MULTIDIMENSIONAL SMALL MEDIUM ENTERPRISES ACHIEVEMENT RATING: IMPROVING TO DATA WAREHOUSING & DATA MINING METHODS

JameelAlsarayrah[1] and AlaaTawfiq AL-Zyadat[2]

[1&2]Department of Management Information System, Albalqa Applied University, Amman, Jordan

ABSTRACT

Multidimensional small medium enterprises achievement rating: improving to data warehousing & data mining methods,the study gives a fully performed expert diagnostic system, which assesses the execution of SMEs on a benchmarking security. The system has been in practice for many years and has run into a consistent and considerably challenging development to suffice both the requirements of analysis and the generation of benchmarking records. The research explains why we chose to improve the system with data warehousing and data mining procedures.

KEYWORDS

Data Warehousing, Data Mining, Diagnosis, Expert Systems, SME

1. INTRODUCTION

In this research, a review of some researchers' work briefly described diagnosis system that was developed for SMEs, which can be viewed as a decisiveness support system[1]. Based on a benchmarking strategy [2], our method produces a multidimensional evaluation of an SME's generation and control actions and evaluates the results of these actions in phases of productivity, profitability, vulnerability as well as capability[1,2]. This system is completely achieved and operational and has been set to utilize on data from real SMEs for the last two of years. Suddenly, in the second part of the article, the Author impel on to the principal topic: revealing why conventional database methods used by our system must be improved with the combination of data warehousing as well as the data mining methods to expand its data intelligence abilities. This part of the article outlines continuing and prospective work. Our work practices place within the context of an analysis institute for SMEs—the Institute's focus mission is to promote theoretically and employed research to encourage the progress of knowledge on SMEs to offer to their improvement. The special lab in which the research has carried the research detailed here is essentially involved with the growth of scientific expertise on the research and modeling of SMEs' performance and risk, including a type of interrelated subjects such as economics, management, data systems, product, technology, etc. Analysis plans carried out at our lab require both analytical and practical perspectives, usually compelling infield knowledge with SMEs. As a result, our analysis plans perpetually strive to present practical answers to real problems facing SMEs.

DOI:10.5121/ijitcs.2018.8201

2. THE SME EXHIBITION EVALUATION SOFTWARE

2.1 A SURVEY OF THE PERFORMANCE, DEVELOPMENT, AND GROWTH SYSTEM (PDG)

The Performance, Development, and Growth (PDG) software system, which works on computers discovered in the lab, assesses an SME from a visible prospect and on a comparable basis in order to perform a diagnosis of its achievement and potential, complemented with appropriate references. Although we normally connect to our system as a diagnostic system, it is, in fact, a hybrid diagnostic credentials system as it not only recognizes the -estimated SME's flaws but it also offers recommendations on how to approach these flaws in order to develop the SME's achievement[3]. A lengthy eighteen-page questionnaire is managed to obtain related information articles on the SME to be assessed. Data deduced from the questionnaire and the financial records are computerized and stuffed into the system. The recent makes a multidimensional evaluation in roughly 3 minutes by distinguishing the distinct SME with a suitable collection of SMEs for which we have previously received related data—this is the crux of the benchmarking method. The output is a completely description in 28 management practices. And this include among other things,which involve, the human resources management, production systems and organization, market development enterprises, accounting, finance and control tools as well, 20 results pointers and 22 overall information pieces are also assessed, that leads to 14 recommendations on short-term that moves the evaluated SME which inturn, could offer in order to increase its overall appearance. The system employs the SAS statistics package, plus Microsoft Excel for several calculations and the production of the last output statement. The output records are regularly observed by a team of multidisciplinary human specialists in order to ensure that recommendations are helpful for the entrepreneurs. This validation stage, which perpetually takes place before the report is given to the SME, is a chance to secure further advances in our system, whenever suitable. It is also a relevant means for the human experts to refresh their individual expertise on SMEs. An intermediary associate is part of the means in order to ensure confidentiality. A recent research was conducted by 307 Canadian manufacturing SMEs that have practiced the evaluation report, including 49 that have done so plentiful than once. Our results show that the specialist benchmarking evaluation enables these corporations to increase their operational display, proving the advantage of benchmarking but also, the value of the recommendations included in the output report regarding short-term effects to change management practices.

2.2 THE KNOWLEDGE IN THE ENGINEERING ASPECTS

The system's expertise is placed in two main segments: the in-depth questionnaire and the benchmarking outcomes analysis module. The first redaction of the questionnaire was generated by a multidisciplinary team of researchers in the following fields, thus providing to the multidimensional nature of the evaluation: business strategy, human resources, information systems, industrial engineering, logistics, marketing, economics, and finance. The questionnaire advancement team was confronted with two significant difficulties that immediately became

Essential goals:

i) find a common language (a shared ontology) that would permit researchers to know each other and, at the same time, would be available to businesspeople when answering the questionnaire, and

ii) recognize long-term achievement pointers for SMEs, as well as difficulty indicators, while enduring contents to a minimum since in-depth evaluation was not enough.The team was capable to reach these two goals by the designation of a "knowledge integrator" performance to the design head. During the 15-month term of its growth, the questionnaire was examined with

businesspeople in order to guarantee that it was easy to surmise in terms of both contents and question formulation and record design and information visualization. Various models were introduced to businesspeople and they displayed a branded interest for graphics and colours.The researchers' expertise was fragile in the description of important erudition that would permit the system to quickly create a general analysis of any manufacturing SME[4]. The investigation also required to be stable and flawless, while being understandable to ordinary entrepreneurs as we looked out before. This was pioneering research work that the entire team was accompanying. Certainly, additional SME diagnosis systems are usually financial and based on real quantitative datafor instead of master systems in finance. The knowledge integrator stated earlier played a major part in this learning engineering and integration rule. Each expert had to recognize training, systems, or devices that had to be performed in a manufacturing SME to guarantee a reliable level of achievement. Suddenly, achievement indicators had to be established in order to estimate to what degree these personal practices, systems, or tools were perfectly performed and enabled the business to reach particular purposes—the association between practices and outcomes is a distinctive feature of our system. Subsequently, every chosen achievement pointer was specified a corresponding weight by the specialist and the knowledge integrator. This weight is applied to form the business being diagnosed with respect to its source group, hence enabling the generation of relevant observations and suggestions. The weight is also employed to create a global evaluation that will be presented in a synoptic table. Contrary to many achievement diagnostic tools in which the business's information is corresponded to patterns and standards the system estimates an operation relative to a source group chosen by the businessperson. Analysis managed by many institutes precariously challenges this use of patterns and standards: it seems to be questionable for SMEs as they completely are too complex to maintain the interpretation of reliable patterns and standards.

Achievement indicators are performed as variables in the system—more accurately in its database, and in the benchmarking outcomes analysis module. These variables are explained in phases of three classes:

- Binary variables, which are linked to yes or no questions;

i) scale variables, which are linked with to the corresponding ranking of the business onward with a 1-5 or a 1-6 scale, depending on the question; and lastly
ii) Continuous (which is also called numerical) variables, which are linked with numerical values such as the export rate or the training budget.

Since variables come in many varieties, they must also be prepared separately at the mathematical level, prominently when calculating the source group employed for benchmarking goals. In line to define the source group with a single value, a central tendency measure that is the agent of the source group's set of measurements is practised. Depending on the variable level and its statistical distribution, means, medians/percentages are utilized in the benchmarking computations[5]. The estimated enterprise's outcomes are outranked and connected with codes that will next be applied to design the several graphics in the benchmarking record. The resulting codes show the estimated enterprise's benchmarking outcome for every achievement indicator. They are then utilized by the report formation module to perform the benchmarking output report, which comprises several coloured graphical illustrations, as well as observations and suggestions. The output report comprises essentially coloured diagrams and simple illustrations that are expressed in plain English/French so that SMEs entrepreneurs can quickly recognize it. The system also does some comparatively old AI methods. For example, the observations provided in the output report are created through a template-based strategy, an early method applied in natural language processing.

3. THE REQUIREMENT FOR GREATER DATA INTELLIGENCE ABILITIES

Our system is presently at a staging where we can now review the introduction of AI techniques in current advancements. We have started to develop innovative modules that will enhance yet more the intelligence features of the system:

a) Improvement of data warehouses and data mining algorithms to promote statistical processing of data and enlarge knowledge uprooting abilities. Such extracted knowledge will be helpful to advance the systems' meta-knowledge level, which could be utilized in the systems' descriptions for example and also to increase human experts' region knowledge. This state is previously in growth and is the principal focus of our prevailing work[6]

b) a large number of database resources and statistical variables managed by the system is amazing. A conceptual taxonomy, copulated with a developed data encyclopaedia, has rapidly grown a necessary addition. For example, the researcher should be ready to get out immediately to what thoughts a circumstantial attribute/variable is connected, to what computations or results it is described, and so on. This condition has newly introduced.

c) Improvement of an expert system to reduce the demand for any human interference in the system. Presently, a human expert must revise all automatically-produced reports before they are sent to the SME. Most of the moment, only minor modifications are needed. The knowledge applied to make this final amendment drives into deliberation individual results provided in different portions of the benchmarking record and examine potential consequences of interrelationships between them in order to ensure that conclusions and recommendations of the appraised SME are both correct and consistent. This is the component of the future work.

d) Augment the PDG system with case-based logic and associated machine learning algorithms In various features of the system, evaluation of the difficulty at hand could be aided if it were reasonable to set associations with related intricacies previously explained earlier [6,7]. Planning the difficulties' striking characteristics to maintain this strategy would further propose sufficient potential to reduce the users' responsibility through the primary data-gathering phase. This state is the portion of the scheduled work.

e) Investigate the potential of tool technology to re-engineer any components of the system, particularly from a resolution support system view [8]. This could be particularly appealing for the modelling and implementation of shared sources of expertise that provide to decision processing. For example, in the system, each source of expertise in the performance evaluation of an SME could be linked with a different agent regulating and accomplishing it is personal information field [8,9]. Interplay and coordination among real agents would be important features of a new system based on a society of collective problem-solving tools.

4. DATA WAREHOUSING AND DATA MINING

4.1 MOTIVATIONS

Our system has pointed to the production of a huge database on several separate SME-related motives. This database is regularly modernized with numerical, qualitative as well as financial historical (yearly) data from evaluated SMEs. The database was invented to promote the generation of the benchmarking statement, but further to encourage scientific investigation on SMEs carried here by researchers and bachelor students. These two certainties, seldom connected with contradictory goals, have created fundamental difficulties in the database itself, more

particularly in the database uploading ("feeding") device and in the applications that manage the data. The scheme means was incremental over several years and we must maintain that, while the project began five years ago, we nevermore envisioned that it would improve to such a scale.

For that purpose, there is no novel, consistent viewpoint behind the prevailing database formation. There are numerous tables and various software agents to saddle and extract data from the database. As a result, preservation of the system is hard, mostly because of the system's inner complexity and the numerous kinds of use we do have to support. Despite these difficulties, our system and its database create a highly valuable reference of information on SMEs. Till presently, only one expert user begot adequate effective understanding of the database to accurately deduce and provide data for planning-related requirements. This person was regularly requested for the whole lab's data requirements. Some researchers manage the database. There are also research professionals and graduate students that require all sorts of data sets—data collections are data uprootings from the database that are specifically developed to investigate an SME-related condition or problem. So far, all these people ought to question the same person for a data collection tailored to their requirements, and developing a single data collection could practice as enough as a week. As these users arrive from diverse settings, they manage to have distinct aspects of the data. Although the primary producers and developers of the system had established a general philosophy, several of them have gone after. All new users hold to be shaped before being capable to appropriately use the least share of the system, and this practice is a difficult duty. To maintaining the norms of incremental growth, we simply taken data warehousing as well as the data mining methods in order to adequately utilize the valuable knowledge held in the database. Hence, the system will adequately satisfy the requirement for assimilation of different data references, to fully satisfy the requirement for a metadata repository, and for the overall advantage of the different mechanisms that are previously available in a DW helped with a database environment before-mentioned as Oracle[10]. The appearance of sparse systems, each providing its personal data to the database, executes it extra challenging to utilize the data without a sufficient procedure. Synthesis of all required references of data must be performed for each distinct project. Amidst a plenty of work, it is reasonable to hold current applications going and to generate new schemes. However, with data warehouse, all entreaties will serve from the associated selection, change and load (ETL) methods. The appearance of blended data decreases the time and quantity of data guidance. The number of mistakes due to guidance is also decreased because the integration is performed in one convenient position for an efficiency of perception and data reusability. The different obstacle that we encounter for each current design is the lack of a consolidated reference of documentation. Our clarification is to generate a repository for metadata in the DW. This will enable users to reach metadata regarding "variables". Most of this documentation previously live, however, the users have to tortuously explore into many records to find what they require. This is why all these records are being streamlined and coordinated in the metadata repository. Conclusively, many techniques are now possible to efficiently employ data in a DW, such as DM devices. After the database is quite extensive, researchers normally perform data collections from a comparatively small subset of the database. The application of data warehousing can possibly practice all accessible data and recognize connections among various issues. For instance, there are several common problems that data warehousing methods will enable us to discuss simply. Is there a connection between the change in an SME and its financial situation? Is it important to have business partnerships to improve sales figure? What affects sales figures? What changes partnerships? Why do these SMEs invent higher than these distinct ones? All the data in the DW will be employed with DM purposes to answer arguments interrogations, and others, and probably reveal earlier hidden guides, laws, events, etc, enabling researchers to gain extra information on SMEs and, ultimately, to create a balanced proper benchmarking evaluation of SMEs. This explains very thoroughly the study-application sequence in which our activity took place. A significant obstacle that we encounter when assessing these methods was the evidence that the uploading rules were created to assist the inclusion of data in

the database and to produce articles. The transcription of the database was not quite well filled with the control of data warehousing devices and applications. This combination had to be modified. For that purpose, we produced an ETL state to generate an intermediary database more rightly adjusted to the requirements of the lab's users and researchers. Moreover, this technique provides much more simply the combination of several references of information, a distinct plus in our study environment.

4.2 FIXING UP THE FOUNDATIONS

To fix up the data warehouse, we need to initial blend all data into a combination form: the extract, transform and load (E.T.L.) modes—see a depiction of our DW foundation in Figure 1 below. Longitudinal investigations were confused with the unconventional database. For example, noknowledge was held as to when or why specific data components were replaced. In the data warehouse, timestamps are attached to the data at loading time and historical information is connected with the data. Most of the data in the system develop from inquiries that are transmitted to our lab where solutions are transcribed in. Seldom, an error (e.g typo) is injected at entrance time. This is why a means of manual confirmation was performed numerous years ago. This method applies banners to indicate to the system when distinct data can be deemed secure, once old-fashioned (manual) verification has been performed. These flags require being prepared by the ETL schemes so that none "unreliable" data is applied in the DW—as, in all DW efforts, data ought to be "clean" as many as potential. This original form is also suitable to renew and achieve the metadata documentation. The arrangement of the tables is investigated and transformed to provide an accessible path for all foreseeable applications of the data. A security rule also guarantees that users are given enough path to the DW. The benchmarking record will shortly be redesigned to avail of the new opportunities given by the DW. An urgent help will be a strong decrease of the time needed to create the report. Currently, it requires about 3 minutes to automatically generate a benchmarking report. This comparatively plentiful time is illustrated in part by the necessary to change all data in the same money (the available data are in US dollars, Euros and Canadian dollars). The complexity of the tables in the database performs it remarkably tough to implement these transformations, and the production of a developed view with a quick refresh was not likely.

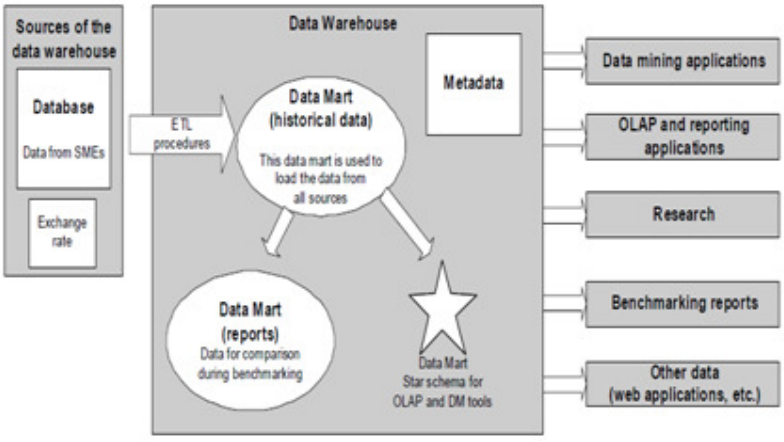

Figure 1. A description of our data warehouse structure

There are other alterations that are also required. One of them is owing to the circumstance that data regarding a particular SME can be administered in several questionnaires;the data has to be adjusted. This has to be done every time a report is requested. The particularities of the database lead to some compromises in achievement. Nevertheless, in the DW, all of these reflections are reaped care of by the ETL procedures. The data is stored in a conventional and much more useful

form responsive to report generation and data mining for investigation targets. Once the DW is developed, we anticipate a conception time of around 30 seconds for benchmarking reports. Besides, various modern means, some of which will be AI-established, will be built possibly to enhance the usefulness of the report. These are some of the numerous advantages that DW draws to our system. The DW sustains a class of mechanisms that enable users an urgent access to the data apparently without the aid of a database expert. We have also formed a custom-formed, user-friendly Web-based mechanism that permits users to generate datasets from the DW. Metadata also enable users to perfectly translate data in the repository. One of the metadata devices is a dictionary of "variables" (that is the questionnaire fields). Aforementioned is a precious device for users considering there are hundreds of variables, and several variables have style arrangements. Also, as discussed earlier, many modifications have happened across time in the questionnaire and in the database. Likewise now, our system remains to develop and new questionnaires are designed. This is why it is so essential to know what a variable describes precisely because even this interpretation can evolve over time, for instance, a problem with three potential answers (A, B and C) in a preceding questionnaire can presently have five answers (A, B, C, D and E). The literature on DW administers with several of these questions, including methodological aspects [11]. There are various ways as to how the centre of the DW should be created. Some practitioners say that using a multidimensional database is the genuine way to boost attainment [10,11], while others say that adopting such a configuration can be the case of great misery if adjustments have to be performed [12]. Furthermore, we preferred to execute a multidimensional data mall based on a player schema. It will be practised for On-Line Analytic Processing as well as DM devices. It is of this data mall that we will be capable to investigate the system's data from a new view. Our database is huge, but not large. Our popular sources are enough to manage the size of data, and we also have a bit of additional volume if necessary. The benchmarking reports could still be created from the fundamental relational database. Although, the multidimensional data mart is of noble regards to the operation in step on data warehouses and data mining. Not all data warehousing devices inevitably require a multidimensional database to work still. We are notwithstanding very positive that this data warehousing and data mining strategy we have exerted lately will cover the way to a current, accurate understanding of many significant phases of SMEs as they are (essentially) modelled in our database. As quoted briefly, the researchers regularly handle simply a subset of all the possible data in their data collection. They examine their beliefs on certain specimens utilizing their expertise to get different points. Nevertheless, by applying data mining tools, they can now instantly examine their hypotheses upon the complete data warehousing[13]. We can require users to be involved in inquiries that are currently needlessly complex to measure with the original database, such as: Since their current evaluation, what changed in the SMEs that adopted our record? What sort of SME judges to get a global accreditation? These are yet a few instances,where various inquiries that will be accessible to prepare with data mining procedures employed to our data warehousing. In the system, there is also a need for different information statements about the data in the data warehousing. The treatments presently employed to produce these records are not accessible to produce and modernize for non-programmers as they are executed in PL/SQL. Usually, so news records have a very precise time of life. For instance, a record displays the topographical pattern of the SMEs we have in the database. Other records confer the area in which the SMEs explore, while others display the date at which the questionnaires were collected at the lab. Seldom, there may be a requirement to identify precisely how numerous benchmarking reports were transferred per month in the USA alone. This sort of report is simple to get with an OLAP application. OLAP applications bequest practice the multidimensional data mart, though will additionally have admittance to all extra data in the DW. Remarkable experimentation has previously been created with the database and technoscientifically algorithms such as automated cluster detection[14]. The development of data collections was challenging as a whole, and the data could not be renewed efficiently.

5. CONCLUSION

We have succinctly shown a completely realized skilful diagnostic system which assesses the achievement of SMEs on a benchmarking footing. The system has remained in practice for many years and has operated within a consistent and considerably challenging growth in series to satisfy both the requirements of SME-oriented analysis and the generation of benchmarking reports for SMEs (nearly 500). We have developed to a case where data warehousing and data mining procedures had to be added, regarding the advantages they contribute to help our work, prominently our investigation on SMEs and, consequently, the performance evaluation conducted out by our system. At the moment of review, we are about to initiate the current data warehouse and begin our first experimentations with data mining methods. Data warehousing has several efficient techniques in our SME-oriented setting and it will, simultaneously with data mining methods, undoubtedly change the way researchers practice the rich data we have received and proceed to gather on SMEs. We expect to significantly enlarge our information on SMEs, and farther develop our evaluation model of SME achievement.

REFERENCES

[1] Bamberger S.K. (1997), "Cooperating Diagnostic Expert Systems to Solve Complex Diagnosis Tasks," Lecture Notes in Computer Science #1303, 325-336.

[2] Bui T. and J. Lee (1999). "An Agent-Based Framework for Building Decision Support Systems", Decision Support Systems, 25, 225-237.

[3] Cassell C., S. Nadin and M.OlderGray (2001), "The Use and Effectiveness of Benchmarking in SMEs", Benchmarking: An International Journal, 8(3), 212- 222.

[4] Denkena B., R. Apitz and C. Liedtke (2003), "Knowledge-Based Benchmarking of Production Performance", Proceedings of the Business Excellence (Performance Measures, Benchmarking and Best Practices in the New Economy) Conference, Guimarães (Portugal), June 10-13 2003.

[5] Golfarelli M and S. Rizzi (1998), "A Methodological Framework for Data Warehouse Design", Proceedings of the First ACM international Workshop on Data Warehousing and OLAP, 3-9.

[6] Hastie T., R. Tibshirani and J. Friedman (2001), The Elements of Statistical Learning (Data Mining, Inference, and Prediction), Spinger

[7] IMT Strategies Inc. (1999), "The Sales and Marketing Imperative: The Impact of Technology on Business Strategy", Data Warehousing: The Ultimate Guide to Building Corporate Business Intelligence, SCN Education B.V., 15-26. Kimball R. (1996), The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses, Wiley.

[8] Matsatsinis N.F., M. Doumpos and C. Zopounidis (1997). "Knowledge Acquisition and Representation for Expert Systems in the Field of Financial Analysis", Experts Systems with Applications, 12(2), 247-262.

[9] Mitchell T.M. (1997), Machine Learning, McGraw- Hill.

[10] Nedovic L. and V. Devedzic (2002), "Expert Systems in Finance—A Cross-Section of the Field", Expert Systems with Applications, 23, 49-66.

[11] O'Neil B., M. Schrader, J. Dakin, and Northern Lights Software Ltd. (1997), Oracle Data Warehousing Unleashed, SAMS Publishing.

[12] Rouge A., J.Y. Lapicque, F. Brossier and Y. Lozinguez (1995). "Validation and Verification of KADS Data and Domain Knowledge", Experts Systems with Applications, 8(3), 333-341.

[13] Schreiber G., H. Akkermans, A. Anjewierden, R. de Hoog, N. Shadbolt, W. Van de velde and B. Wielinga (2002), Knowledge Engineering and Management:. The Common KADS Methodology, MIT Press.

[14] Stamelos I. and I. Refanidis (2002), "Decision Making Based on Past Problem Cases", LecturNotes in Artificial Intelligence #2308, 42-53.