Integrating Artificial Intelligence into Data Warehousing and Data Mining

Nelson Sizwe. Madonsela, Paulin. Mbecke, Charles Mbohwa

Abstract— **Knowledge engineering is key for enhancing organizational capabilities to gain a competitive edge and adapt and respond to an unpredictable market environment. Such knowledge can be generated from collected data which is often considered complex. Organizations are collecting vast amounts of data to transform them into real-time information in order to attain successful decision-making support systems. It is more than likely that such processes can be challenging; yet such knowledge must be extracted from thoughtfully designed and implemented data warehousing, and mined to obtain the required information. This paper explores appropriate techniques, technologies and trends to facilitate the integration of artificial intelligence into data warehousing and data mining. It provides an insightful overview of data warehousing and data mining, and it highlights the techniques and the limitations of analyzing and interpreting enormous data.**

***Index Terms*— artificial intelligence, data warehousing, data mining, knowledge discovery, business intelligence.**

I. Introduction

Information in the 21st century has become the main source of gaining competitive edge. [4, p.1] are of the opinion that there is "nothing new under the sun", since there are "lots of old things we don't know".

Manuscript received 21 March 2015; revised 03 April 2015
Nelson Sizwe Madonsela is a Masters student at the University of Johannesburg (UJ), Faculty of Engineering and the Built Environment (FEBE). He works for the Department of Justice and Correctional Services as an IT trainer and the Department of Higher Education as a Part-time Lecturer. Tel: +27721345265, Email: nel_madon@yahoo.com
Prof Paulin Mbecke is currently a postdoctoral research fellow in the Department of Public Administration & Management, School of Economic & Management Sciences at the University of South Africa and a visiting Associate Professor in the Faculties of Political, Administrative and Social Sciences (Department of Community Development) and Economic and Management Sciences at the University du Moyen Lualaba (UML) in the DRC. Email: pmbecke@gmail.com
Prof Charles Mbohwa is the Vice-Dean: Postgraduate Studies, Research and Innovation at the University of Johannesburg (UJ), Faculty of Engineering and the Built Environment (FEBE). He serves as a Professor in the Department of Quality and Operations Management. Email: cmbohwa@uj.ac.za

Firms are collecting vast amounts of data daily and developing advanced data warehouse systems to secure the data. Their intention is transforming it into vital information, or knowledge, for developing decision support systems (DSS). According to [10] there are still challenges with regard to the techniques of analyzing and interpreting the exact meaning of enormous data, in fact, to integrate artificial intelligence (AI) into data warehousing. It is much more likely that critical information is often overlooked or not tapped into from these vast amounts of data, while organizations invest huge amounts of financial resources on collecting, storing and securing the same data. [5, p. 27] emphasize that:

> "It is crucial for business to acquire a better understanding of the commercial context of their organization, such as their customers, the market, supply and resources, and competitors. Business intelligence (BI) technologies provide historical, current and predictive views of business operations. Examples include reporting, online analytical processing, business performance management, competitive intelligence, benchmarking, and predictive analytics."

To take a case in point, there is clear evidence that, it is imperative to establish such a broad and in-depth understanding of the incorporation of concepts like data warehousing, data mining, AI and BI, to name but few. [10] maintains that is difficult to incorporate AI into data warehousing or BI. Even so, many researchers strongly believe that for business to gain competitive advantage, the usage of data warehousing is key [13], [10], [9]. However, [11] stresses that there are constraints such as the understanding of the required information, organizational culture and funding. These authors further emphasize that a data warehouse should not be viewed as a "static database"; but rather as "a dynamic decision support framework that is, almost by definition, always a work in progress" [11, p. 652]. Consequently, establishing a broader understanding of data warehousing and its basic elements is necessary to address all issues in relation to data warehouse designing and implementation. Thereafter, the techniques and technologies of integrating AI into data warehousing can be incorporated. This paper explores appropriate techniques and technologies to facilitate the integration of AI into data warehousing. The important objectives are as follows:

- To establish the techniques of analyzing and interpreting enormous data.

ISBN: 978-988-14047-2-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

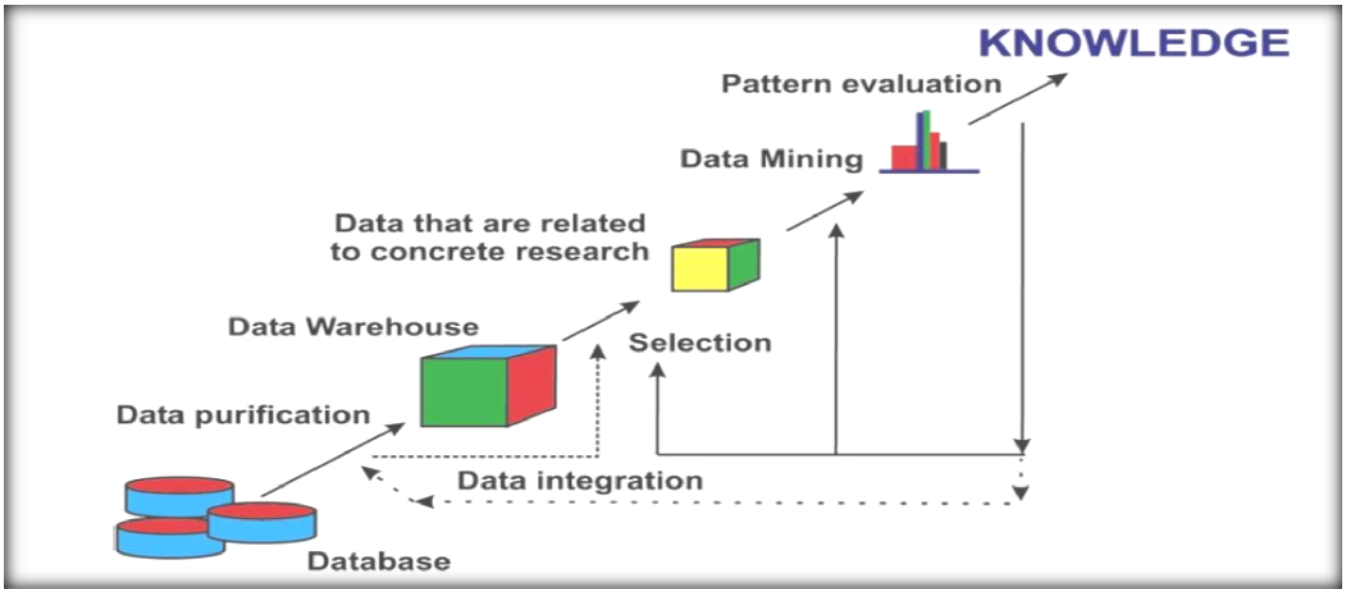

Fig 1, Process of Knowledge Discovery **Source:** Joseph (2013)

- To acquire ways of integrating AI into data warehousing.

It is important to address these essential basic elements of data warehousing, since it is considered to be a fundamental (DSS) tool. An organization with insufficient information is likely to make inadequate decisions that can hinder its success within the marketplace. Thus in-depth understanding of the techniques of analyzing and interpreting huge data into information is imperative to the establishment of techniques and technologies to integrate AI into data warehousing.

II. AN OVERVIEW OF DATA WAREHUSING AND DATA MINING

It is necessary to delineate the difference between data warehousing and data mining since these concepts are interlinked. "Reference [8, p. 329] states the combination of data warehousing and data mining technology has become an innovative idea in many business areas through the automation of routine tasks and simplification of administrative procedures". According to [10, p. 5], a "data warehouse is a database that collects and stores integrated data from several databases, usually integrating data from multiple sources and providing a different way of looking at the data than do the databases being integrated."

The author of Fig 1 has granted us permission the use the graph. [2, p. 2] describes data mining as a "process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques." This implies that data warehousing entails the integration of data from different resources, while data mining processes the data to meaningful information for DSS, and [8] states that data mining is a process of knowledge discovery. (See Fig 1).

Fig 1 provides a clear picture of the relationship between data warehousing and data mining since these concepts are more often confused as being interchangeable. In this regard, the authors are of the view that a simply understanding of data warehousing and data mining has been established: therefore, it is important now to acquire adequate techniques for analyzing and interpreting enormous data and to highlight the implications.

III. THE TECHNIQUES OF ANALYZING AND INTERPRETING ENORMOUS DATA

According to [6], the market environment is continuously changing and demands adequate decision-making from economists - which depends on the application of information technologies. Integrating data and extracting knowledge from the market environment is always complex; in fact, this requires sufficient modeling techniques. [5, p. 5] maintain that the world "is data rich and information poor", due to the vast amounts of data which are collected but not transformed into information. If this is the case, it could be due to failure to use adequate techniques and technologies to analyze and interpret such data. Thus the data mining implementation phase is critical since there are several techniques that can be considered, especially with the incorporation of AI. To look at different experts' proposed approaches through an AI lens, [6, p. 69] argue that AI "deals with problems of classification, prediction and optimization incorporating processes that can be called intelligent in decision making etc, where the problems cannot be simply formalized", in fact, "AI systems are designed to adapt and learn". From a business perspective most theoretical economists [6], [1], [12], [10] state that there is an increasing application of AI into data mining analysis. [14, p. 25] declares that there are three AI categories with regard to data mining, namely:

1) **"Knowledge representation.** Data mining seeks to discover interesting patterns from large volumes of data. These patterns can take various forms, such as

ISBN: 978-988-14047-2-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

association rules, classification rules, and decision trees, and therefore knowledge representation becomes an issue of interest in data mining, for instance, trend discovering.

2) **Knowledge acquisition.** The discovery process shares various algorithms and methods with machine learning for the same purpose of knowledge acquisition from data, or learning from examples, for instance, inductive logic and decision trees.
3) **Knowledge inference.** The patterns discovered from data need to be verified in various applications and so deduction of mining results is an essential technique in data mining applications", for instance, prediction or forecasting.

These three techniques are considered the most common ones associated with AI. This paper established that the most important factor is the integration of artificial intelligence into data warehousing or rather the methodologies that embody AI perspectives. [6], [13] observed the following applications (machine learning tools) in relation to AI: Expert (knowledge) systems (ES), Artificial Neural Networks (ANNs), Genetic Algorithms, Fuzzy Systems and Agent-based Computational Economics (ACE). Machine learning tools are becoming popular for data mining, based on an ability to automatically learn to observe patterns from past events or experiences, and make intelligent decisions based on these observations. These applications are appropriate for specific domains; therefore an organization must ensure that it applies the relevant application. Failure to do so could result into incorrect information being collected and wrong decisions taken. Since AI systems have the ability to learn and adapt to uncertain environments, this is exactly what organizations need to rapidly respond to the unpredictable global market. [6] recommend ANNs (Artificial Neural Networks) as being the appropriate technique. They claim that ANNs are capable of functioning in a vague environment and have the "ability to study the nonlinear relation between variables" (p. 72). [5, p. 24] urge the use of machine learning which results in an ability to "automatically learn to recognize complex patterns and make intelligent decisions based on data". Clearly, there is constant knowledge being discovered regarding data warehousing and data mining methodologies, technologies and tools which makes it difficult to itemize them all. Nonetheless, all these applications incorporate the element of AI and the ability to extract knowledge from vast amounts of data. At this point, the authors would like to declare that this paper subscribes to [6]'s view of ANNs technique of analyzing and interpreting enormous data. These authors observed that ANNs has been applied for economic forecasting and has proved to be accurate for both short and long periods according to the key performance indicators. [7, p.43] explains that ANNs:

> "are non-linear data driven self adaptive approaches as opposed to the traditional model based methods. They are powerful tools for modeling, especially when the underlying data relationship is unknown. ANNs can identify and learn correlated patterns between input data sets and corresponding target values. This feature makes such a computational model very appealing in application domains where one has little or incomplete understanding of the problem to be solved but where training data is readily available".

To put it another way, the critical point is that ANNs possess an element of AI in relation to their ability to adapt and learn from the unknown environment and predict the future, which is exactly what is needed in the world of knowledge engineering to gain competitive advantage. [7, p. 42] states that NNs (Neural Networks) exhibit, or embody the following features:

- "The NNs exhibit mapping capabilities, that is, they can map input patterns to their associated output patterns.
- The NNs learn by example. Thus, NN architectures can be 'trained' with unknown examples of a problem before they are tested for their 'inference' capability on unknown instances of the problem. They can therefore identify new objects, although previously untrained.
- The NNs possess a capability to generalize. Thus, they can predict new outcomes from past trends.
- The NNs are robust systems and are fault tolerant. They can recall full patterns from incomplete, partial or noisy patterns.
- The NNs can process information in parallel, at high speed, and in a distributed manner."

[7] illustrates exactly what is needed to approach the unpredictable market environment, since making decisions in an uncertain environment can be detrimental, with organizations making executive decisions based on incorrect forecasting. According to [12], NN architectural structures are called neurons, which illustrate how the ANNs model (perception) addresses classified challenges. In brief, "a perception computes its inputs value, $\hat{y}$, by performing a weighted sum on its inputs, subtracting a bias factor t from the sum, and then examining the sign of the result" [12, p.247]. ANNs techniques are "more computerized" and have a "higher degree of accuracy" in terms of long term forecasting, as it has been proven in Tokyo and United States of America [6]. Notwithstanding, it is arguable that integrating AI into data warehousing and data mining still has limitations, especially when it comes to analyzing and interpreting enormous data. ANNs applications have limitations when it comes to recognizing the input variable's impact; and that, with all applications, the degree of accuracy will not reach 100%. [6] argues that the constant changing market environment demands accurate real-time decision-making processes, and "decision making with AI support can partly eliminate the bounded rationality of a decision maker to make better decision with more relevant data and information [6, p. 62]". This implies proper AI application since there are different divisions, in fact, selecting the right application for the exact problem. Researchers noted that organizations lack ability to identify

ISBN: 978-988-14047-2-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

the right technology to address the encountered problem, which consequently makes the problem more complex [15]. Early on [6] highlighted a few AI applications that are often applied within the financial markets and therefore it is imperative that an organization establishes the relevant AI application. It is important to mention that adopting AI applications requires a skilled workforce with ability to apply the AI application. Interestingly, it is likely that the adoption of AI application is perceived as its ability to address the inadequacy of human expertise regarding the integration of AI into data warehousing and mining. However, there is no doubt that humans cannot be eliminated within these processes. This is made evident by [16, p. 1] who suggests that there is a gap in relation to AI transformation from "computational discipline" into "highly transdisciplinary". These authors argue that today there is a misunderstanding of AI; in fact, AI is seen as an algorithms rather to establish the meaning part of the intelligence. In other words, it is possible that this could be the reason why organizations have not taken advantage of AI to accomplish competitive advantage. Furthermore, they claim that literature suggests that the AI perspective has been inadequately explored recently comparing to the 60s. Therefore, we profoundly concur with [4] when they maintain that there is information that has not been discovered yet.

The controversial issue has been whether AI or humans is the answer to discover such information. Moreover, literature indicates some limitations on both AI and humans' expertise regarding unpacking the exact understanding of "perception, locomotion and manipulation [16, p. 3]". It seems that both sides need to be investigated further and also to reconsider that these issues can be addressed by combining both aspects. In addition, we should agree that both aspects are complementary and inseparable, as [16, p. 3] pointed that AI systems cannot handle "common-sense knowledge". In this paper framework, our argument is that to gain a competitive edge and adapt and respond to an unpredictable market environment, organizations need knowledge engineering. This can be attained through the application of appropriate techniques and technologies to facilitate the integration of artificial intelligence into data warehousing and data mining. We acknowledge the limitations that have been pointed out by other researchers [6], [16]. Although this is true, the usage of AI has been reported as a success in many sectors, especially the financial markets, in relation to forecasting. In addition, there is a degree of accuracy in AI when it comes to decision-making compare to the human aspect. [17, p. 14] provides a broader understanding of AI and data mining (DM) by declaring that "AI & DM enables us to process, model, and use real-time data streams, build accurate prototypes of sophisticated reservoir-simulation models". This also reaffirms the application of AI in different area of our life in terms of real-time prediction. The aspect of the reservoir-simulation model is vital to operations management decisions since its decision making relies heavily on real-time feedback in order for any actions to be taken. In terms of human analysis, there is a good possibility that this might take days, and by then the damage could have already occurred. Recently in South African we have seen two manufacturing companies plantation experiencing a devastating event because of the flow of fluids (gas), which resulted on fire. In brief, the incorporation of AI is imperative to prevent any unpleasant situations in an organization which could be financially disastrous.

Organizations have invested in data warehousing and data mining, but there is a possibility that the same data has not been adequately used. The emphasis is that to address this perceived inability requires the integration of AI into data warehousing and data mining. Adequate literature has proven the ability of AI regarding environment prediction in different areas of our real world. In addition, decision makers cannot adequately interpret the vast amounts of data that are collected daily to make accurate and real-time decisions. Wrong decisions might be taken which might lead to disastrous consequences. To summarize, the advancement of technologies has created a new era of artificial intelligence. This addresses challenges that are beyond human beings' understanding, while, trying to provide some degree of understanding why things occurring the way they do. In a business context, organizations cannot forecast the environment markets they operate in on their own without the use of AI.

IV. CONCLUSIONS

Knowledge engineering in the 21st century is critical given the huge sums of money invested in business. In business contexts, decision-making requires rich information that can be attained through decision-making support systems. AI has the ability to extract information from the data. Failure to implement such systems can result in organizations making executive decisions based on false alarm prediction, with repercussions that might be detrimental and irreversible. The paper emphasizes the appropriate approach toward designing and implementing AI into data warehousing and data mining. Evidence indicates that there are challenges in relation to knowledge discovery from data, due to lack of application of AI techniques to analyze and interpret the vast amounts of collected data. Notwithstanding, AI techniques that have the ability to address these challenges were highlighted. In fact, it was put forward that these techniques are applicable as per domain since each has been developed to address problems of a specific domain: for example, financial domain, health domain and agricultural domain.

ACKNOWLEDGEMENT

N. S. Madonsela thanks Dr Neil Barnes for editing the manuscript and the co-authors for their inputs. In addition, I thank my peer Mr. Kudakhwashe Mushore and Mr. Adelino Chissale for their support. Lastly, the peer review team for their positive feedback, which allowed us to refine the paper.

REFERENCE

[1] A. Fred, L.G. J. Dietz, K. Liu, J. Filipe. Knowledge discovery, knowledge engineering and knowledge management. First International Joint Conference, IC3K, 2009. Funchal, Madeira, Portugal, October

2009, revised selected papers. Springer. 2011, New York.
[2] D. T. Larose. Discovering Knowledge in Data. An introduction to data mining. New Jersey, 2005.
[3] P. Tan, M. Steinbach, V. Kumar. Introduction to data mining. 2006. Boston: Pearson Addison Wesley.
[4] G. S. Linoff, M.J.A. Berry. Mining the web. Transforming customer data into customer value. 2001.
[5] J. Han, M. Kamber, J. Pei. Data mining. Concepts and techniques. 3rd ed. United States of America.
[6] K. Hilvska, P. Knocz. "Application of Artificial Intelligence and Data Mining Techniques to Financial Markets", ACTA, vol. 6, no. 1, 2012.
[7] G. Kumar. Artificial neural networks and its applications, 2007, Available: http://www.iasri.res.in/ebook/EBADAT/5Modeling%20and%20Forecasting%20Techniques%20in%20Agriculture/5-ANN_GKJHA_2007.pdf Accessed date 05 March 2015
[8] M. V. Joseph, "Significance of data warehousing and data mining in business applications", International Journal of Soft Computing and Engineering (IJSCE), New York vol. 3, no. 1, pp. 329-333, Mar. 2013. .
[9] S. Kudyba. Managing data mining. Advice from experts, 2004. United States of America.
[10] Z. Chen. Intelligent data warehousing: From data preparation to data mining. United States of America, 2002.
[11] P. Rob, C. Coronel. Database Systems. Design, implementation and management, 2002, United States of America.
[12] P. N. Tan, M. Steinbach, V. Kumar. Introduction to data mining, 2006. Pearson Education, Inc.
[13] B. Thuraisingham. Data mining. Technologies, Techniques, Tools, and Trends, 2000. United States of America.
[14] X. Wu. "Data mining: An AI perspective", IEEC Computational intelligence bulletin, vol. 4, no. 2, December 2004.
[15] R. Abdullah, M. G. Hassan. "Advanced Manufacturing Technology: The Perceived Impact of Producer's Value". The 3rd International Conference on Technology and Operations Management. Bandung-Indonesia, 4-6 July, 2012, pp. 23-30
[16] R. Pfeifer, F. Idia. "Embodied Artificial Intelligence: Trends and Challenges. Available: http://people.csail.mit.edu/iida/papers/PfeiferIidaEAIDags.pdf Accessed date 05 March 2015
[17] S. D. Mohaghegh. "Artificial Intelligence and Data Mining: Enabling Technology for Smart Fields". vol. 5, no. 3. 2009

BIBLIOGRAPHY

Nelson Sizwe Madonsela is a Masters student and Senior Tutor at the University of Johannesburg (UJ), Faculty of Engineering and the Built Environment (FEBE). His research addresses the impact of information and communication on service design and operation at ArcelorMital South Africa. He earned a diploma in Information Technology (Software Development) from the Tshwane University, South Africa. In addition, he earned his bachelor's degree in Quality from the University of South Africa. Before joining the Faculty of Engineering and the Built Environment, Nelson served as IT trainer for the Department of Justice and Correctional Services and the Department of Higher Education as a Part-time Lecturer. He is the author of the best paper session from the Planetary Scientific Research Organization, 2013. ***Entitled: Improving the South African Manufacturing Sector through the adequate usage of ICT.***

Prof. Za-Mulamba Paulin MBECKE worked as Senior Researcher and Senior Project Manager from July 2002 to September 2012 for the Crime Prevention and Safety and Security Research Groups at the Council for Scientific and Industrial Research (CSIR). Prof Mbecke is currently a postdoctoral research fellow in the Department of Public Administration & Management, School of Economic & Management Sciences at UNISA and a visiting Associate Professor in the Faculties of Political, Administrative and Social Sciences (Department of Community Development) and Economic and Management Sciences at the Université du Moyen Lualaba (UML) in the DRC. Prof. Mbecke is a contract lecturer at the Universities of the Witwatersrand and of Johannesburg where he supervises students in the Masters programmes in Management – field of Security (Wits) and Masters of Technology – Quality and Operations Management (UJ). He is fluent in both English and French.

Prof Charles Mbohwa is Vice-Dean: Postgraduate Studies, Research and Innovation at the University of Johannesburg's (UJ) Faculty of Engineering and the Built Environment (FEBE). He assumed duty on 1 July 2014 and also serves as a Professor in the Department of Quality and Operations Management, which forms part of FEBE's School of Mechanical and Industrial Engineering. As an established researcher and professor in the field of sustainability engineering and energy, his specializations include sustainable engineering, energy systems, life cycle assessment and bio-energy/fuel feasibility and sustainability with general research interests in renewable energies and sustainability issues. Professor Mbohwa has presented at numerous conferences and published more than 150 papers in peer-reviewed journals, 6 book chapters and one book. He holds a Masters in Operations Management and Manufacturing Systems from the University of Nottingham and completed his doctoral studies at Tokyo Metropolitan Institute of Technology in Japan. Prof Mbohwa was a Fulbright Scholar visiting the Supply Chain and Logistics Institute at the School of Industrial and Systems Engineering, Georgia Institute of Technology. He is a fellow of the Zimbabwean Institution of Engineers and is a registered mechanical engineer with the Engineering Council of Zimbabwe.

ISBN: 978-988-14047-2-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)