Applying Data Mining Hybrid Classification Techniques and Decision Tree approach for Government Employee's Performance Prediction

Nirvesh Kumar[1,*], Dr. Raghav Mehra[2,*]

[1]Research Scholar, Computer Science & Engineering Department, Bhagwant University Ajmer (Rajasthan)
[2]Associate Prof., Computer Science & Engineering Department, Bhagwant University Ajmer (Rajasthan)

Abstract

The precious awareness can be revealed through data mining process. In data mining, Taxonomy is one of the key tasks to communicate acquaintance from massive amount of data. Consequently right employee at right position is essential, and is the crate for executing a system using amalgam course of action based on Data Clustering and assessment Tree of Data Mining methods to assist the authority in decision making process for selecting the employee for posting and transfers. This anticipated research work projects how data clustering method can be applied for evaluating the employee recital. Different performance evaluation factors like personality, social sensitivity, efficiency, experience, punctuality, domain assignment of previous postings, ratings of previous work assignment etc. The consequence of the proposed research work predict the number of gov.employees selected for different types of work-assignment and discharged accordance to their performance

Keywords- Data Mining, Hybrid Classification Techniques,Government Employee's Performance, Prediction.

Introduction: Nowadays, in the K-Era, acquaintance is a valuable asset and among the crucial issues to address. There are a lot of fields adapted by government sector department ,customer relationship, education and some others. Therefore, the intention of this paper is to advocate the possible arrangement techniques for employee future performance through some experiments using the selected classification algorithms.

The second section describes the associated work on DM in HRM; arrangement in Data mining; and the possible methods for classification. The 3 segment confer the experiment setup in this learning. Section 4 shows some experiment results and analysis. Finally, the paper ends at Section 5 with the ultimate remarks and future investigate directions.

Related Work:

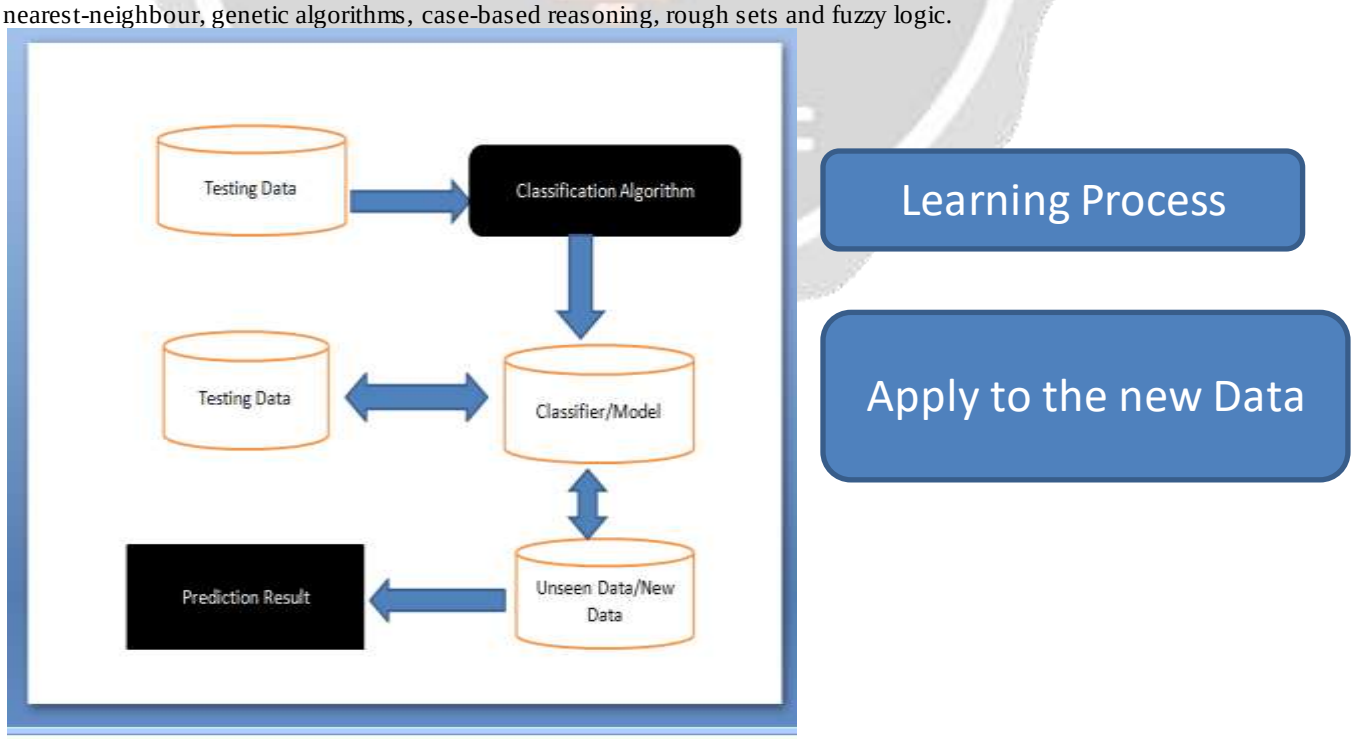
Data Mining in Human Resource: Data mining chores are generally categorized as clustering, association, sorting and (SVM) support vector machine, DM technique has been functional in a lot of fields, but its purpose in Human Resource Management (HRM) is very rare(Chien & Chen, 2008). Table I lists several of the applications in HR that use data mining tasks, and it shows that there are the minority discussions on that. Besides that, there are very few discussions or research on the uses of data mining, project assignment and talent recruitment.

Data Mining Task	Activity in HRM
Classification	Personnel selection (Chien & Chen, 2008),
	Job attitudes (Tung, Huang, Chen, & Shih, 2005)
	Personnel Selection – Recruit and Retain Talents (Chien & Chen, 2007)
Association	Training (Chen, Chen, Wu, & Lee, 2007)
Classification and Prediction	Project Assignment (Huang, Tsou, & Lee, 2006)
Classification and Association	Personnel Selection (Tai & Hsu, 2005)

Table I: Data Mining Task in HR Applications

Classification and forecast are several of the methods that can fabricate able decision. If the accuracy is considered acceptable, the model can be functional to the new data to be acquainted with the prediction result (Figure 1).

There are many techniques that can be used for classification such as decision tree, association rule mining, k-nearest-neighbour, genetic algorithms, case-based reasoning, rough sets and fuzzy logic.



Mining Process for Figure 1: Fundamental Data

Table II summarizes the potential techniques of assessment tree, neural network and nearest neighbour algorithm.

Data Mining Techniques	Classification Algorithm
Decision Tree	• *C4.5* (Decision tree induction – the target is nominal and the inputs may be nominal or interval. Sometimes the size of the induced trees is significantly reduced when a different pruning strategy is adopted). • *Random forest* (Choose a test based on a given number of random features at each node, performing no pruning. Random forest constructs random forest by bagging ensembles of random trees).
Neural Network	• *Multi Layer Perceptron* (An accurate predictor for underlying classification problem. Given a fixed network structure, we must determine appropriate weights for the connections in the network). • *Radial Basic Function Network* (Another popular type of feed forward network, which has two layers, not counting the input layer, and differs from a multilayer perceptron in the way that the hidden units perform computations).
Nearest Neighbor	• *K*Star* (An instance-based learning using distance metric to measure the similarity of instances and generalized distance function based on transformation.

Table II: Potential Data Mining Classification Techniques

Experiment Setup:

The decision tree is a 'divide-and–conquer' approach from a set of independent instances and the nearest neighbor is based on the distance metric (Table II). The process of classification uses input variables i.e. performance factors; and the outcome is the employee's performance pattern that shows the status of promotion. The attributes for dataset in this experiment as shown in Table III.

Attribute	Description
Category	P – Professional, S - Support Staff
Gender	Male and Female
Qualification	Doctorate, Master, Bachelor, Diploma and Certificate
$PK_{1...6}$	Work Outcome (50%)
$PM_{1...6}$	Knowledge and Skill (25%)
$KP_{1...6}$	Individual Quality (20%)
$KS_{1...6}$	Activities and Contribution (5%)
$YEAR_{1...6}$	Evaluation mark (100%)
Target/Class	Recommendation for promotion (Yes or No)

Table III: Attributes for Performance Prediction

The aspects for training dataset are selected based on the interrelated factors for employee concert (Executive Brief, 2008) as illustrated in Figure 2. The classification rules will demonstrate us about the interesting or imperative characteristics for the dataset. Besides that, the forecasting model will be used to conclude whether the employee is recommended for promotion or not based on their performance. In this experiment, the training dataset contains 33 related attributes from background information and concert factors which is demonstrated in Table III.

Result and Discussion:

In this experiment, the accuracy of classification techniques is measured by averaging the accuracy from 10 fold cross validation datasets. The average accuracy obtained from each of the algorithms for the dataset is shown in Table IV. The result of this experiment showed that all of the classifiers have moderate accuracy, which is more than 70%. In many cases, the moderate accuracy is considered as an acceptable accuracy. In this experiment, the dataset produced acceptable models for each of selected classification algorithms. In order to choose the suitable classifier in data mining, the precision of the model is used to settle on the most appropriate classifier for the dataset. As revealed in Table IV.

Table IV: The Accuracy of Classifier

Classifier Algorithm	Accuracy(%)
C4.5 /J4.8	79.49
Random forest	72.21
Multi Layer Perceptron(MLP)	70.25
Radial Basis Function Network	70.53
K-Star	79.34

This technique can produce rules in tree structures and rule sets ; and construct a tree for the purpose of improving the prediction accuracy (Becerra-Fernandez, Zanakis, & Walczak, 2002; Delen, Walker, & Kadam, 2005). Besides that, the C4.5/C5.0/J48 classifier is among the popular and powerful decision tree classifier (Becerra-Fernandez et al., 2002; Delen et al., 2005; Kumar & Ravi, 2007; Tso & Yau, 2007). By using C4.5/J4.8 classifier, a part of generated classifier or rules is shown in Figure 3.

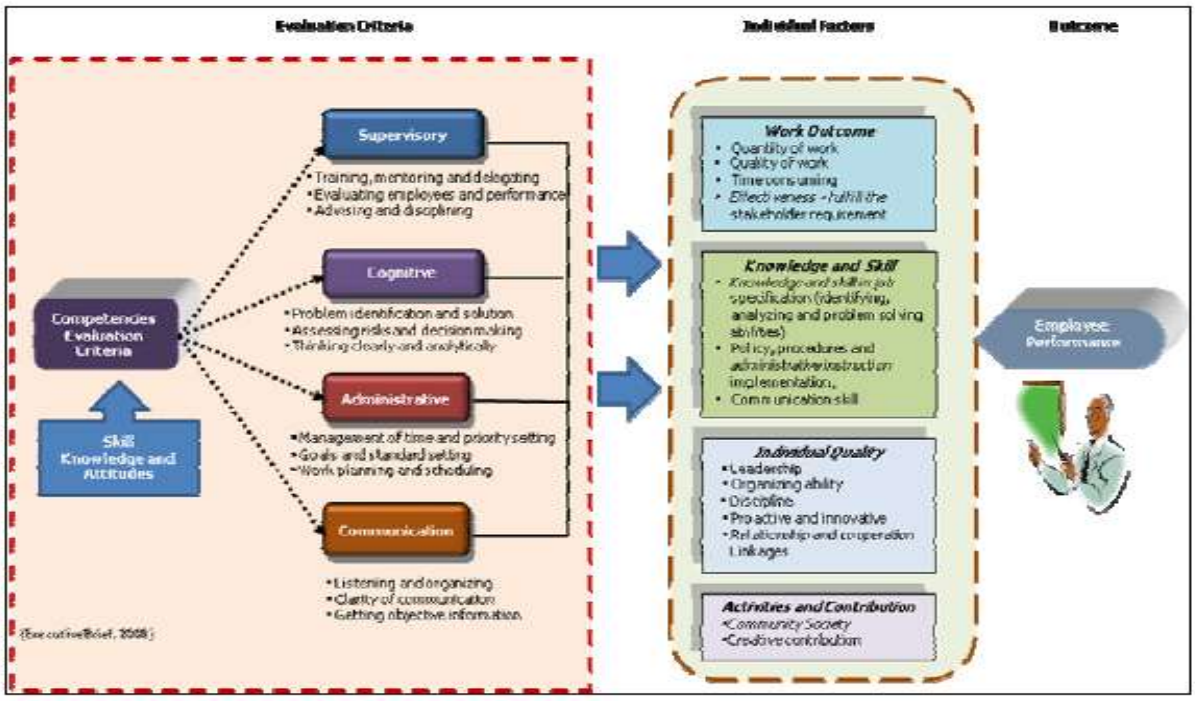


Figure 2: Related Factors for Employee Performance

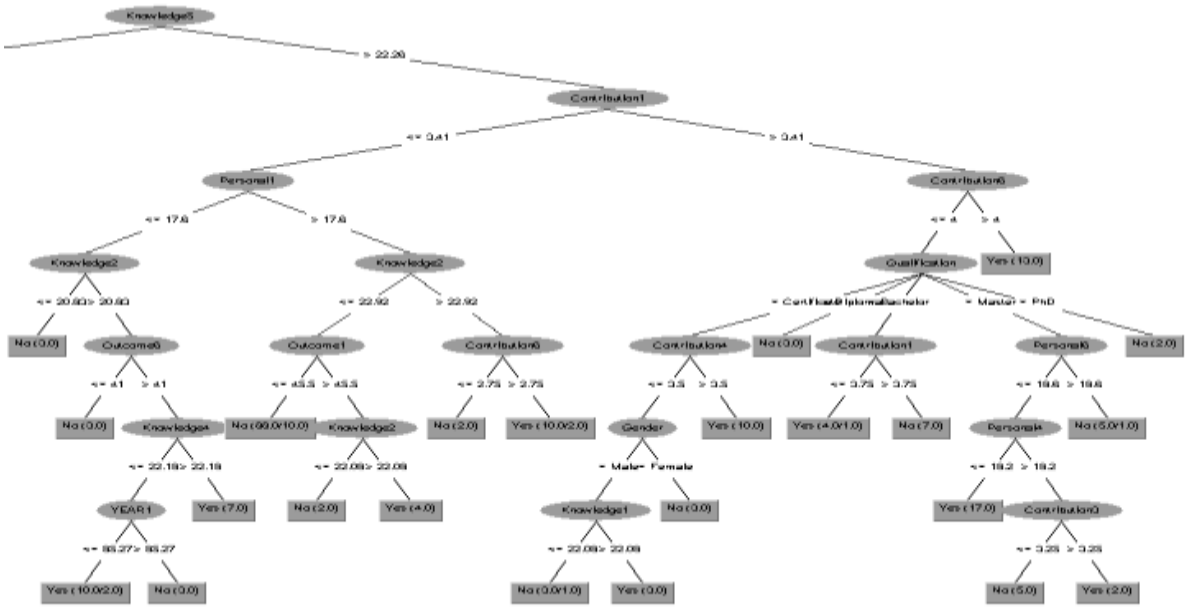


Figure 3: Sample of Rules using Decision Tree

The C4.5 classifier uses two heuristics criterion to rank possible tests. One is information gain by using attribute selection.

Factors	Attributes Hit	Attribute Name
Category(1)	1	*Category(1)*
Gender(1)	1	*Gender(2)*
Qualification(1)	1	*Qualification(1)*
Work Outcome (6)	4 (Year1, Year2, Year4, Year6)	*Outcome1(2), Outcome2(2), Outcome4(2), Outcome6(2)*
Knowledge and Skill (6)	6 (Year1, Year2, Year3, Year4, Year5, Year6)	*Knowledge1(2), Knowledge2(4), Knowledge3(1), Knowledge4(2), Knowledge5(1), Knowledge6(2)*
Individual Quality (6)	3 (Year1, Year4, Year6)	*Personal1(2), Personal4(1), Personal6(1),*
Activities and Contribution (6)	6 (Year1,Year2, Year3,Year4, Year5, Year6)	*Contribution1(2), Contribution2(1), Contribution3(2), Contribution4(1), Contribution5(1), Contribution6(2)*
Year Evaluation mark (6)	1 (Year 1)	*YEAR1*

For that reason, C4.5 classifier also can be worn to conclude the important or interesting attributes from the dataset. In this swot up, the important attributes are identified through the number of hits for each of attributes in the generated classification rules.

For that reason, in future work, other decision tree techniques such as NB Tree, Simple Cart, REP Tree, BF Tree and others will be tested to support this finding.

Conclusion:

This article has described the data mining techniques such as (SVM) Support Vector Machine, Fuzzy logic, (AIS) Artificial Immune System and many others should also be chew over for future work on classification techniques using the same data set .As we can see, the C4.5/J4.8 classifier has highest accuracy in the experiment. Thus, C4.5/J4.8 classifier algorithm is considered as a potential classifier for future work. In this experiment, the generated classification rules from organization process to ascertain whether he/she has the possible for promotion or not. In conclusion, the prediction in HR research has become the major contribution to data mining in HRM.

References:

1. ExecutiveBrief. (2008). 12 Competencies: Which Ones Should Your People Have. Retrieved 13/11/2008, from http://www.executivebrief.com.
2. Amin Zollanvari, Refik Caglar Kizilirmak, Yau Hee Kho, and Danieal Hernandex-Torrano, "Predicting Students' GPA and Developing Intervention Strategies Based on Self-Regulatory Learning Behaviors", 2017, IEEE.
3. AI-Radaideh,Q. A., AI-Shawakfa, E.M., and AI-Najjar, M. I.,"Mining Student Data using Decision Trees", International Arab Conference on Information Technology(ACIT'2006),Yarmouk University, Jordan, 2006.
4. Jantan, H., Hamdan, A. R., & Othman, Z. A. (2010). Human talent prediction in HRM using C4. 5 classification algorithm, International Journal on Computer Science and Engineering, 2(08-2010), pp 2526-2534
5. Subhamkarande, Ajay shelake, sivagami M, Sharon Sophia, Prediction of Employee retention using Cassandra and Ensemble learning,International Journal of Recent technology and Engineering(IJRTE) Vol. 8, No.1, May 2019.
6. Snikhan, Khera, Divya,Predictive Modelling of Employee Turnover in Indian It Industry using machine learning techniques, Journal published india,2019.
7. ZarminaJaffar, Dr.Waheed Noor, ZartashKanwal, Predictive Human Resource Analytics using Data mining classification techniques,International Journal of Computer, 2019.
8. Xiang Gao, Juhhao Wen, Cheng Zhang,An Improved Random Forest Algorithm for predicting Employee turnover,Hindawi, Mathematical problems of Engineering, 2019.
9. Amirah Mohamed Shahir, Wahidah Husain, Nuraini Abdul Rashid,A review on predicting students performance using data mining,The third information systems international conference, 2015.
10. Saurabh pal, Evaluation of Teachers performance: A Data mining Approach, January 2014.
11. Ernest H, ObooyleJr, Ronald H, Humpherey Jeffery M, Pollack Thomas H, Hawer, Paul A.Story, The relationship between job performance and emotional intelligence.
12. Brijesh Kumar Bharadwaj, SaurabPal,Mining Educational Data to analyse Students performance,January 2012.
13. Hameed, A., &Waheed, A.,"Employee Development and it's Affect on Employee Performance a Conceptual Framework". International journal of business and social science, 2(13), 2011.
14. Azar, A., Sebt, M. V., Ahmadi, P., &Rajaeian, A.,"A Model for Personnel Selection with A data Mining Approach: A Case Study in A Commercial Bank: Original Research".SA Journal of Human Resource Management, 11(1), 1-10,2013.
15. Amin Zollanvari, Refik Caglar Kizilirmak, Yau Hee Kho, and Danieal Hernandex-Torrano, "Predicting Students' GPA and Developing Intervention Strategies Based on Self-Regulatory Learning Behaviors", 2017, IEEE.