International Association of Scientific Innovation and Research (IASIR)
(An Association Unifying the Sciences, Engineering, and Applied Research)

ISSN (Print): 2279-0063
ISSN (Online): 2279-0071

International Journal of Software and Web Sciences (IJSWS)

(Open Access, Double Blind Peer-reviewed, Refereed and Indexed Journal)

www.iasir.net

Predictive Data Mining Model for Classifying Employee Performances (PDM-EP)

Ejiofor C. I[1] & Onuodu .F[2]
Department of Computer Science,
University of Port-Harcourt, Port-Harcourt. Nigeria[1,2]

Abstract: *This research paper has shown the usefulness of data mining in extracting useful patterns for classifying employee performance. The designed model: Predictive Data Mining Model for Classifying Employee Performances (PDM-EP) comprises of six model components: Input (IPu), Data Normalization (DNor), Rule Classification (RCliff), Rule Splitter (Rspit), Knowledgebase (Kb) and Output (Otp). The model was designed using an object oriented tool: Unified Modeling Language (UML). This tool portrayed the behavioral functionalities of the model using sequence diagram.*

Keywords: *Data mining, Employee, Performance*

I Introduction

Data mining; a process associated with intelligent systems has been employed and adapted in solving numerous Artificial Intelligent (AI) problems. This processbroadlyis a sub-domain of computer sciences. Data mining; deals with the discovering of non-visible patterns existing within large dataset (Han, et al., 2011). The process of data mining intertwined largely with machine learning, computational processes, statistical evaluation and even database system (Han, et al., 2011; Kantardzic, 2003). The ability to extract and transform hidden patterns into model base understandable structures is the fundamentals of data mining (Clifton et al., 2010). Data mining also entails data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualizationand online updating (Kantardzic, 2003). Data mining has found its application across automated prediction of trends and behaviors, which automates the predictive information found in large systems. This involves the usage of preexisting data in predicting and forecasting future occurrences. Data mining can also be implemented in automated discovery of previously unknown patterns which involves finding and identifying previously hidden patterns which could be applied in solving problems (*http://www.thearling.com/text/dmwhite/dmwhite.htm*). Data mining techniques includes Artificial Neural Network (ANN), a non-linear approach in modeling predictive model through visualized training. Decision Tree (DT), which generate decision rules have also been widely applied in data mining as a classification base. Genetic Algorithm (GA) optimization techniques which combine natural process such as combination, mutation and natural selection has found it application in data mining.Nearest Neighbors (NN) has also been applied repeatedly as a classification tools under data mining in addressing perceived hidden patterns (Haughton et al., 2003; Berson, 2005). Rule induction, largely deals with the implementation of IF-Then –Rules (Lukasz and Petr, 2006). These techniques are having been applied in most fields: financial sector, educational, agricultural, education and manufacturing. It usefulness and versatility have indeed made it a useful tool of implementation (Óscar et al., 2009).

Employee performance is integral facets of an organization which must be implemented in other to enhance employee performance as integral organizational resources. Knowing this, organization has employed various incentives in enhancing employee performance (Aliya et al. 2013; Nassazi, 2013). These incentives include promotion, bonus, increased salary and even regular leaves. Although this incentive have been implemented, it is difficult to ascertain employee performance verbally, orally, or manually (Nassazi, 2013). Therefore several research methods have been implemented in ascertaining employee performances. Therefore, it is the intent of this research paper in providing of a predictive data mining model for classifying employee performances.

II. Review of Related Researches on Data Mining

This selection provides a brief review of related works on data mining which has been explored over the years. This selection provides and identifies certain limitations opening the way for future progression. These limitations provide an avenue in implementing a novel techniques; model based on identified limitation. Tables 2.1 provide a brief review of related literatures pertaining to data mining.

Table 2.1: Review of Related works on Data mining

SN	Author (Year)/Title	Goal	Strength/ Finding	Limitation/ Weakness	Further Research
1.	Alex et al. (2005) An Overview of Data Mining Techniques	Descriptive overview of Data Mining techniques	Various techniques were discussed	No model based solution was implemented	Model base solution will be a plus focusing on employee performance
2.	Mansi and Shivani, (2015) Data Mining - Techniques, Methods and Algorithms: A Review on Tools and their Validity	Summarization of data mining tools and techniques	Comparative Analysis of varied tools	No model based solution was implemented	Model base solution will be a plus focusing on employee performance
3.	NaheedAzeem, ShaziaUsmani (2011), An analysis of Data Mining Based Software Defect Prediction Techniques	Descriptive Analysis of fault prone modules	Comparative Analysis of varied tools employed in analysis fault prone modules	No model based solution was implemented	Model base solution will be a plus focusing on employee performance
4.	Leisa et al. (2002) The application of data mining techniques to characterize agricultural soil profiles	i. Relationship between meaningful soil data ii. Comparison between various data mining techniques	Comparative Analysis of varied tools	No model based solution was implemented	Model base solution will be a plus focusing on employee performance

The research work reviewed from Table 2.1 captures some fundamental issues which will prompt future research on data mining. This literature was mostly descriptive with little attention paid to model design and development implementing data mining. Therefore base on the limitation of these literatures, it is the intent of this research paper to provide a model for classifying employee performance.

III. Predictive Data Mining Model for Classifying Employee Performances (PDM-EP)

The Predictive Data Mining Model for Classifying Employee Performances (PDM-EP) is an intelligent data mining classification model, built with the aim of classifying employees based on a predefined organization classification rules using attributes and labels. This classification determines the productivity of each employee determining organization incentives. Figure 3.1 portrays the Predictive Data Mining Model for Classifying Employee Performances with six main components which include: Input (IPu), Data Normalization (DNor), Rule Classification (RCliff), Rule Splitter (Rspit), Knowledgebase (KB) and Output (Otp).

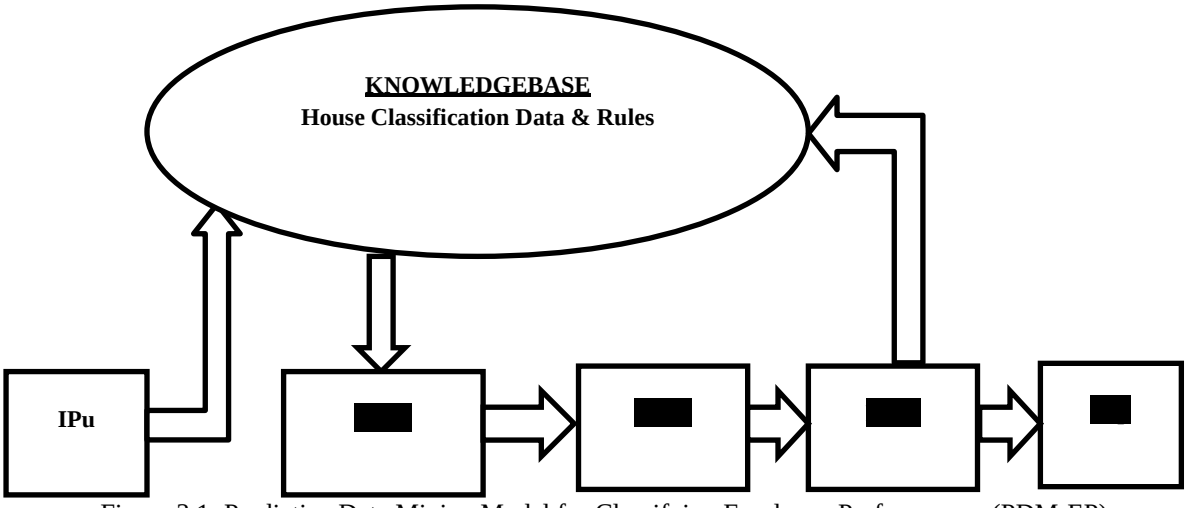

Figure 3.1: Predictive Data Mining Model for Classifying Employee Performances (PDM-EP)

The six model components are elucidated with the following:

Input (IPu): The model input classification data comprises of data derived from various organizations with various employees. This research focuses on Small and Medium Enterprises (SME). This input was collect using a survey approached streamlined toward quantitative and qualitative questionnaires. This questionnaire comprises of demographic section and the questionnaire body. The demographic information identifies personalized

information while the body was streamlined toward employee character condition. A sizeable number of questionnaires are distributed in obtaining numerous classification data.

Data Normalization (DNor): Data normalization is the same as data preprocessing or data preparation. This section is relevant due to the variation of data collected. This data must been placed in a universal scale format enabling data classification with ease. The data normalizations help to eliminate some irrelevant attributes and associated labels, hampering data optimality. Attributes like illnesses, marriage and travel where eliminated while labels complementing these attributes such as very ill, slightly ill, not ill, quickly married, delay in marriage, not married were eliminated along with their attributes. The data was arranged based on best attributes and labels such as time worked, time resumed, hand on equipment and performance rate. For ease of operation each employee was given an employee ID. The data was final categorized within the range of 0-1and performance indicator as excellent, good and poor.

Rule Classification (Rcliff): The rule classification provides the classification rules based on associated attributes tied to associated labels. The labels categorize each attributes and determine seriousness or lack of seriousness of each employee attributes. The rules classifications are inbuilt within the model and determine the model final output categories: excellent performance, good performance and poor performance. The perimeters S and the associated labels J was permutated into $\mathbf{J^S}$providing the total number of model classification rules.

Rule Splitters (Rsplt): The rule splitter simply identifies poor result formation and ultimately output rules.The poor rule formation identifiesrules which must be reclassified. Therefore these rules are sent back to the knowledgebase for reformation. The output rule identifies rules formation with associated attributes and labels conforming to the requirement set within the models. These rules ultimately will determine the rule output: “Excellent Performance, “Good Performance” and “Poor Performance. These rules are sent output as output.

Knowledge base (KB): The initial classification data received as input are housed with the knowledgebase which is subsequently transmitted to the normalization domain. The poor classifications are held also in the knowledgebase prior to rule reformation.

Output (Otp): The output simply receives the output rules as employee performance indicator, which is subsequently used within the organization for enhancing decision making and managing organizational resources.

IV. Design of Predictive Data Mining Model for Classifying Employee Performances (PDM-EP)

The model designed was captured using a standard object oriented designed tool specifically: Unified modeling Languages (UML) due to simplicity, objects acceptance and large user communities. UML sequence diagram showcase the integral model functionalities, exposing the exchange of message between model objects. Figure 4.1 portray Sequence Diagram identifying the Predictive Data Mining Model for Classifying Employee Performances (PDM-EP). The sequence diagram comprises of six main objects: input, data normalization, rule classification, rule splitter, knowledgebase and output. These objects interact in a synchronous manner, with the *input* as an object providing the first input data which is passed on to the next object: *dataNormalization*. The *dataNormalisation*; normalized the received data and interact with the *ruleClassification* by transmitting these normalized data. The *ruleClassification* provide classification rules which is subsequently analyzed by the *rulesplitter*determining *knowledgebaseData storage* and *outPut data*.

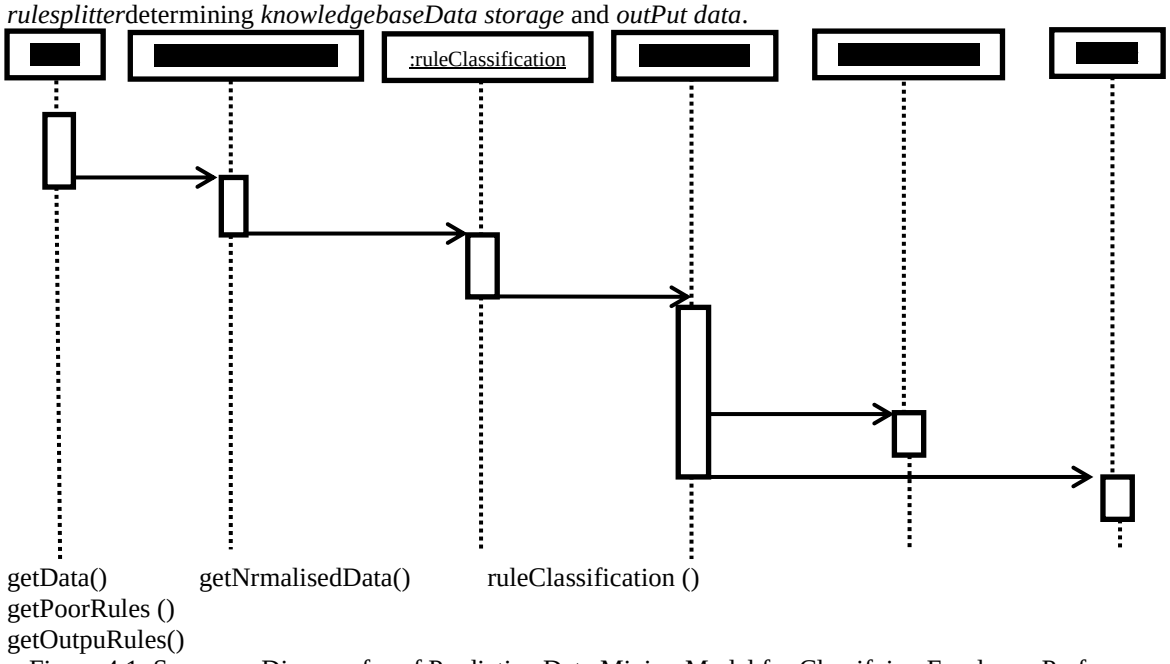

Figure 4.1: Sequence Diagram for of Predictive Data Mining Model for Classifying Employee Performances (PDM-EP)

V. Model Discussion

ThePredictive Data Mining Model for Classifying Employee Performances (PDM-EP) model will enhance efficiency within the work environment through the identification of diligent workers using it three performance indicators: “Excellent Performance, “Good Performance” and “Poor Performance”. These performance indicators determine employee remuneration, rewards and benefits. The model on full implementation will obtains the following benefits:

a. Identification of efficient employee
b. Proper employee reward package
c. Management of manpower resources
d. adequateorganization forecasting

VI. Conclusion

The development, design and elucidation of a Predictive Data Mining Model for Classifying Employee Performances (PDM-EP) have been achieved. The model comprises of six main components: Input (IPu), Data Normalization (DNor), Rule Classification (RCliff), Rule Splitter (Rspit) and Output (Otp). The model integral functionalities were exposed using object oriented design methodologies specifically Unified Modeling Language (UML), portraying the behavior of the model. The overall organization policy pertaining to employee will be enhanced with this model.

References

Alex B., Stephen S., and Kurt T. (2005),An Overview of Data Mining Techniques,weber.itn.liu.se/~jimjo94/courses/TNM048/documents/DM-Techniques.pdf

Berson A. (2005), An Overview of Data Mining Techniques,weber.itn.liu.se/~jimjo94/courses/TNM048/documents/DM-Techniques.pdf

Clifton, C. (2010). "Encyclopædia Britannica: Definition of Data Mining". Retrieved 2010-12-09.

Han, K., Pei, J., Micheline, J. (2011). Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann. ISBN 978-0-12-381479-1.

Haughton, D.; Deichmann, J.; Eshghi, A.; Sayek, S.; Teebagy, N.; and Topi, H. (2003); A Review of Software Packages for Data Mining, The American Statistician, Vol. 57, No. 4, pp. 290–309

Leisa J. Armstrong, D. Diepeveen, R. M. (2002), The application of data mining techniques to characterize agricultural soil profiles, retrieved online from *citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.300.2147&rep=rep1...pdf*

Lukasz K. and Petr M. (2006); A survey of Knowledge Discovery and Data Mining process models. The Knowledge Engineering Review. Volume 21 Issue 1, March 2006, pp 1–24, Cambridge University Press, New York, NY, USA

Kantardzic, M. (2003). Data Mining: Concepts, Models, Methods, and Algorithms. John Wiley & Sons. ISBN 0-471-22852-4. OCLC 50055336.

Mansi G. and ShivaniGoel (2015), Data Mining - Techniques, Methods and Algorithms: A Review on Tools and their Validity, *International Journal of Computer Applications (0975 – 8887) Volume 113 – No. 18*, Pp. 22 -29

Naheed A. and Shazia U (2011), Analysis of Data Mining Based Software Defect Prediction Techniqueshttps://globaljournals.org/GJCST_Volume11/1-Analysis-of-Data-Mining-Based-Software.pdf

Nassazi A. (2013), Effects Of Training on Employee Performance, retrieved online from https://www.theseus.fi/bitstream/handle/10024/67401/THESIS.pdf?sequence=1

Óscar M., Gonzalo M. and Javier S. (2009); A Data Mining & Knowledge Discovery Process Model. In Data Mining and Knowledge Discovery in Real Life Applications, Book edited by: Julio Ponce and AdemKarahoca, ISBN 978-3-902613-53-0, pp. 438–453, February 2009, I-Tech, Vienna, Austria.

Aliya I. Maiya I., Farah L., Hina M. (2013) Factors Affecting the Employee’s Performance: A Case Study of Banking Sector in Pakistan, www.ejbss.com/Data/.../ejbss-1656-15-factorsaffectingtheemployeesperformance.pdf