Employee assessment using data mining techniques: modeling individual capability to improve the competency of companies

Prajak Chertchom

November 29, 2018

Employee assessment using data mining techniques: modeling individual capability to improve the competency of companies

Prajak Chertchom

Faculty of Information Technology, Thai-Nichi Institute of Technology

1771/1 Pattanakarn Rd., Suanluang, Bangkok 10250, Thailand

prajak@tni.ac.th

Abstract—**This paper presents an applying of data mining technique in analyzing the assessment result of staff from big company in Thailand. The study used data mining algorithms with simulator tool in Rapid Miner to model discovered information into three aspects; firstly, the study did a comparison of three algorithms in terms of accuracy and speed. Secondly, the paper presented a modelling of discovered information from three algorithms. Finally, the paper suggested the appropriated model and data visualized tool for HR in monitoring and adjusting for human resource planning and developing.**

Keywords—*data mining, individual capacity, Rapid Miner, Decisions Tree, Naïve Bayes, Generalized Linear Model, and Deep Learning*

I. INTRODUCTION

Today data analysis tools is one of emerging science that is applied to extract information, and patterns, hidden in data. Data science is a combination of techniques, mining, machine learning, predictive and data visualization. If company creates a monitoring systems that can measure individual capacity, individual skill etc. Then, a monitoring systems will provide a guideline of an individual capacity in each the state of development process within the organization. (Mapping with organization's key core competency)

Data sets are knowledge assessment results from big company in Thailand with a data mining technique. The study applied a simulator tool from Rapid Miner to establish, derive and come up with a model of individual capacity and framework related to individual capacity. The knowledge from data mining and visualization of simulator tool can guide a HR & Management for the purpose of establishing a dimensional indicator to identify the level of individual capabilities, to better understand and improve an individual capacity and its organization competency.

Individual capability in organization development process at organizational level is a critical factor in sustaining a company's competitiveness.[1] However, competency in knowledge is more complicated than merely getting the job done properly and with the fact that no significant evidence in the measurement of individual's capacity and no existing measuring tool at individual level.[2] In his book, Kollock[3] an American well-known business management practitioner, 'What Gets Measured Gets Done' and "Thriving on chaos". He wrote that

measurement of an individual is a very important methodology in business process. Most organizations set up vision, mission and target every year with the expectation of quality, efficiency and productivity. By the way, all processes need a controlling and monitoring for identifying potential gaps in key information. Good predictive simulator can help and focus on individual performance requirements and results. However, in big organization with staffs exceeding 200 people (4,705 firms). [4] It is difficult for the HR department and the management to measure the individual capacity and to understand what kind of competency their staffs should possess in order to accomplish a company task.

This paper is organized as follows. Section 2 describes a short brief of performance appraisal' concepts, a data mining techniques used in this study such as Naïve Bayes, Generalized Linear Model, Deep Learning and Decisions Tree and its application. Section 3, specifically, presents comparison results from several data mining techniques and applying of a simulator tool. Section 4 describes a discussion and consideration. Section 5 explains the conclusion and future work.

II. LITERATURE REVIEW

A. Performance Appraisal' concepts

1) A performance assessment is a procedure that used to evaluate the job performance and individual capacity such as individual's quality, work quantity, knowledge, cost of project, and time management typically by the manager. Normally organization uses a performance assessment for a tool to evaluate the performance of an employee from the past job, to guide and to manage skill and career development.

Meesutha and Sutchukorn[5] explained in their book that the methods generally used in the individual performance assessment process include the following:

a) Management by objectives (MBO)

b) 360-degree appraisal

c) Behavioral observation scale

d) Behaviorally anchored rating scales

2) Knowledge Assessment: Performance appraisal is general system or tool or method for measuring an employee's

job performance on recent successes and failures. Generally, the traditional system measures an individual based on the areas for example: skill, knowledge, teamwork, character, and leadership effectiveness. Knowledge Assessment helps orgnanization to build their capacity to access and use knowledge to strengthen their competitiveness. The activity supports HR to design and develop a realistic and achievable knowledge-based competency strategies. It helps company assess how employee compare with others in their ability to compete in the knowledge economy and to identify appropriate traning policies to help them achieve their businsess goals. Assessment results also offers recommendations for HR development to reform on the knowledge gap accompanied by complementary advice on what the orgnaization needs to do to develop appropriate capacity to deliver on business target.

B. Data Mining and its Tool

Data Mining is the techniques of extracting enormous data to discover useful information and patterns. There are several data mining techniques. In this study, we considered more frequently used by industrial people[6]; Decisions Tree, Naïve Bayes, Generalized Linear Model, and Deep Learning.

- Decision Tree: This algorithm is a technique that constructs classifiers to predict the value of a target variable based on several input variables. Classification technique in D3 is a technique for grouping information with defined attributes. This technique is suited for predictive modeling to unseen the future. It is also called "supervised learning". There are two types of classification. D3 is normally applying "Entropy" to measures the homogeneity (purity) of a set of examples. The entropy for S is defined as

$$\text{Entropy}(S) = - (p/t).\log2(p/t) - (n/t).\log2(n/t)$$

- Naïve Bayes [7] is a conditional probability model. It is one of the most commonly used equations in machine leaning. This algorithm is generally applied for segmentation of customers, Sentimental Analysis and Identify object etc. The following figure shows Naïve Bayes formula.

$$P(c \mid d) = \frac{P(d \mid c)P(c)}{P(d)}$$

Fig. 1.   Naïve Bayes formula

- Generalized Linear Model [8] is an algorithm for predictive analytics that used to compare past successes and failures, and then used those formulas to predict future outcomes.

$\eta_i = \beta_0 + \beta_{1x1i} + ... + \beta_{pxpi}$ and two functions

-a link function that describes how the mean, $E(Y_i) = \mu_i$, depends on the linear predictor

$$g(\mu i) = \eta_i$$

-a variance function that describes how the variance, $\text{var}(Y_i)$ depends on the mean

$$\text{var}(Y_i) = \varphi V (\mu)$$

where the dispersion parameter φ is a constant

Fig. 2.   Generalized Linear Model (GLMs)

- Deep Learning (Neural Network) [9-10]: This model are generally based on an artificial neural network(ANN), It is an algorithm that attempts to create a model to represent a high level of data by creating an information architecture that consists of several substructures. Each of these is derived from nonlinear transformations.
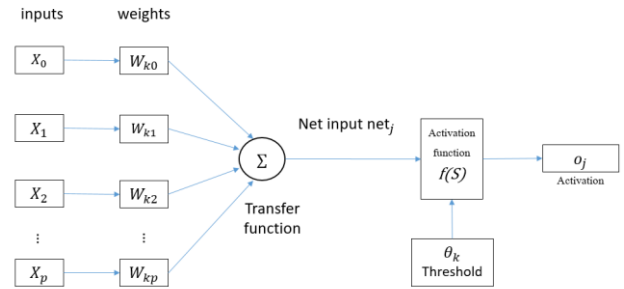

Fig. 3.   Deep Learning

Thus, from the theory background and problems statements, this study applied the three DM techniques and tool as described above for comparing for analyzing data set to predict and finding pattern of Individual Capacity.

In addition, simulator in application of Rapid Miner was applied to show how it good for HR staff in their task to manage an individual competency in an organization.

III.   RESEACH METHODOLOGY

A. Data Sets

The data obtained from knowledge competency test result from big company in Thailand. The test contains of 4 dimensions with 80 items (Planning, Organizing, Leading and Monitoring) and final score of 217 employees. The datasets consist of Employee code, Position, Planning test score, Organization test score, Leadership test score and monitoring test score. Table 1 shows some information of the datasets.

TABLE I.   SAMPLE INFORMATION FROM DATASETS

Emp. Code	Position	Plan	Organization	Leader	Monitor
48419 45	Group Leader	85	85	85	75
69097 80	Part Leader	60	85	70	85
48423 91	Group Leader	65	80	60	85
78918 63	Group Leader	95	80	60	90
38054 89	General	65	70	60	80

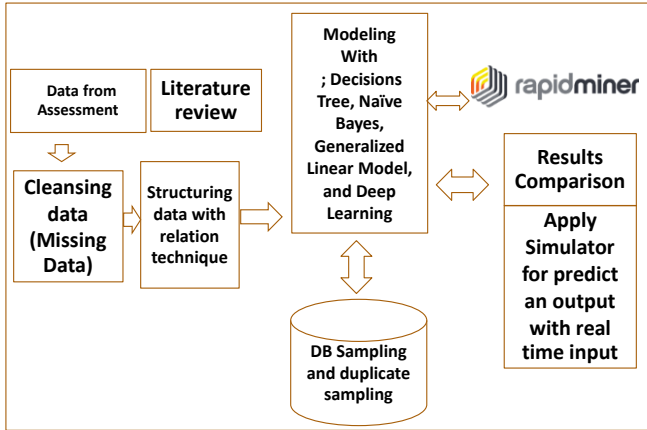Then, we organized data using MS SQL and MS Excel to suit this study's specific purposes as shown in figure 4.

Fig. 4.   Research Methodology Framework

This study used Rapid Miner Application for modeling and computation. Comparative techniques were applied in the experiments that are, Decision Tree, Naive Bayes, Generalized Linear Model, and Deep learning. Moreover, we use simulator of this application to change the inputs to a model and view the output for shows predictions and find the best value to advise a HR for knowledge score to fit with an organization competency.

IV. RESULT AND DISCUSSION

A. Accuracy

For the first result, the study shows an overview of an accuracy as shown in figure 5.
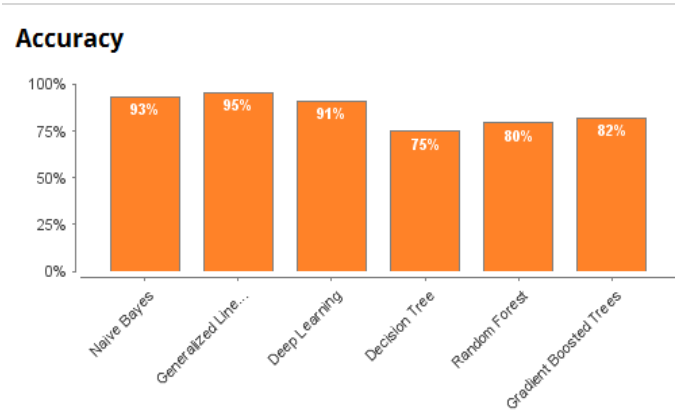

Fig. 5.   Accuracy of each Algorithms

From our experimental, the assessment data is suitable for using "Generalized Linear Model" in prediction the model of individual capacity with 95% follows by Naïve Bayes (93%), Deep Learning (91%) and Decisions Tree (75%) respectively.

B. Runtime (ms)

From figure 6, it shows that Naive Bayes took the shortest time in computation with 41 milliseconds (ms) followed by Decision Tree (69 ms), Deep Learning (217 ms) and Generalized Linear Model(317 ms) respectively.
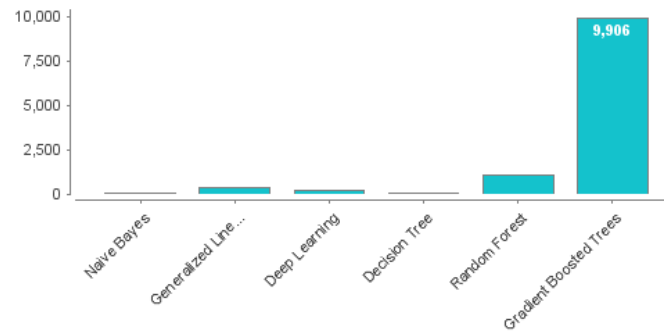

Fig. 6.   Runtime (ms)

It reveals that the most accuracy model took a longest time from selected algorithms (excluded Random forest and Gradient Boosted Trees which were applied for D3 differentiation). Thus, from our conclusion, we can assume that with a big data computation if we apply Generalized Linear Model, we need a time and enough CPU for processing.

C. Modeling with Tree

From figure 7, we can see that this tree model imply for HR that monitoring, planning, leading and organizing are the attribute predicted class or value of target variables by learning decision rules inferred from prior data respectively.
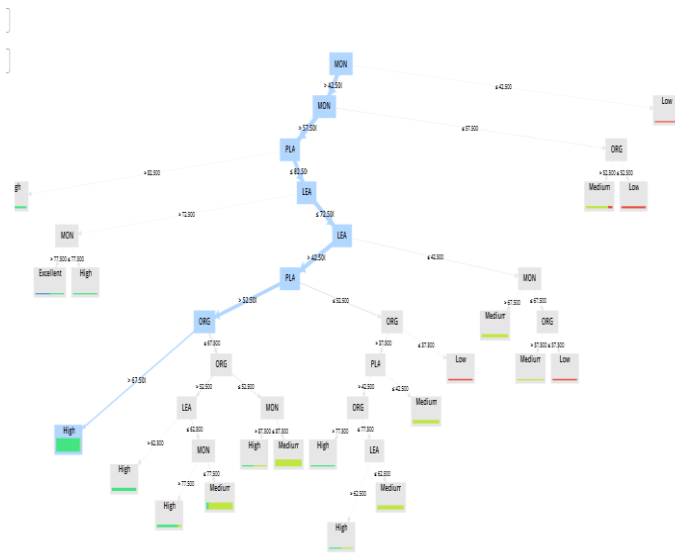
Fig. 7. Model with Decission Tree

With this decision tree result, company must focus on a knowledge of monitoring skill and planning which mostly impact to an organization competency

D. Figures Generalized Linear Model

TABLE II.   GENERALIZED LINEAR MODEL

Attribute	Coefficients		
	Excellent	*High*	*Medium*
PLA	0.432	0.251	-0.116
ORG	0.398	0.269	-0.115
MON	0.269	0.360	-0.040
LEA	0.472	0.293	-0.117
Intercept	-110.013	-78.298	22.242

From table 1, the Generalized Linear Model reveals that the excellent capacity of staff depends on leading (0.47), planning (0.43), organizing (0.39) and monitoring (0.27)   respectively.

Thus, for develop individule capacity, company must foucs on knowledge of leading first follows by planing. Hence, this score model help HR in evaluating staff for knowledge gap training.

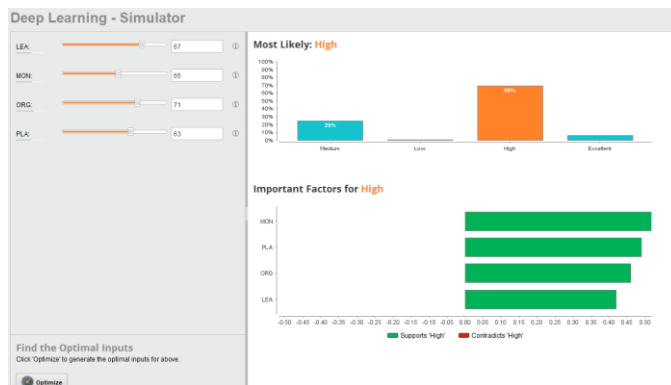## E. Simulator from Rapid Miner

Fig. 8.   Simulator from Rapid Miner

Simulator from Rapid Miner provides an easy-to-use, real-time interface to change the inputs to a model and view the output. It shows predictions, confidences, and explanations for those inputs. Thus, we recommend a company to use deep learning- Simulator for easy adjust an input from an assessment result. In case of big data from organization, it can help an organization in planning easily from this tool for competency planning.

V. CONCLUSION AND FUTURE WORK

After modelling with three algorithm by Rapid Miner, simulator function has been apply for an interactive benchmarking tool that allows organizations to identify the problems and opportunities that they face in making the transition to the key competency that they set for building competitiveness, and where they may need to focus on HR development attention to encourage future investments for employees. The Knowledge Assessment Methodology has been applied via various industry from education to production. Assessment can be used by anyone with an internet connection so it can store vast amount of data that can be used to analyze and prediction. The information technology system can generate reports that reveal how skill of employee compares with other organization on various aspects of the knowledge economy.

Graphically by simulator offers several options and pre-set display modes for complex visual representations, such as basic scorecards that reveal similarities, gap, strengths, and weaknesses across organization. Thus, for future work, researcher would like to collect a knowledge and skill assessment result from industry, then, apply data analytic with application development to create a system function that help industry in each sector to visualize their knowledge and skill competency. In addition, with this system can provide real-time interface to change the inputs of knowledge and skill need to a model and view the potential competitiveness of their own.

REFERENCES

[1] H. Vinarski-Peretz, G. Binyamin, and A. Carmeli, "Subjective relational experiences and employee innovative behaviors in the workplace." Journal of Vocational Behavior, vol. 78, no. 2, 2011, pp. 290-304.

[2] C.H. Caroline, G.P. Hammond, C.I Jones, and R.C. McKenna. "Enabling technologies for industrial energy demand management." Energy Policy vol. 36, no. 12, 2008, pp. 4434-4443.

[3] P. Kollock, "The economies ol online cooperation." Communities in cyberspace vol. 220, 1999

[4] "Number of establishments by size of establishments" The National Statistical Office (NSO) . [Online]. Available: http://web.nso.go.th/en/census/indus/data/Census_Executive_12.pdf. [Accessed: 10-Jul-2018]

[5] A. Meesutha and S. Sutchukorn, "Performance Appraisal by competency assessment", TPA Publishing, 13th edition, 2008

[6] "7 Important Data Mining Techniques for Best results" [Online]. Available: https://www.educba.com/7-data-mining-techniques-for-best-results/ [Accessed: 15-Aug-2018]

[7] "Naïve Bayes" [Online]. Available: http://www.cs.uvm.edu/~icdm/algorithms/10Algorithms-08.pdf . [Accessed: 15-Aug-2018]

[8] "Generalized Linear Fomula," [Online]. Available: https://slideplayer.com/slide/9449268/. [Accessed: 15-Aug-2018]

[9] "Deep learning" [Online]. Available: https://www.analyticsvidhya.com/blog/2017/05/gpus-necessary-for-deep-learning/. [Accessed: 18-Aug-2018]

[10] L. Deng and D. Yu, "Deep Learning: Methods and Applications". Foundations and Trends in Signal Processing, vol. 7 (3–4), pp. 1–199, 2014, doi:10.1561/2000000039.