

# INFORMATION RETRIEVAL THROUGH MACHINE LEARNING

<sup>1</sup>Nasir Naveed, <sup>2</sup>Muhammad Abubakar Siddique and <sup>3</sup>Muhammad Tariq Pervez

<sup>1</sup>Department of Computer Science, Virtual University of Pakistan  
nasir.naveed@outlook.com

<sup>2</sup>Department of Computer Science, Virtual University of Pakistan  
[MS140400077@vu.edu.pk](mailto:MS140400077@vu.edu.pk)

<sup>3</sup>Department of Computer Science, Virtual University of Pakistan  
Tariq\_cp@hotmail.com

**ABSTRACT:** Now a days of business age & the managers are in wait of concise and accurate information about the consumer's likeness trends even the information is in bits, they concern only with the accuracy and relevancy of information about a particular product / service or event by the consumer. The astonishing thing is to track the users' information & interests from social media, surveys and from users search queries. This is a real fact, but the question arises about how to gain this information? Has it worthy or worthless? We are proposing system that is using different flavors of data include the meta-data. Also discussing about how to system works for data recovery / mining and will enable to perform proper forecasting and in-time analysis so that optimal decision could be taken. Another question arises i.e. how to retrieve & achieve the relevancy of to acquire destination research result in optimal time span? As well as the support to machine decision process of artificial intelligence. Third one question is about how to interpolate the data from social medial review web sites & search engines etc, we are discussing on a processing model with the collaboration of expert system which is performing data exploration to retrieval of feasible and relevant information. The discussion is about the different techniques & methods of data explorations & mining. Our model is fully sensible and by using clever algorithm it will predicts / forecast on our under study information.

**Keyword:** Business age, managers; meta data; organized; optimal; decision; artificial intelligence, expert system, collaboration, exploration; operations; feasible; predicts; forecast.

## 1. INTRODUCTION

In the history of computers, numbers of transistors per dense of an integrated circuit doubles every two years and their prices becomes low [MOORE Law].

This analogy also has context with database. The size of information become doubles day to day. The database size has been increasing astonishingly. The use of computer inventories from grocery store to university & birth place to graveyard is occurring frequently. In these days every bit of life aspect is depended and its transaction is accordingly recorded. This is the story of lay man. On the other side the databases of research institutions & governments also increasing. Another interesting factor is that the research investigations are being recording the old decade's manual data into computer readable format for analysis purpose. NASA [1] has assessed the alternative mission human Mars exploration by using multi criteria decision making approach. For this purpose, NASA has uses the data mining for getting optimal decision within stipulated time span. In bio technology for studying of large DNA are quite easy with data mining techniques with inference of statistical analysis. Machine based intelligence is the technique to sieve / mine the desired information. All the companies are working to find information about consumer interests so that they have to work on those fields for acquiring better outcomes of their investments. In that research paper we are also working on the process of digging out requisite & valuable form of large patterns of database. The sieved / mined information is placed in data mode & artificial intelligence techniques are used for finding required data patterns. Typically required data pattern & rules are sieved from 2<sup>nd</sup> dimension view of database. The data structures & relations are not concerns by the data engineers in this view of database. By using the meta data model the interested data patterns are easily traceable.

Especially expert systems are used for intimating requisite patterns [2]. This system focuses on the followings for process of acquiring desired results:

**Patterns:** The patterns mean the situation where we need the number of users aged 10-50 years who like to eat chicken and are surfing on Facebook. We want to prepare a advertising campaign on the social media specially on Facebook about new launching chicken products by our firm [3].

### Authentic

The accuracy of data sample is mandatory because on the basis of sample the decision could be made. It includes integrity, size & degree of accuracy. Without authentic data the sample results are misleading. E.g. on a social media a campaign about fat burner coffee was advertised which shows about thousands of likes and hundreds of gratitude's by the consumers. The outcome of good advertisement is that the consumer got impressed from the splendid comments and purchased the produced but all in vain after using the product. The fake fabulous comments insists him to purchase the product after that about experience of the product he may commented against the product, the admin of the social media page, removes all the anti comments, then the question arises what type of action will be taken. Ultimately the consumer would post on the other media groups. But when our processing model scans / mines / sieve the information it never give importance to anti comments as these are very less in quantity as compare with the fake comments. Hence, it is big fear for acquiring the unauthentic information. Every problem in the world has a solution, for the aforesaid problem we are proposing for use of strategy introduced by the [12] regarding fast spread of bad news on the twitter.

### Useful / Relevant

The collecting data is not a racket science but the filtration of relevant data is required some technique. Our goal is to discard irrelevant data and acquire relevant data for better forecasting of results on the basis of sample [13].

### Effective and Efficient

For effective and efficient data acquisition the algorithm is used [11].

## 2. DATA PROCESSING MODEL

The processing model that should be used by a expert system [14] is depicts in the below Figure 1:

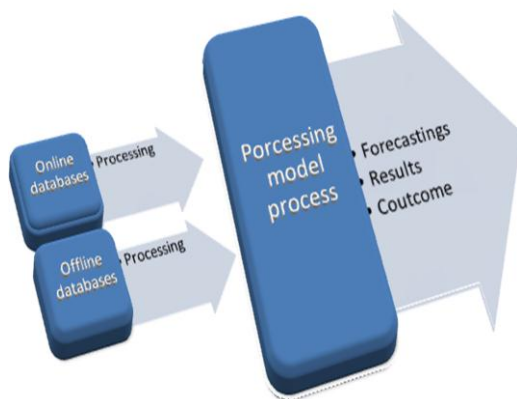


Fig.1: Proposed data processing model.

We forwarded the requisite database in the processing model. The database is the collection of heterogeneous data with multiple extensions; our processing model has the capability of to convert all extensions files into a single extension (file type).

The process is start from the meta data & the first stage is file type/ category analysis. This analysis has a major role in processing model. Its collects the all information that received from multiple vendors with fully completed all aspects e.g. their relationship along with dependencies. This information performs soul action for expert system; on the basis of same the experiments can be done. A file type editor is used for discrepancy verification from existing data. And the 2<sup>nd</sup> phase i.e. experiment editor used to run multiple expert systems over multiple instances of database.

Meta Data has a key role in data validation. Our fully devotion is with the meta data. The process model should helpful for the expert system for categorization of data types. Now our key component is from that we retrieve our results from database. The database is collection of logically correlated data exists in single or multiple files organized form feasible retrieval of requisite data or information. The data is organized in the form of tables & its each field is said record. The information about that fields and about the data contained in it is save in another file said data about meta data or data dictionary. And the database management

system is the collection of logically correlated functions that performs data / information, manipulation, retrieval & storing operations. The [7] expert systems are using these databases by maintaining them into single file of fixed length vectors. We are extract data from them by using an algorithm & represent it in the form of output. A lot of data is uncompleted and noise effected, a lot of time has been consumed to overcome on these anomalies. For discussion on this topic we must have to introduction with some techniques detailed below:

**Database Management System:** the DBMS has procedures for storing, acquiring & modification of data. Decisions can be made according to this information retrieved through database management system. Day by day new data base systems i.e. objected oriented data management system & other are developing, that are capable of automatic retrieval of user usable information without any extra inference of user or data base administrator.

**Expert System:** Expert system has a major role in the machine learning and artificial intelligence. The famous problem of assembly line during the production of vehicle has been now tackled by the expert systems. In their optimal solution of finding the less time wastage or the methods to get high proficiency work in less time [4]. The heuristic are also used by the expert system to solve the industry bases problems solving and decision making. Firstly the expert systems reads / examine the all aspects / knowledge of problem situation & then it's used its knowledge based (depends on multiple inputs & heuristics) to solve the issues or to examine any case and provide its solution according to the knowledge it have [3]. The fully operation and successful models/examples of the expert systems are automatic clinical lab test machines, that are do the operations on samples as per precompiled instructions. Second one is the modern automatic hepatitis & pregnancy identification kits, when the sample is placed on that kits, the sample had come to reaction with already available chemicals on the kit, resultantly the kit / stick changes its color. In case of sugar level checking the blood sample read out by the data reader and after performing pathological operation its processor calculates the sugar level in the blood and displays it on the output screen. Thirdly the example of automatic blood pressure apparatus when its cloth was beard with the patience arm it itself pumps it with air and start reading of blood pressure with heart beat rate & produces the result on output screen. Fourthly is the example of simple thermometer, when the patient places it his / her mouth its built-in sensors are staring reading the mouth temperature and in case of analog thermometer the mercury level get rise on the scale, in case of digital thermometer it will display the temperature output on its built-in screen. These all the simple and real life expert system that are working on predefined installed instructions, now we have to start work on more complex situations to solve the unseen issues and that is so though but much interesting [4].

### Statistical Model

The statistics has a vital role and ample work in the modern research days. Also for our research purpose the use of statistical model is inevitable; it has a major impact on research analysis. The results of researches may be got

urgently by assuming the hypothesis and heuristics can easily be derived. Here our aim is to make out meta data and data sets so strong that can fruitful for acquiring hundred percent accurate results from any sudden situation. Our statistical model will be so strong and expert can do operations on the worse situations and give effective and efficient results so that beneficial for optimal decision making. We also get / check variance of information by using statistical models for tracking of users interests on the social media by extracting information from working on by calculating number of followers + callings + likeness [13].

For capturing the desired results by the experts systems or any other machine or algorithm the base is to proper explore / find or search out the meta data or data dictionary, then we are able to solve / trace out the solution of desired results.

Above diagram illustrate about the methodology used in our data processing model to acquire out desired results. The main role is the algorithm of finding relevant patterns. On that we to work for the statistical model that will examine / read and search out the desired and discarding the irrelevant data streams. When the data is inputted in the processing model this include all databases, but he system has the function to focus on the related patterns so that the operations could be done on the basis of that data. After processing the out enables the user / expert system to forecast their results and do the optimal decision on that basis of information. Our processing model has based on the following functionalities.

- (1) Exploration of databases.
- (2) Information exploration process.
- (3) Techniques for information exploration.
- (4) Proper acquisition of results for end user.

### 3. DATABASE ISSUES

The data processing model is fully depended on the input database, the without the input there is no concept of processing and output. As already discussed that the main issue is that to acquire a well organized form of data that capable our processing for mind blowing decision especially in machine learning it is big issue but still not impossible. Another challenge is how to tackle the multiple databases that are also static, dynamic, partial and noise influenced, by tackling this situation we may able to complete our fifty percent work. The second challenge is that how to filter or trace that under observation data that is relevant to our desired domain or not. As we know that the contents of dynamic databases or not stable and ever changing after time to time. Also there is time management factor for maintaining data e.g. the amount of ice in the northern areas of Pakistan has been decreased in summer season & when this ice melt and produced in the form the water the water level of rivers rise up and that water also captured by the irrigation system, dams and lakes and to some amount will wasted in the sea and a minor effect will happen in the sea level as well as the same amount of water , ice and water level of sea & river vary in season to season, hence it is very careful job for the humans as well as the expert system to give opinion or decision on the specific situation. Here is the use of statistical model for variance calculation [13]. In this data some things remain static, as the rivers, seas, mountains

& canals data that could not be changed easily even the out of turn atmosphere situation as in current days.

### Irrelevant Data Fields

The data / grains must be feasible relevance to as in the student database the name and roll numbers must be correctly entered in the relevant data fields, e.g. if the all the correct the data must be semantically and syntactically correct, as in the primary class students the maximum age is of 5 to 12 years, it not feasible that the age of primary student is 50 years (same case are occurs in non formal education systems not in formal education systems) Any data operation by the data mining engineers or experts systems must be done on the basis of analogy of same heuristics.

### Ambiguous and Noised Data

Some times in the situation wrong data types were selected / chose by the data entry operators at the time of initial stage e.g. where we are required the integers data type wrongly we selected the real data type, as we now that both have quite different attributes this minor error results a very huge difference in the interim or regular data analysis. Second major cause of ambiguous data is wrong identification of inheritance relationship and wrong demonstration of entity relation data high level of data garbage. Thirdly depend of uncompleted analysis and results are the next other big cause for the same. Forth main cause of unambiguous data is non-removal of deviations from the statistical produced data as per defined standards (careless behavior for updating existing record). The fifth main cause is computation of non- standard numeric data, because 64 bit operating systems produce more near & accurate decimal results instead of 16 bits, when the mixed up data of 16 bit and 64 bits operating system calculation is combined and analysis made on its basis leads to enormous biased results.

### Incomplete Records

As we know that the database is a collection of correlated data. Sometimes the data is entered incomplete, as the name of student is writing but its age has been missed from data entry. As in the lab test the urine sample has been taken of male person but its pregnancy has been showed in the data base which is impossible. Hence these types of case are the example of uncompleted records. A heuristic approach is required to detect such problems.

### 4. INFORMATION RETRIEVAL FROM DATABASE

As already discussed above that a database is a collection of logically inter correlated data and it is also a correlation with data dictionary on the basis of meta data. The data dictionary has a vital importance as its also helps for finding relevant patterns but also classified the data into its types, e.g. given the field of date of birth in the form of days, months & years in Asian countries and months, days & years in the Europe countries, this date type format varies from region to region. The data is collection of raw material the data become database after come sort of operation, this operation is not possible to do without the use of data discovery, because without its classification database existence is not possible. The retrieved information from existence data is said to be domain information. The meta data in data dictionary knowledge has the basis for further information retrieval, updation and removal.

## Discovered Data Information

In this part the multiple portion of discovering information are discussed as under:

1. Information Shape & Representation of data.
2. Accuracy standard.

### 1. Information Shape and Representation of Data

The First thing is about the information derived from that data also classified about the meta data classification. All the records are correlated with their related data/information e.g. as for the Banker it is necessary for him to reach in the office at 9 o'clock AM. and left the office at 8 o'clock PM., else in case of late reaching or leaving the seat without getting approval from competent authority, there will be a strict disciplinary action can be taken out by the respective Bank administration. Hence, in this example the bank discipline & timings are correlated with the bank employee. This scenario is like a tree where each branch is depended over other on a prescribed standards e.g. neural networks etc. The retrieved data & information can also logically visualize in tree shapes or entity relationship diagrams.

### 2. Authenticity of Data

The accuracy or authenticity of data is mandatory but it is quite difficult to maintain its accuracy as it is totally or partially based on probabilistic data. There is another big bang comes in the universe if the respective data might be wrong or miss-interpreted, and are taking decisions upon it retrieved information [9]. We have to acquire that data & information that has maximum chances of accuracy. The hundred percent accuracy & authenticity required for human language interpreters and translators and in that situation the role of fuzzy set theory is inevitable. Our proposed processing model used the fuzzy logic for achieving maximum degree of accuracy. To achieve maximum degree of accuracy is quite difficult and interesting. E.g. there is 25% probability of storm in the sea at the departure of cargo ship and after 2 hours of its departure the probability increases from 25% to 80%, at that stage the Caption of the ship decision matters either be a risk lover or risk averse. If the attitude is risk averse he cannot travel in the deep sea, in case of risk lover he instantly start travelling in the middle deep sea but not remember to plane to tackle the any emergency situation. In this example it is also very difficult for a expert system to take decision for traveling, but using fuzzy set theory or on the other hand the large amount of mathematical induction and statistical variance distribution analysis are used for taking feasible decision. But as we know that our processing model has to be work with multiple & bulky database management systems it is not feasible to do analysis operations on full database management system. Hence, at that stage our data processing model have to quickly sort the samples with the help of statistical sampling distribution. At that stage for sample collecting the fuzzy set theory is used for acquiring samples from database managements systems and after that statistical sampling distribution analyses made on the samples to acquire feasible results, then on the basis of these samples results the optimal decision can be taken. Here is again

the decision and analysis is done that the results taken from samples are accurate or not or there is needed to take more samples for most accurate assessment. By doing work with multiple techniques our processing model will able to take such sort of elegant decisions [5].

## 5. METHODS FOR DECISION MAKING

The hub of our processing model is the methods or algorithms for acquiring 100% accurate information this can be possible by two ways.

### 1. Capturing of Relevancy Degree

The process of capturing degree of relevancy among the records, information, fields, calcifications of information.

### 2. Expanding Procedure

In expanding procedure, the acquired information is further summarized by the processing model with the user's influence.

### 1. Capturing Relevancy Degree

The capturing degree of relevancy means that to find out the attributes common in average with other values and functions. The modern web spy-wares also belong to this category. These spy-wares are get installed with the hidden / cryptogram method in user's PC or cell phone and look after the actions & activities of the user on the web and simple uses by capturing that in a log file. After a stipulated time span these log files transferred to hacker's PC through any cipher method. Where the interests of the users are captured through some analysis as already discussed in this paper & according to that information the hackers as well as the marketing companies will target the users with the advertising of products that he has interested. This strategy is a crime by law but now a day's all the modern advertisement and famous companies are supporting and doing on this analogy in order to maximize their profits and sale targets. Because when they are get familiar with the consumer interests of a specific geographical area, it is bit easy for the multinational companies to capture the consumer. E.g. the firms / companies are also arranged surveys about what are the interests of females aged 25-50 in Lahore City, and know from that their interest is to dress up her with the cloths of famous brands, and the same age groups of Multan city females have keen interest in cooking. So the firms have doing their market advertisement according to the geographical area people interests. Now days, it is observed that on surveys a lot of finance and time spent. So the firms are hiring the data mining engineers to help out them for acquiring feasible information about the consumer interests that also involves less time & finance. The multinational firm's interests in this field are a big cause for propagation of law crimes also. We have here discussed the rising demand of information need & requirement is the big cause in increase in cyber crimes. We also used the Euclidean Geometry proposed processing model to capture the degree of user interest.

### 2. Expanding procedure

After finding the degree of relevancy the information is required to be expanding here we are using the pure artificial

intelligence. Here we are using the decision trees and neural networks [10].

### 5.1 Capturing Degree of Relevancy Tasks

The main tasks for capturing degree of relevance are listed below:

1. Aggregation.
2. Identification.
3. Analysis.

#### 5.1.1 Aggregation

The aggregation of all the records on the basis of relevancy of attributes, e.g. all the warranty claims of tyres of Hino Buss for the financial year 2013-14. In this example we are requiring the warranty claims of tyres class that are using by the Hino Buses class, there are two classes are being used. The second example is the consumption of chicken in all the hotels of Lahore City. In this example one class is of chicken and the other class is the hotels of Lahore city, on that grounds the relevant data could be aggregated.

#### 5.1.2 Identification

The identification has a major impact in data mining engineering as the use of "Chaddar Culture" is in very high range in the Lahore City in the winter season. On the other hand the use of jackets and long shoes are high demand of consumers in the Islamabad and northern regions of Pakistan as well as the in these area the use of meat & dry fruits increase rapidly as compare to south region for country. Hence the identification of information according to the geographic situation and elite and middle class consumer's areas shopping trends, education, and hotelling, wearing & traveling absolutely differ from each other. E.g. the elite people of southern region in summer season go to live in Murree, Abott Abad and swat areas, while the poor ones enjoy the summer season by taking baths in canals.

#### 5.1.3 Analysis

The analyses are the major key of discussion in our paper as well as in the modern world. About every aspect of life especially, in the business world depends on the analysis. As we already discussed about the user interests. The firms are also having a keen knowledge & information about the consumer variance about the products from season to season. We are also working to mine better data in order to get success in our analyses. As on the side our processing model in digging out data and we are simultaneous applying multiple operations on it and converting it into useful information format, other side our processing model also working on that about the analysis that our analysis is going on right direction or otherwise we are require to extract other data and do other analysis factors or not. We have to say that each iteration of information extraction, this process becoming more atomic for getting hundred percent accurate results in every aspect in order to survive in the 24 hours working & simultaneously competing world.

#### Heavy Computation

Now we are coming on the turning point the practical, actual theoretical work is so easy in contrast with practical work. Our aims are very high so to dig out the useful information from corpus / heave database management system a huge computation is required. Too much complexity occurs. For digging out requiem information and

data algorithm required according to the data requirement the complexity occurs as the complexity occur the computation cost get raises second to second. We increase our work to get pure information then we have to trade off with finance. Our main target is to dig feasible information with low computation, for this we must use the fuzzy set theory and statistical sampling distribution.

### 6. CONCLUSION.

The access to data bases is increasing rapidly, above is discussion; some actions on some issue are required to take on emergent basis. The data exploration is emerging field and our discussion spread over the paper is the data mining and exploration is emerging and so interesting but not easy as seems. The research is also applied for all type of databases of daily life usage application and facing issues. It requires a hard work and strong heuristics approach for obtaining fruitful outcomes. This is also useful for mobile companies operation for towers installation so that proper signal propagation will happen, as well as helpful for robotics & drones for taking quick and less time consuming algorithms. The data and information is beneficial if the work of computer scientists for developing of strong logic and fuzzy set theory so that new horizons can be explored. At that situation the work on finding effective and efficient algorithms is required, that works in less time and produces long time benefits. Here is also need of incremental models as applied in software engineering processing model in which a single phase completed and examined by beta versions and after some due course of time final version will produced, same analogy of work is required at that stage and further required work is discussed below in future hone & work agenda.

### 7. FUTURE WORK

The future hope & work agenda the real life examples for exploration of information for analysis of feasible results are listed below:

1. Retrieval Model for capturing user's interest from Tweeter [12].
2. Financial analysis
3. Biotechnology Analysis
4. Social media analysis.
5. Agriculture analysis.
6. Insurance & warrantee claim Analysis
7. Engineering problem solving
8. Atomic reaction simulation
9. Astronomical analysis
10. Military & intelligence analysis
11. Budget comparisons
12. Town Planning.

#### 1. Retrieval Model for Capturing User's Interests from Tweeter

[12] uses the classical vector space model on which they firstly use the issues of sparsity and its effect on length normalization & secondly used the quality of contents by

capturing the tweet sense / interest regarding the users. Actually this is emerging field, in which users' interests predicted / captured on the basis of their likes & re-tweets on the social media. This is a big consumers' interests forecasting field. They also trained their model to check / validate the users interests on the basis of re-tweets [12].

### 2. Financial Analysis

Financial analysis is predicted for future business planning, especially for the forecasting of feasible stock marketing analysis & optimal purchasing [15].

### 3. Biotechnology Analysis

Modern DNA, cancer prediction, treatment of chronic disease. Now the NGO's are working to help the humanity with fall of disease in the third countries & three countries Pakistan, Afghanistan, & Nigeria are high risk rate of polio spread. Now days in Pakistan the polio surveys are done and Prime Minister has a special focus on removal of chronic diseases from our beloved country. Without contribution of data exploration techniques International NGO's & countries are not able to do the valuable works for human beings.

### 4. Social Media Analysis

All the business strategies are developed on the basis of modern research tools the social media analysis are one of them example to seek the social media users interest so that they could developed new games and operations works on the basis of users [14].

### 5. Agriculture Analysis

In America the 60 Metric Ton average of per acre of wheat but Pakistan in 30 Metric Ton the difference is that the Americans have take research analysis to improve their harvests progress. They find out that the hybrid seed and proper spread of fertilizer agents. So they improved their production targets by applying the research techniques.

### 6. Insurance and Warrantee Claim Analysis

The warrantee claim is deeply studied by the firms in order to improve their product quality, and reduce cost of replacement hence, firms good will saved. The insurance companies work on the optimal situation. They cannot take the life insurance policy from any person of age 60 years, because they have the information about that average age of a Pakistan citizen in 60 years, if they will take a person aged 60 years life insurance policy it means the insurance company has taken 100 percent risk. In case the person has medically fit and not has suffering from any chronic disease e.g. sugar & blood pressure. But the morality rate in fact is 60 year's age and the traffic accidents, terrorist activities increases day to day basis. So it is a optimal decision by the insurance company to never get this type of insurance cases.

### 7. Engineering Problem Solving

I already discussed the assembly line problem issues such type of problems are rising day by day in the industry and also their solution hidden the data mining operations and forensics analysis are taken by the engineers & scientists to solve that issues.

### 8. Atomic Reaction Simulation

The reactor simulation can be made with the help of exploration of information from atomic reactions database and proper safety measure find out so that big disasters like Char-noble & Bhopal incidents could have avoided.

### 9. Astronomical Analysis

I already discussed about role of data mining operations in the mission mars. All the spare explorations programs achievements are impossible without the contribution of information exploration.

### 10. Military and Intelligence Analysis

I have already studied in the operation research subject that this subject is founded during to Second World War by the British scientists for saving their supply lines from the Japans forces and kemo flagging their important places [6]

### 11. Budget Comparisons

By using budget forecasting mythology, the organization look into the aspects where they are wasting their resources and working on aspects to less incur and much earn with the help of data forecasting [15]

### 12. Town Planning.

The Town planning have been done with the data simulators, requirements of Masajids, Schools, Colleges, roads & park etc.

## REFERENCES.

1. Tavana, M., "A Subjective Assessment of Alternative Mission Architectures for the Human Exploration of Mars at NASA Using Multicriteria Decision Making," *Computer and Operational Search*, **31**(7):1147-1164(2004).
2. Cleary, J., Geoffrey, H., Sally J. C. and Ian, H. W. "Meta Data for Database Mining," *Proceedings Of The First IEEE Metadata Conference*, Maryland(1996).
3. Tzung-Pei, H., Chai-Ying, L., "Induction of Fuzzy Rules and Membership Functions from Training Examples," *Fuzzy Sets and Systems*, **84**(1): 33-47(1996).
4. Joseph, B. and Jacob, R., "A Weighted Approach for Assembly Line Design with Station Paralleling and Equipment Selection," *IIE Transactions*, **35**(1): 73-85(2003).
5. Brian, R. G. and Paul, C., "Induction of Ripple-Down Rules Applied to Modeling Large Databases," *Journal of Intelligent Information Systems*, **5**(3): 211-228(1995).
6. Mirela, D., "Design of an Expert System for Efficient Selection of Data Mining Method," *Bluetinul Institutului Politehnic Din IAȘI, Publicat de Universitatea Tehnică, Gheorghe Asachi" din Iași Tomul LIV (LVIII)*, 49-73(2008).
7. Bonnet, A. and Cluad, D., "Oil-Well data interpretation Using Expert System and Pattern Recognition Technique," *International Joint Conferences on Artificial Intelligence Organization*, pp. 185-189(1983).
8. Nii, H. P. H., Esward, A. F. and Jhon, J. A., "Signal to Symbol Transformation :Hasp/Siap Case Study," *The AI Magazine*, **3**(2): 23-35(Spring1982).
9. Bennell, J. A., Crabbe, D., Thomas, S. and Gwilym, O. A., "Modelling Sovereign Credit Ratings: Neural Networks versus Ordered Probit," *Expert Systems with Applications*, **30**(3): 415-425(2006).
10. Hoti, S. and McAleer, M., "Country Risk Ratings: An International Comparison, *Seminars of Department of Economics of University of Western Australia*(2002).
11. Naveed, N. and Thomas, G. and Staab, S., "Sentiment Diversification of User Generated Reviews: The FREuD

- Approach”, *Association for the Advancement of Artificial Intelligence*, (2013).
12. Naveed, N. Thomas, G., Jérôme, K. Arifah, C. A., “Bad News Travel Fast: A Content-based Analysis of Interestingness on Twitter,” *In Proceedings of the 3rd International Web Science Conference*, Article no. 8(2011).
  13. Arifah, C. A., Thomas, G., Jérôme, K. and Nasir Naveed, “LiveTweet: Monitoring and Predicting Interesting Microblog Posts,” *In Proceedings 34<sup>th</sup> European Conference on IR Research, ECIR(2012)*.
  14. Naveed, N., Sergej, S. and Steffen, S., “ATT: Analyzing Temporal Dynamics of Topics and Authors in Social Media,” *3<sup>rd</sup> International Conference on Web Sciences*,(2011).
  15. Butler, M., “An Artificial Intelligence Approach to Financial Forecasting using Improved Data Representation, Multi-objective Optimization, and Text Mining,” *Submitted in partial fulfillment of the requirements for the degree of Masters of Electronic Commerce at Dalhousie University Halifax, Nova Scotia* (August 2009).