important and stamped as the crucial field of data mining Advancements. Different MNC's and unfathomable affiliations are worked in better places of the interesting nations. Each one spot of operation may make far reaching volumes of data. Corporate manager oblige access from every such source and take pivotal choices .The data movement center is utilized inside the essential business respect by enhancing the sufficiency of administrative choice making. In a faulty and to a great degree intense nature's turf, the estimation of essential data structures, for case, these are easily seen however in today the earth, sufficiency or rate is by all record by all record by all account not the only key for forcefulness. This sort of colossal measure of data's are accessible as tera- to peta-bytes which has without a doubt changed in the extents of science and building. To take a gander at, manage and settle on a choice of such sort of huge measure of data that oblige structures called the information mining which will changing in different fields. This paper presents degree of the information mining which will obliging in the business enclosure.

Index Terms— Data mining, Business Intelligence, Data mining Methods, Data mining applications, Data Mining Tools, Educational Data Mining, Knowledge Discovery.

I. INTRODUCTION

In the 21st century the people are utilized within the diverse innovations to sufficient in the general public. Every single day the people are utilizing the unlimited data and these data are in the diverse fields It may be in the form of documents, may be graphical formats, may be the video, may be records(differing exhibit).As the data are accessible in the distinctive arrangements so that the best possible move to be made. To examine these data as well as take a decent choice and keep up the data. As and when the client will obliged the data ought to be recovered from the database and settle on the better choice. This system is really called as a data mining or Knowledge Hub or basically KDD (Knowledge Discovery in Database).the critical reason that pulled in a lot of consideration in data innovation the revelation of valuable data from extensive accumulations of data

industry towards field of "Data mining" is because of the impression of "we are data rich however data poor". There is immense volume of data yet barely ready to turn them into helpful data and learning for managerial choice making in business.

To create data it requires enormous accumulation of data. It may be diverse arrangements like sound/feature, numbers, content, figures, hyper text groups. To exploit data, the data recovery is basically insufficient, it obliges an instrument for programmed rundown of data, extraction of the quintessence of data put away, and the finding of examples in crude data. With the huge measure of data put away in documents, databases, and different archives, it is progressively essential, to create influential device for dissection and understanding of such data and for the extraction of fascinating learning that could help in choice making. The main reply to all above is 'Data Mining'. Data mining is the extraction of concealed

Author for Correspondence:

¹Department of Computer Science and Engineering, Nandha Engineering College, Erode, Tamilnadu, India.
Email:sanmaheshri@gmail.com

²Department of Computer Science and Engineering, Nandha Engineering College, Erode, Tamilnadu, India.
Email:umamurthi@gmail.com

prescient data from vast databases; it is a compelling innovation with extraordinary potential to help associations concentrate on the most vital data in their data stockrooms. Data mining devices foresee future patterns and practices, helps associations to make proactive information driven choices. The mechanized, prospective examines offered by data mining move past the breaks down of past occasions gave by prospective apparatuses normal of choice help supportive networks. Data mining apparatuses can address the inquiries that customarily were excessively time intensive to purpose. They get ready databases for discovering concealed examples, discovering prescient data that specialists may miss in light of the fact that it lies outside their desires.

Data mining, prominently known as Knowledge Discovery in Databases (KDD), it is then on trifling extraction of verifiable, formerly obscure and possibly helpful data from data in databases. It is really the methodology of discovering the shrouded data/example of the storehouses. The improvement and application of data mining calculations obliges the utilization of effective programming instruments. This paper[1] is composed as takes after: the first segment verifiable improvement and state-of-the-symbolization highlights the recorded advancement of data mining programming until present; the criteria to analyse data mining programming are clarified in the second area criteria for analysing data mining programming. . In this paper[2], can open these discovery DM demonstrates by utilizing a novel visualization approach that is focused around an affectability investigation (SA) technique. Specifically, can propose a worldwide SA (GSA), which develops the appropriateness of past SA routines (e.g. To characterization errands), and a few visualization Techniques (e.g. Variable impact trademark bend), for evaluating data significance and impacts on the model's reactions. A set of true research endeavours has been led to test the structures, with some of them informed in this paper [3].

The most ordinary/regular undertakings in the instructive environment that have been determined through data mining systems lastly the absolute most guaranteeing future lines of examination are discussed[9]. This paper [10] presents a survey of — and order plan for — the

writing on the application of data mining systems for the recognition of budgetary extortion. This study [11] examines the online inquiries and visit messages consequently recorded by a Live Video Streaming (LVS) framework utilizing data mining and content mining strategies. The notion investigation can be found, since an accepted strategy for shopper objection identification, is lacking for discovering, ordering, and prioritizing vehicle deformities examined in online gatherings, and portray and assess another methodology and choice help supportive network for car imperfection distinguishing proof and prioritization.

II. RELATED WORKS

Data is an indispensable crude material of the data economy, much as coal and iron mineral were in the Industrial Revolution. Anyway the business world is simply starting to figure out how to process everything. In the previous 10–15 years, data mining has turned into an engineering in it right, is settled additionally in business intelligence (BI), and keeps on exhibiting relentlessly expanding imperativeness in innovation and life sciences parts. For instance, data mining was a key component supporting methodological leaps forward in heredity. It is a guaranteeing innovation for future fields, for example, content mining and semantic hunt engines, adapting in independent frameworks as with humanoid robots and autos, chemo informatics and others [1].

A novel visualization methodology focused around a Sensitivity Analysis (SA), which is a basic strategy that measures the impacts on the yield of a given model when the inputs are shifted through their scope of values [2]. Immediate mining for discriminative examples has been highlighted, for example, in Harmony, model-based pursuit tree, and rising difference designs. Consolidated mining helps this classification as well. The post analysis and post mining of educated examples is a generally utilized methodology, case in point, to prune tenets, lessen repetition, and compress scholarly runs the show. Not quite the same as post analysis-based systems, the majority of the joined together examples introduced [3].

A procedure administration framework (PrMs) gives non specific methodology help works

and takes into consideration differentiating methodology rationale from application code. For this reason, the procedure rationale must be unequivocally characterized focused around the demonstrating examples gave by a methodology meta model. At runtime the PrMs then coordinates the methodologies as per the characterized rationale. For every business methodology to be backed, a procedure sort spoke to by a procedure model *S* must be characterized. In this paper, a methodology model is spoken to as controlled chart, which contains a set of hubs – either speaking to process steps (i.e., exercises) or control connectors (e.g, And-/Xor-Split) – and a set of control edges between them. The last point out priority and also circle retrogressive relations. Besides, it can assume that process models are square structured [4].

Model transparency identifies with human capacity to comprehend what the model comprises of, heading preferably to the capability to apply it to new perceptions (which may term transportability). On the off chance that a model is transparent, it could be transported. A few models have reliably turned out to be solid in their capability to fit data, for example, neural system models, however to have low transparency or transportability [6]. The current writing on hostile to segregation in software engineering essentially explains on data mining models and related strategies. A few recommendations are turned to the disclosure and measure of segregation. Others manage the avoidance of discrimination [8]. There are numerous applications or undertakings in instructive situations that have been determined through DM. Case in point, Baker recommends four key ranges of use for EDM: enhancing learner models, enhancing space models, contemplating the pedagogical backing gave by learning programming, experimental examination into learning and learners; and five methodologies/routines: forecast, bunching, relationship mining, refining of data for human judgment and disclosure with models[9].the writing inquiry focused around the descriptors "fiscal extortion," "data mining" and "business intelligence." The utilization of Boolean articulations to apply these terms to a pursuit of online databases, which initially delivered roughly 1200 articles. The survey and order procedure was deliberately and freely confirmed by the co-writers,

and just articles that were identified with data mining and FFD were included [10].

As indicated by the instructive data mining group site, instructive data mining (EDM) is characterized to be "a developing order, concerned with creating systems for investigating the interesting sorts of data that originate from instructive settings, and utilizing those strategies to better comprehend understudies, and the settings which they learn in." Furthermore, a few heading EDM specialists (Baker, 2009; Baker & Yacef, 2009; Romero & Ventura, 2010) arrange work in EDM into a couple of classifications, for example, facts and visualization, forecast (grouping, relapse, and thickness estimation), bunching, relationship mining, outlier discoveries, and content mining[12].

In the notion dissection writing, it is assumed that vigorously negative postings (protestations) will be demonstrative of item deserts. Then again, whether this assumption – that negative slant predicts imperfection presence – is valid for auto imperfections has not been tried in earlier research [13]. Temporal Abstraction (TA) is a manmade brainpower procedure, which incorporates area learning into the data dissection process. TA diagrams the evolutionary procedure of fleeting data through aqualitative presentation mode, for example, level movements, times of strength and trends [14]. A Web based business application programming model of the framework portrays different segments of the framework and how they are interfaced for mining procedures utilized for concentrating significant data. The application building design model demonstrates that, a guide of different data and practical modules[17].

III. MINING APPROACH FOR BUSINESS INTELLIGENCE

An effective Building Applications is one which can enhances bits of knowledge of business intelligence by offering results among different modules currently compelling choice making by utilizing essential mining methods as a device. In many tools: decision trees, clustering, regression, data cleaning, data filtering, feature extraction, principal component analysis, factor analysis, advanced feature evaluation and selection,

computation of similarities, artificial neural networks, model cross validation, and statistical relevance tests. In some tools: fuzzy classification, association learning and mining frequent item sets, independent component analysis, bootstrapping, complexity measures, model fusion, support vector machines, k-nearest-neighbour methods, Bayesian networks, and learning of crisp rules. Comparison of DM tools based on their features and approaches along with the mining techniques they can perform is represented in table 1.

VISUALIZATION: The fundamental destination of data visualization is the general thought regarding the data mining model. In data mining the majority of the times are recovering the data from the vaults which are in the concealed structure. This is the troublesome task for a client. So this visualization of the data mining model helps us to give most extreme levels of understanding and trust. Since the client does not know beforehand what the data mining methodology has found, it is a much greater jump to take the yield of the framework and make an interpretation of it into a significant answer for a business issue. The data mining models are of two types: Predictive and Descriptive. The predictive model makes prediction about unknown data values by using the known values. Ex. Classification, Regression, Time series analysis, Prediction etc. The descriptive model identifies the patterns or relationships in data and explores the properties of the data examined. Ex. Clustering, Summarization, Association rule, Sequence discovery [20].

ASSOCIATION RULE MINING: In ARM a couple set of association rules are applied on elements of web site structure to identify relationships among various modules of web engineering application. Association rule mining is finding all association rules with support and confidence values that are greater than or equal a user-specified minsup and minconf respectively. In general, the process of extracting interesting association rules consists of two major steps.

Step 1: The first step is finding all item sets that satisfy minsup (known as Frequent-Item set generation).

Step 2: The second step, is generating all association rules that satisfy min conf using item sets generated in the first step.

After generating frequent item sets, association rules that are greater than or equal to min conf are generated. Those rules are called interesting association rules. Those rules can be invested in many different applications. One of those applications is improving the structure of the company's website that the mined database belongs to. This is done during the website's design phase by creating links between items that seem to be sold together, or highlight those links if they are already exist, and/or create index pages which are pages that have direct links to some products that may be of interest for some group of customers. All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified [17].

CLASSIFICATION: Classification is dividing an existing set of events or transactions into other predefined sets or classes based on some characteristics. In web usage mining, classification is used to group users into predefined groups with respect to their navigation patterns in order to develop profiles of users belonging to a particular class or category. A web mining strategy for web personalization based on a novel pattern recognition strategy which analysis and classifies users taking into account both user provided data and navigational behaviour of the users[17].

CLUSTERING: Clustering is the process of partitioning a given population of events or items into sets of similar elements, so that items within a cluster have high similarity in comparison to one another, but are very dissimilar to items in other clusters. In web usage mining there are two main interesting clusters to be discovered: usage clusters and pages clusters [17].

In a comparative analysis of multiple prediction models, it is a common practice to split the complete data set into training and testing sub sets, and compare and contrast the prediction models based on their accuracy on the test data set. In splitting the data into training and testing dataset one can choose to make a single split (e.g., half of the data for training and other half of the data for testing) or multiple splits, which is commonly referred to as k-fold cross validation. The idea behind k-fold cross validation is to minimize the bias associated with the random sampling of the training and holdout data samples [6]. There are

many applications or tasks in educational environments that have been resolved through DM.

For example, Baker suggests four key areas of application for EDM: improving student models, improving domain models, studying the pedagogical support provided by learning software, scientific research into learning and learners; and five approaches/methods: prediction, clustering, relationship mining, distillation of data for human judgment and discovery with models. Castro suggests the following EDM subjects/tasks: applications dealing with the assessment of the student's learning performance, applications that provide course adaptation and learning recommendations based on the student's learning behaviour, approaches dealing with the evaluation of learning material and educational web based courses, applications that involve feedback to both teacher and students in e-learning courses, and developments for detection of atypical students' learning behaviours [9].

Prediction estimates numeric and ordered future values based on the patterns of a data set. Han and Kamber note that, for prediction, the attribute for which the values are being predicted is continuous-valued (ordered) rather than categorical (discrete-valued and unordered). This attribute can be referred to simply as the predicted attribute. Neural networks and logistic model prediction are the most commonly used prediction techniques [10]. A CRM system is a repository of customer information which contains all customer profiles. In addition to the traditional database roles, it has the capability of personalising needs of individual customers by differentiating products or services for each unique customer. Popular strategies recommended to improve CRM include the use of BI for price discrimination, lock-in/high switching costs, and BI tools [16].

Business understanding consists of determining business objectives, assessing the situation, determining the data mining goals and producing the project plan. In the following, setting of the business model under these activities. None of these activities in themselves already violate privacy, but the choice of objectives, goals and plans may imply later business practices that are prone to do so [19]. PETs are tools and mechanisms which, when integrated or used in conjunction with

online services or applications, allow users to protect their data provided to and handled by such services or applications. They provide encryption (to prevent eavesdroppers from seeing the content of transferred information), and/or anonymization/pseudonymization (to prevent the identity of the communication partners from being released) [19].

IV. CHALLENGES IN BUSINESS INTELLIGENCE

In 1989, Howard Dresdner proposed "business intelligence" as an umbrella term to portray "ideas and strategies to enhance business choice making by utilizing actuality based help supportive networks. It was not until the late 1990s that this use was far reaching. Frequently BI applications use data accumulated from a data distribution center (DW) or from a data shop, and the ideas of BI and DW now and again join together as "BI/DW" or as "BIDW". A data stockroom holds a duplicate of explanatory data that encourages choice backing. Be that as it may, not all data distribution centres serve for business intelligence, nor do all business intelligence applications oblige a data stockroom.

Starting a business intelligence (BI) method could be scary for little endeavours and departmental arrangements. BI is regularly seen as a convoluted, language filled coliseum that obliges bunches of IT mastery and assets. There is additionally a pressure between conveying quick comes about monetarily and making the best choice for long haul development and adaptability. This paper furnishes littler associations with a schema to assess and select the right BI result and approach that meets their necessities today and tomorrow.

DIFFICULTIES MAY CONFRONT:

A. Guaranteeing development – Gainful and maintained development is discriminating for a little venture. Without it, chance either being made unimportant by speedier contenders or being swallowed by a bigger organization. Productive development obliges pulling the right levers on a reliable premise, so there ought to be clear experiences into your business. Case in point, knowing which items, sections, and clients return higher benefits will help to dispense more assets to

them. On a continuous premise, there ought to have the capacity to screen and approve if your incremental ventures are in fact conveying a normal expand in deals and benefits.

B. Accomplishing pace – Capability to quickly exploit another business open door is one of your greatest favourable circumstances. Nonetheless, more modest associations frequently have restricted assets. Need to comprehend what is working admirably, so it could be immediately promoted, and what is not living up to expectations well, so it might be quickly tended to. Know the income, costs, and use for as of late presented items. This can prompt all the more effectively recognize poor entertainers and rapidly movement use far from them. Without such experiences, it may take more time to settle on such choices and keep wagering on poor entertainers, or it may settle on wrong choices that are wrongly affected by one data point.

C. Staying Focused – Not at all like substantial organizations, little associations offer a slender arrangement of items and administrations and typically have constrained assets. Proceeded with achievement lies in getting everybody in the group on the same page by utilizing the same set of presumptions focused around the same variant of reality. Along these lines, if building, deals, fund, and showcasing associations utilize the same examination on pipeline, income, and edges, they are more inclined to have the same conclusions and be adjusted on necessary.

V. DATA MINING OUTLOOK IN BUSINESS INTELLIGENCE

The right time to take business intelligence (BI) to the following level by including data mining and propelled investigation. This is a level of BI brilliance that numerous associations never figure out how to develop to, however the criticalness of pushing ahead with cutting edge competencies can't be disparaged – they can give a really manageable playing point and empower your association to amplify both its productivity and viability. Data Mining is the procedure of running data through refined calculations to uncover genuine examples and connections that may overall be stowed away. These might be utilized to help you comprehend the business better furthermore misused to enhance future execution through prescient investigation.

Case in point, data mining can caution you there's a high likelihood a particular client won't pay on time focused around an investigation of clients with comparative aspect. The below represented graph.1 shows the rate of data mining used in various different applications.

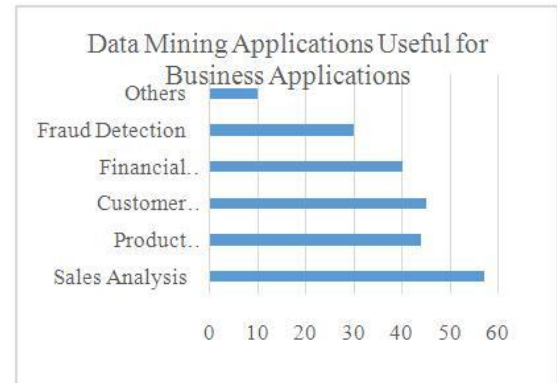


Figure 1. Data Mining Applications Useful for Business Applications

To help you completely use data mining for hierarchical playing point, the Pentaho BI Undertaking group has worked in conjunction with the advancement and business groups to coordinate standard BI capacities with cutting edge data mining. Pentaho Data Mining is separated by its open, guidelines agreeable nature, utilization of Weka data mining engineering, and tight joining with center business intelligence capacities including reporting, examination and dashboards. Other data mining offerings fail to offer this level of sophistication and integration.

In this report, blanket the business profits of integrating data mining as a component of business intelligence arrangement, together with the how's and why's of data mining to furnish you with a strong understanding of this theme. Gives knowledge into concealed examples and connections in your data. An exemplary sample of data mining is a retailer who reveals a relationship between offers of bread and screw driver on Sunday evenings – two things you wouldn't typically consider as connected. The clarification is that individual who are conveyed to get a crisp supply of breads are additionally prone to get screw drivers while they happen to be in the store – something that hadn't been perceived as a critical deals driver before data mining uncovered it.

Empowers you to adventure these connections to enhance hierarchical execution proceeding with the sample above, all the time retailers follow up on the connections they find by utilizing strategies, for example, putting interfaced things together on end-of-isle shows as an approach to goad extra buys. All associations can profit from acting in a comparable manner – utilizing newfound examples and connections as the premise for making a move to enhance their effectiveness and viability.

"The individuals who don't gain from history are destined to rehash it" is an acclaimed quote from savant George Santayana. On account of data mining, having the capacity to foresee conclusions focused around noteworthy data can drastically enhance the quality and results of choice making in the present. As a straightforward sample, if the best marker of whether a client will pay on time ends up being a blend of their business portion and whether they have paid past bills on time, then this is data you can helpfully profit from in settling on present credit choices.

CRM as depicted by Strategic Customer Relationship Management System (SCRMS) which gathers, coordinates and determinations different client related data from distinctive operation frameworks in divisions inside an endeavor characterizes CRM as streamlines values as gainfulness, income and client fulfillment (what and why) by sorting out around client sections, encouraging client fulfilling practices and actualizing client driven business models.

V. DISCOVERY OF DATA FROM LARGE TRANSACTIONAL DATABASES

Item set mining is an exploratory data mining technique widely used for discovering valuable correlations among data [22]. The first attempt to perform item set mining was focused on discovering frequent item sets, i.e., patterns whose observed frequency of occurrence in the source data (the support) is above a given threshold. Frequent item sets find application in a number of real-life contexts (e.g., market basket analysis, medical image processing, biological data analysis. However, many traditional approaches ignore the influence/interest of each item/transaction within

the analyzed data. To allow treating items/transactions differently based on their relevance in the frequent item set mining process, the notion of weighted item set has also been introduced. A weight is associated with each data item and characterizes its local significance within each transaction.

Two novel qualities measures are considered to drive the IWI mining process. Infrequent item sets that do not contain any infrequent subsets have been considered. Experiments can be performed on both synthetic and real life data sets, show the effectiveness of the considered approach. In particular they show the characteristics and usefulness of the item sets discovered from data coming from benchmarking and real. To reduce the computational time the residual trees are introduced. The item sets that are both high frequent and high utility can be obtained using the method. The customer relationship management is incorporated into the system by tracking the customers who are frequent buyers of the different kinds of item set.

A number of relevant algorithms have been proposed in recent years for the fast access of data from the database. The situation may become worse when the database contains lots of long transactions or long high utility item sets. The efficient utility pattern algorithm is used for mining high utility item sets. A frequent pattern in a database may be a set of items, or a subsequence, a sub-tree or a sub-graph that appears across the database with a frequency that is not less than a pre-defined threshold value. Two algorithms, named utility pattern growth (UP Growth) and UP-Growth+, and a compact tree structure, called utility pattern tree (UP-Tree), for discovering high utility item sets and maintaining important information related to utility patterns within databases.

High utility item sets can be generated from UP-Tree efficiently with only two scans of original databases. Several strategies are there for facilitating the mining processes of UP-Growth and UP-Growth+ by maintaining only essential information in UP-Tree. By these strategies, overestimated utilities of candidates can be well reduced by discarding utilities of the items that cannot be high utility or are not involved in the search space. The proposed strategies can not only

decrease the overestimated utilities of PHUIs but also greatly reduce the number of candidates. Different types of both real and synthetic data sets can be used in a series of experiments to compare the performance of the algorithms with the state-of-the-art utility mining algorithms.

A large number of candidate item sets degrades the mining performance in terms of execution time and space requirement such problems can be avoided. Multiple time database scanning for the generation of candidate item sets can be reduced using frequent and high utility pattern. The traditional association rule mining (ARM) is used to identify frequently occurring patterns of item sets. ARM model treats all the items in the database equally by only considering if an item is present in a transaction or not. The frequency of occurrence may not express the semantics of applications, because the user's interest may be related to other factors, such as cost, profit, or aesthetic value. To overcome the problem of the existing model a new tree structure called hybrid tree (H-Tree) which combines the Incremental High Utility Pattern Lexicographic Tree (IHUPL-Tree) and Incremental High Utility Pattern Transaction Frequency Tree (IHUPTF-Tree) can be proposed.

CONCLUSION

This concludes with Data Mining is one of the utilitarian social affairs among the comprehensive undertaking class investigative functionalities offered within Business Development. In the past 10–15 years, information mining has transformed into a development in it right, is settled furthermore in business intelligence (BI), and continues showing tenaciously extending vitality in designing and life sciences divisions. This business gathering use information digging as an instrument for settling monetarily essential business applications, for instance, customer relationship organization, coercion distinguishment, and so forth. This field is basically secured by a mixture of business gadgets offering support to databases with tremendous datasets, and significant fuse in the association's work process. For preparing at universities, data mining mechanical assemblies should be to a great degree common, with a satisfactory natural customer interface, and modest.

Conceivably high utility item sets are produced utilizing utility example with two database filters. This joins with the regular example to give better execution and gives best answer for time utilization. Apriori calculation obliges different time databases checking. To discover long examples it may require an excess of database checking that is very time intensive. Interim, while preparing datasets that contain long examples, it produces an excess of hopefuls and sub sequences of continuous examples. To take care of these issues high utility example can be utilized, which dodges the expensive competitor era and requires just two times database examining. The principal pass discovers every regular thing, and the second pass builds reduced information structure utilizing the high utility things which are utilized for putting away compacted, vital data about high utility patterns. In future Customer Relationship Management (CRM) is consolidated into the framework by following the clients who are successive purchasers of the various types of item sets.

REFERENCES

- [1] Ralf Mikutand Markus Reischl, "Data mining tools", Data Mining and Knowledge Discovery, Springer, Volume 00, January/ February 2011.
- [2] Chen Li, Manfred Reichert, Andreas Wombacher "Mining Business Process Variants: Challenges, Scenarios, Algorithms", Data & Knowledge Engineering, Elsevier, 2011.
- [3] JozefZurada and SubhashLonial, "Comparison Of The Performance Of Several Data Mining Methods For Bad Debt Recovery In The Healthcare Industry", The Journal of Applied Business Research, Springer, Volume 21, Number 2, 2005.
- [4] Andrew KusiakandZijun Zhang, "Adaptive Control of a Wind Turbine With Data Mining and Swarm Intelligence" IEEE TRANSACTIONS ON SUSTAINABLE ENERGY, VOL. 2, NO. 1, JANUARY 2011.
- [5] Sara Hajian, Josep Domingo-Ferrer and AntoniMartínez-Ballesté, "Discrimination Prevention in Data

- Mining for Intrusion and Crime Detection”, IEEE, Computational Intelligence In Cyber Security, 2011.
- [6] Xi Chen, Indranil Bose, Alvin Chung Man Leung, ChenhuiGuo, “Assessing the severity of phishing attacks: A hybrid data mining approach”, Decision Support Systems, Elsevier, Volume 50, Issue 4, March 2011.
- [7] Wu He, “Examining students’ online interaction in a live video streaming environment using data mining and text mining”, Computers in Human Behavior, Elsevier, Volume 29, Issue 1, January 2013.
- [8] Alan S. Abrahams, Jian Jiao, G. Alan Wang, Weiguo Fan, “Vehicle defect discovery from social media”, Decision Support Systems, Elsevier, Volume 54, Issue 1, December 2012.
- [9] Jinn-Yi Yeh, Tai-Hsi Wu, Chuan-Wei Tsao, “Using data mining techniques to predict hospitalization of hemodialysis patients”, Decision Support Systems, Elsevier, Volume 50, Issue 2, January 2011.
- [10] FRANCESCO BONCHI, CARLOS CASTILLO, ARISTIDES GIONIS, and ALEJANDRO JAIMES, “Social Network Analysis and Mining for Business Applications”, ACM Transactions On Intelligent Systems And Technology, Volume 2 Issue 3, April 2011.
- [11] Dien D. Phan, Douglas R. Vogel “A model of customer relationship management and business intelligence systems for catalogue and online retailers”, Information & Management, Elsevier, Volume 47, Issue 2, March 2010.
- [12] TapanNayak, Prof. B Lakshma Reddy, “A Mining Approach for Web Engineering In Respect Of Business Intelligence Application”, International Journal of Modern Engineering Research, Vol.3, Issue.2, March-April. 2013.
- [13] Xuezhong Zhou, Shibo Chen, Baoyan Liu, Runsun Zhang, Yinghui Wang, Ping Li, YufengGuo, Hua Zhang, Zhuye Gao, Xiufeng Yan, “Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support”, Artificial Intelligence in Medicine, Elsevier, Vol.3, Issue.2, March-April. 2013.
- [14] Bettina Berendt, “More than modelling and hiding: towards a comprehensive view of Web mining and privacy”, Data Mining and Knowledge Discovery, Springer, Received: 11 January 2010 / Accepted: 20 January 2012 / Published online: 15 February 2012 © The Author(s) 2012.
- [15] NeelamadhabPadhy, Dr.Pragnyaban Mishra, and RasmitaPanigrahi, “The Survey of Data Mining Applications and Feature Scope”, International Journal of Computer Science, Engineering and Information Technology, Vol.2, No.3, June 2012.
- [16] Liu. Y, Liao. W, and Choudhary. A, “A Fast High Utility Itemsets Mining Algorithm,” Proc. Utility-Based Data Mining Workshop, 2005.
- [17] Luca Cagliero and Paolo Garza, “Infrequent Weighted Itemset Mining Using Frequent Pattern Growth” IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 4, April 2014.
- [18] Tao. F, Murtagh. F, and Farid. M, “Weighted Association Rule Mining Using Weighted Support and Significance Framework,” Proc. ACM SIGKDD Conf. Knowledge Discovery and Data Mining (KDD ’03), pp. 661-666, 2003.
- [19] Vincent S. Tseng, Bai-EnShie, Cheng-Wei Wu, and Philip S. Yu, Fellow, IEEE “Efficient Algorithms for Mining High Utility Itemsets from Transactional Databases”, IEEE transactions on knowledge and data engineering, Vol. 25, No.8, August 2013.