International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 02 | Feb 2020 www.irjet.net p-ISSN: 2395-0072

Employee Performance Prediction System using Data Mining

Tejas Raut[1], Priya Kale[2], Rashmi Sonkusare[3], A. K. Gaikwad[4]

[1,2,3]Student, Dept. of Computer Science & Engineering, DES'S COET, Dhamangaon Rly
[4]Professor, Dept. of Computer Science & Engineering, DES'S COET, Dhamangaon Rly

---

Abstract - Human capital is of a high concern for companies' management where their most interest is in hiring the highly qualified personnel which are expected to perform highly as well. Human Resources Management (HRM) has become one of the essential interests of managers and decision makers in almost all types of businesses to adopt plans for correctly discovering highly qualified employees. Accordingly, managements become interested about the performance of these employees. Results show that professional skill development programs are needed in order to prepare employees to perform their tasks more efficiently. The knowledge flow model of the Open source tool is also used to frame the elements. To get a highly accurate model, several experiments were executed based on the previous techniques that are implemented tool for enabling decision makers and human resources professionals to predict and enhance the performance of their employees.

***Key Words*: HRM, Employee Performance Prediction, Data Mining, KNN, Dataset.**

1. INTRODUCTION

HRM has a leading role in deciding the competitiveness and effectiveness for better continuation. Organizations consider HRM as "people practices". Therefore, it becomes the responsibility of the HRM to allocate the best employees to the appropriate job at the right time, train and qualify them, and build evaluation systems to monitor their performance and an attempt to preserve the potential talents of employees. A young yet promising of this kind is data mining. It standouts due to its wide-array of techniques from the different domains such as statistics, artificial intelligence, machine learning, algorithms, database systems and visualization. These influences served as groundwork for its applications to business for which the academic institution is unexceptionally classified. Generally, regardless of discipline, data mining has gained popularity due to its tools with potentials to identify trends within data and turn them out into knowledge mostly with predictive attributes that could significantly lead to better and strong bases for decision making with a wide range of Open source tools availability.

Data mining is a young and promising field of information and knowledge discovery. It started to be an interest target for information industry, because of the existence of huge data containing large amounts of hidden knowledge. With data mining techniques, such knowledge can be extracted and accessed transforming the databases tasks from storing and retrieval to learning and extracting knowledge. Decision tree is one of the most used techniques, since it creates the decision tree from the data given using simple equations depending mainly on calculation of the gain ratio, which gives automatically some sort of weights to attributes used, and the researcher can implicitly recognize the most effective attributes on the predicted target. As a result of this technique, a decision tree would be built with classification rules generated from it.

Previous studies specified several attributes affecting the employee performance. Some of these attributes are personal characteristics, others are educational and finally professional attributes were also considered. Chein and Chen (2006) used several attributes to predict the employee performance. They specified age, gender, marital status, experience, education, major subjects and school tires as potential factors that might affect the performance. Then they excluded age, gender and marital status, so that no discrimination would exist in the process of personal selection. As a result for their study, they found that employee performance is highly affected by education degree, the school tire, and the job experience. Most recently, the prevalence of intelligent machine learning algorithms in the field of computer science has led to the development of robust quantitative methods to derive insights from industry data.

Supervised machine learning methods—wherein computers learn from analyses of large-scale, historical, labelled datasets—have been shown to garner insights in various fields, like biology and medical sciences, transportation, political science as well as many other fields. Owing to the advancements in information technology, researchers have also studied numerous machine learning approaches to improve the outcomes of human resource (HR) management. A detailed listing of recent studies in using supervised machine learning on employee turnover is described in Table 1, and lists the data included and related machine learning algorithms that were used therein, including decision tree (DT) methods, random forest (RF) methods, gradient boosting trees (GBT) methods, extreme gradient boosting (XGB), logistic regression (LR), support vector machines (SVM), neural networks (NN), linear discriminant analysis (LDA), Naïve Bayes (NB) methods, K - nearest neighbor (KNN), Bayesian networks (BN) and induction rule methods (IND)

2. RELATED WORK

In general, Classification contains some steps to complete its process. The first step is called the learning step where in the model; predefined classes are built by analyzing a set of training dataset variables. Each variable is assumed that has a relation and regards to a predefined class. The second step is responsible for estimating the accuracy of model or classifier (validating the model) through testing the model using a different dataset. If the classifier's accuracy was considered acceptable, the model or classifier can be used to apply to new unseen data to give prediction about specific unknown label class and this is considered the third step. Therefore, the model acts as a classifier in the process of decision-making. There are various classification techniques have been used in the prediction process such as DT, Naïve Bayes, SVM, etc.

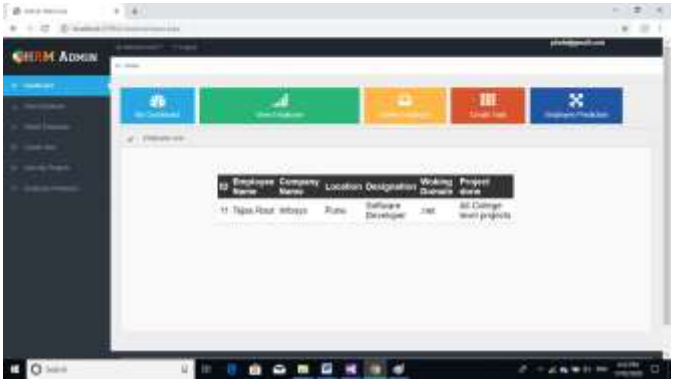
Human capital is of a high concern for companies' management where their most interest is in hiring the highly qualified personnel which are expected to perform highly as well. Foremost, the Human Resource (HR). Management plays the role of ensuring this by closely adhering to the standards set by the higher management or by some heuristic needs of applicants with distinctive qualifications and potentials. However, oft-quoted factors that may affect employee performance are attributed to educational backgrounds, working experiences, as well as personal qualities. These when converged provide a picture of how well an employee performs his tasks. Assessment of human resource performance is a sensitive task. To avoid partiality, an efficient tool to deal with various data and assist managers to make decisions and plans is of great help. In data mining, historical data such as those attributes that influence performance could be exploited as learning experiences. These can be used to predict future circumstances and rich resource of knowledge and decision supports.

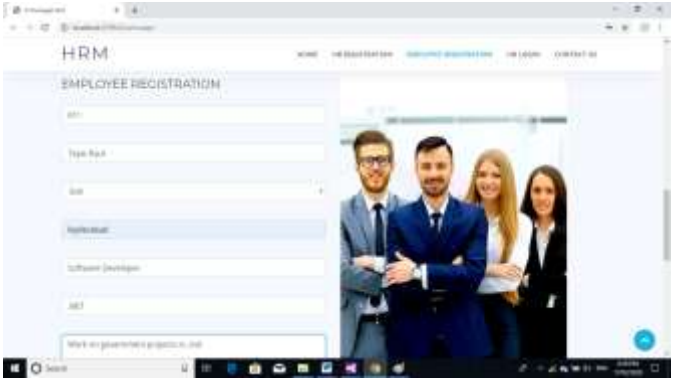
In general, data classification is a two-step process. In the first step, which is called the learning step, a model that describes a predetermined set of classes or concepts is built by analyzing a set of training database instances. Each instance is assumed to belong to a predefined class. In the second step, the model is tested using a different data set that is used to estimate the classification accuracy of the model. If the accuracy of the model is considered acceptable, the model can be used to classify future data instances for which the class label is not known. At the end, the model acts as a classifier in the decision making process. There are several techniques that can be used for classification such as decision tree, Bayesian methods, rule based algorithms, and Neural Networks.

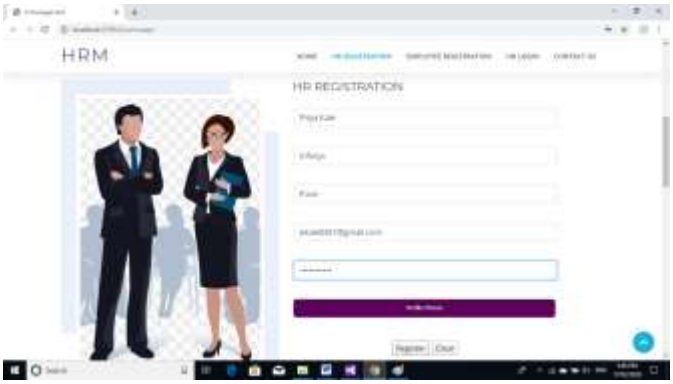
3. IMPLEMENTATION

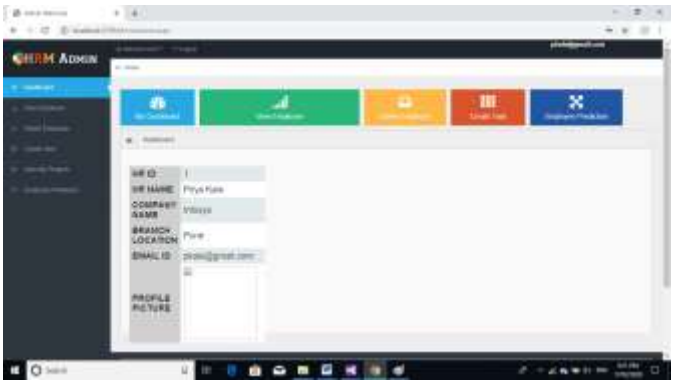
Feature selection is a one of the main concepts of DM and Machine Learning. Where, it is a process of selecting necessary useful variables in a dataset to improve the results of machine learning and make it more accurate. At which, Using too many numbers of variables in a dataset reduce predictive performance. The data set may contain too many features; some of them do not promote the prediction accuracy, and thus make the predictive model excessively complicated.

Therefore, unnecessary useless variables must be avoided to make the model efficiently works. Deciding which unnecessary variable to avoid can be done by a manual manner using domain knowledge or it can be done automatically. In the section of Business understanding with proper endorsement and approval of some academic administrators, questions as to how the (DM) data mining functionalities are best applied in any Technological Institute has been identified. Recent studies had However, oft-quoted factors that may affect employees performance are attributed to educational backgrounds, working experiences, as well as personal qualities. These when converged provide a picture of how well an employee performs his tasks. Assessment of human resource performance is a sensitive task. To avoid partiality, an efficient tool to deal with various data and assist managers to make decisions and plans is of great help. In data mining, historical data such as those attributes that influence performance could be exploited as learning experiences.









4. CONCLUSION

This paper has concentrated on the possibility of building a classification model for predicting the employees' performance. Applying the DM techniques in the different problem domains in the HRM field is considered as an important and urgent issue. Especially, at the public sector in Egypt. In addition, increasing the horizons of academic and practice research on DM in HR for reaching a government sector with a high performance. This study further outlooks for applications of results to analyze enhancement programs for senior employees of any organization and to identify patterns affecting both teacher and student performance using other data mining techniques such as association rules.

REFERENCES

[1] Islam, M. J., Wu, Q. M. J., Ahmadi, M., and Sid-Ahmed, M. A., (2010), "Investigating the Performance of Naive- Bayes Classifiers and K- Nearest Neighbor Classifiers" Journal of Convergence Information Technology Volume 5, Number 2, April 2010.

[2] Al-Radaideh, Q.A., Al-Nagi, E., (2012). "Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance", International Journal of Advanced Computer Science and Applications, 3(2), pp 144 – 151.

[3] S.Yasodha and P. S.Prakash, (2012), "Data Mining Classification Technique for Talent Management using SVM," the International Conference on Computing, Electronics and Electrical Technologies, 2012.

[4] Hua Hu, Jing Ye, and Chunlai Chai, (2009), "A Talent Classification Method Based on SVM", in International Symposium on Intelligent Ubiquitous Computing and Education, Chengdu, China, 2009, pp. 160-163.

[5] Kirimi JM, Motur CA (2016), "Application of Data Mining Classification in Employee Performance Prediction". International Journal of Computer Applications, Volume 146 – No.7, July 2016.

[6] Desouki M. S., Al-Daher J (2015), "Using Data Mining Tools to Improve the Performance Appraisal Procedure, HIAST Case". International Journal of Advanced Information in Arts, Science & Management Vol.2, No.1, February 2015.

[7] ZHAO Xin (2008). A Study of Performance Evaluation of HRM: Based on Data mining. International Seminar on Future Information Technology and Management Engineering

[8] Yan Huang (2009). Study of College Human Resources Data Mining Based on the SOM Algorithm. 2009 Asia Pacific Conference on Information Processing.

[9] Chen Xiaofan, and Wang Fengbin (2010).Application of Data Mining on Enterprise Human Resource Performance Management. 3rdInternational Conference on Information Management, Innovation Management and Industrial Engineering

[10] Honglei Zhang (2009). Fuzzy Evaluation on the Performance of Human Resources Management of Commercial Banks Based on Improved Algorithm. 2009 2nd International Conference on Power Electronics and Intelligent Transportation System.