INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

Predicting Faculty Performance based on Self-Appraisal Form using Classification Algorithms

M.Saravanan, R.Geethamani, B.Kalaivani, A.Jothi, I.Ravi

Assistant Professor, Department of Computer Science, K.R College of Arts and Science, Kovilpatti,

Tamilnadu, India.

ABSTRACT: The main aim of this paper is to predicting the College Faculty's Teaching Excellence and Assessment performance and raising the salary and incentives for yearly basis using machine learning algorithms for data mining. Data Mining is used for information discovery of interest in Human Resource Management. Faculties are evaluated from their Appraisal form basis.This System enhances and predicts the performance of Faculties in colleges and used to fix the salary and incentives regarding their performance. The proposed idea in this paper is to perform an analysis the number of parameters for prediction indicators needed for faculty's performance evaluation and monitoring. Thispaper is used to predict the quality and potential of faculty across various disciplines which will enable higher level authorities to take decisions and understand certain patterns of teacher's motivation, satisfaction and progress of the faculties.

KEYWORDS: - Classification Algorithms, C4.5 Algorithm, C5.0 Algorithm, Weka tool

I. INTRODUCTION

Data mining has fascinated lot of attention in the research industry and in society as a whole in recent years, due to enormous availability of large amount of data and the need for turning such data into useful information and knowledge. Data mining, also called Knowledge Discovery in Databases (KDD), is the field of discovering new and potentially useful information from huge databases.

Human source has become one of the main concerns of organizers in almost all types of industries which include private companies, educational institutions and governmental organizations. Business Organizations are really interested to settle plans for correctly selecting proper employees. After hiring staffs, managements become anxious about the performance of these staff were management build evaluation systems in an attempt to conserve the good performers of Staff. 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 subtly 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.

II. RELATED WORK

Faculty Assessment is a process and a result: a way to determine goals, to appraise the processes for reaching students and to assess the extent to which they have been met. The Academic Senate encourages faculty evaluation to be done in a non-punitive, collegial atmosphere.

Faculty evaluation is a complex process; no single source of data is adequate. The evaluation process should begin with a written self evaluation. This encourages discussion and goal setting, giving increased value to the process.

The combined appraisals of students, colleagues, administrators, and faculty member's self-assessment are required for reasonably reliable and valid judgments.

Purposes Of faculty Evaluation

* Improvement of teaching and learning

* Retention/Promotion

* Systematic and regular review

* Staff Progress

* Professional growth and development

The information obtained from the evaluation may be used for any of the above purposes. For some purposes, such as tenure, the evaluation results are shared information; with others, such as professional growth or improvement of instruction, the information should remain confidential.

1. **Self Evaluation**: A good evaluation process begins with a written self-evaluation. This should be based on previous year evaluations and reflect continuity and reexamination of the goals. The faculty member being evaluated should examine the methods used for effective instruction, the strategies he or she uses to remain current in the discipline, participation in extracurricular activities as well as required activities and any other pertinent factors related to employment.
2. **Peer Evaluation** The purpose of a peer evaluation is to provide objective, professional assessment of the performance of role responsibilities, professional growth and extracurricular activities. The peer evaluator and evaluate should meet to discuss the written self-study and to decide how to conduct the remaining portions of the evaluation process.
3. **Administrator Evaluation** Administrators offer a different perspective on the faculty member being evaluated. They should be consulted for information that will contribute to the evaluation in the area of their expertise. Usually quality of service in the discipline is best left to the faculty in the discipline. Administrators are generally best equipped to evaluate the faculty member's service to the broader campus goals and off-campus professional activities.
4. **Student Evaluation:** "The quality of student learning is directly--although not exclusively--related to the quality of classroom teaching. Teachers need to receive specific comprehensible feedback on the extent to which they are achieving their goals and objectives."

In this paper, we discuss the teachers' performance evaluation using data miningclassification techniques at college teachers. The proposed model considers the various aspects of performance measures of teachers, like Students' Feedback,Self-evaluation, Peer-Evaluation, and Administrator Evaluation.Proposed model is designed to combine the knowledge and expertise of staff experts with reasoning capabilities that will provide a great support to the institution for decision-making in educational institutions

III. C4.5 ALGORITHM IN DATA MINING

C4.5 is the successor to ID3 and removed the constraint that features must be categorical by dynamically defining a discrete attribute that partitions the continuous attribute value into a discrete set of intervals. C4.5 converts the trained trees (i.e. the output of the ID3 algorithm) into sets of if-then rules. This accuracy of each rule is then evaluated to determine the order in which they should be applied. Pruning is done by removing a rule's precondition if the accuracy of the rule improves without it.

The C4.5 algorithm is a famous algorithm in Data Mining. The C4.5 algorithm acts as a Classifier in Decision Tree. C4.5 is a data mining algorithm and it is used to generate a decision tree. The C4.5 algorithm is very helpful to generate a useful decision that is based on a sample of data.

When we generate the decision trees with the help of C4.5 algorithm, then it can be used for classification of the dataset, and that is the main reason due to which C4.5 is also known as a statistical classifier.

So, before starting the C4.5 algorithm, you must revise the Decision Trees and how Decision Trees can be used as classifiers in data mining.

Algorithm to generate C4.5 decision tree

Input: an attribute –valued dataset D

1. Tree={ }

2. if D is "pure" or stopping criteria met then

3. terminate

4. else if

5. for all attribute a∈D do

6. Compute information-theoretic criteria if we split on a

7. end for

8. a_{best} = Best attribute according to above computed criteria

9. Tree = Create a decision node that tests a_{best} in the root

10. D_v = induced sub –datasets from D based on a_{best}

11. For all D_v do

12. $Tree_v$= C4.5(D_v)

13. Attach $Tree_v$, to the corresponding branch of tree

14. end for

15. return Tree

IV.C5.0 CLASSIFIER

C5.0 algorithm is an extension of C4.5 algorithm which is also extension of ID3. It is the classification algorithm which applies in big data set. It is better than C4.5 on the speed, memory and the efficiency. C5.0 model works by splitting the sample based on the field that provides the maximum information gain. The C5.0 model can split samples on basis of the biggest information gain field. The sample subset that is get from the former split will be split afterward. The process will continue until the sample subset cannot be split and is usually according to another field. Finally, examine the lowest level split, those sample subsets that don't have remarkable contribution to the model will be rejected. C5.0 is easily handled the multi value attribute and missing attribute from data set.

Algorithm to generate C5.0 decision tree

Input

a. Data partition, D, a set of training tuples and their associated class labels

b. attribute_list, the set of candidate attributes c. attribute_selection_method, a procedure to determine the splitting criterion partitions the data tuples into individual classes. This criterion consists of a splitting_attribute and ,either a split-point or splitting subset

Ouput: C5.0 decision tree Method:

1. create a node N

2. if tuples in D are all of the same class, C, then 3. return N as a leaf node labelled with the class C

4. if attribute_list is empty, then

5. return N as a leaf node labelled with the majority class in D

6. apply attribute_selection_method(D, attribute_list) to find the best splitting_criterion

7. label node N with splitting_criterion

 8. if splitting_attribute is discrete-valued and multiway splits allowed then

9. attribute_list attribute_listsplitting_attribute

10. For each outcome j of splitting_criterion Let D_j be the set of data tuples in D satisfying outcome j if D_i is empty then attach a leaf labelled with majority class in D to node N else, attach the node returned by Generate C5.0 decision tree(D_j , attribute_list) to node N

11. Return N

In this paper, C4.5 classification and C5.0 classification have been implemented on the education institutions to evaluate the faculty performance and compared with its accuracy. Even though C5.0 is similar to C4.5, C5.0 handles all types of data like continuous, categorical, dates, times and timestamps. It mainly supported boosting to improve the classifier accuracy.

V. DATASETS

Data is collected from the institutions. In order to collect the required data, a question format was Prepared and distributed to faculty working in an institution. The inquiry form was filled by 450 Staffs. After the forms were collected, the process of preparing the data was accomplished. The data in the questionnaires has been transferred to Excel sheets. Then, the types of data has been reviewed and modified. These files are prepared and converted to (ARFF) format to be compatible with the WEKA data mining toolkit.

VI. WEKA TOOL

An open source software provides tools for data preprocessing, implementation of numerous Machine Learning algorithms, and visualization tools so that you can develop machine learning techniques and apply them to real-world data mining problems. What WEKA offers is summarized in the following diagram −
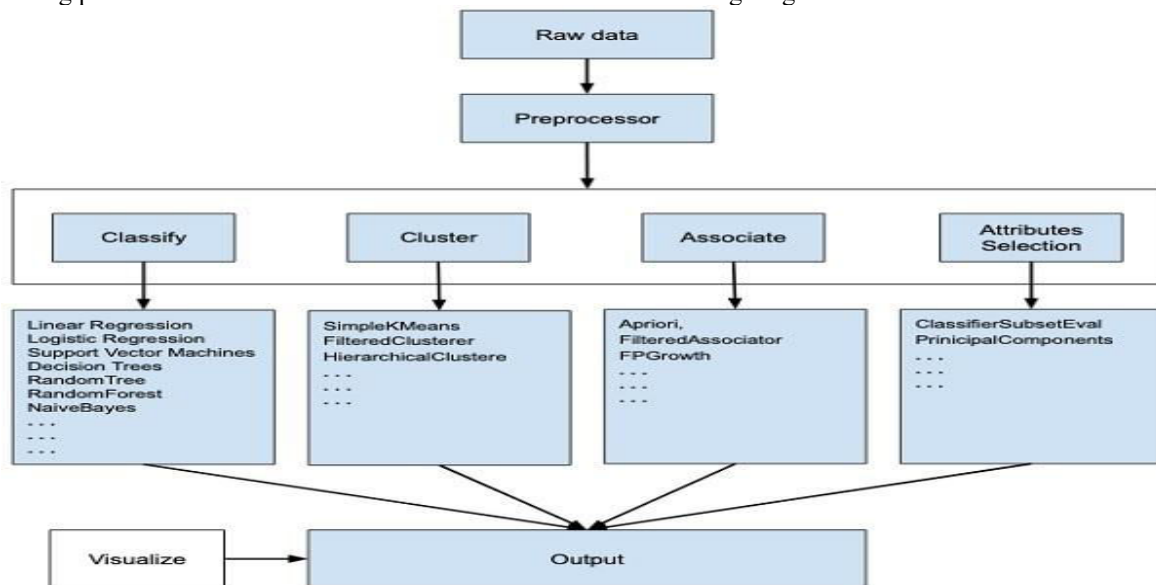
In Weka tool, inserted some sample data in the form of ARFF file. The data areStaff_Name, Level, Marks obtained and Grade. Level Attribute consists of two fields like L1 and L2. L1 represents the above 15 years of experience and L2 represents below 10 years of experience.

Marks attribute denote the total marks scored in all the fields and also in competency level (interview). Grade attributes denotes the overall performance value like Good, OK and Satisfied. The sample data are represented in the ARFF file format.

@relation staff

@attribute StaffName {Saravanan,Ravi,Kalaivani,Geethamani,Jothi,Gandhi}

@attribute Level {L1,L2}

@attribute Got_Marks numeric

@attribute grade {Good,OK,Satisfied}

@data

Saravanan,L1,35,Good

Ravi,L1,32,Good

Kalaivani,L2,31,OK

Geethamani,L2,30,Satisfied

Jothi,L2,32,OK

Gandhi,L2,30,Satisfied

Table 1: Staff Data set

Attribute Name	Value	Description
Lectures, Seminars, Tutorials, Practical's, Contact Hours	Numeric Value	Workload of a particular staff
Teaching Load in excess of institute norms	Numeric Value	Extra work performed by the staff
Percentage of each subject taught	Numeric Value	Semester wise percentage level
Ph.D. completion	Yes/no	Upgrading the staff research interest
NET/SET/JRF Online course(Npteletc)	Yes/no	Passing Eligibility test and upgrading skill is required
No.of papers presented in National/International conferences	Numeric Value	To know the recent research areas
No.of papers published in Journal	Numeric Value	To know the recent research platforms
Student's Feedback Mark	Numeric Value	Students will give the feedback about understanding of subject, communication level of the staff (given as form)
Marks obtained from Competency level	Numeric Value	Subject Experts will evaluate and give marks for each staff
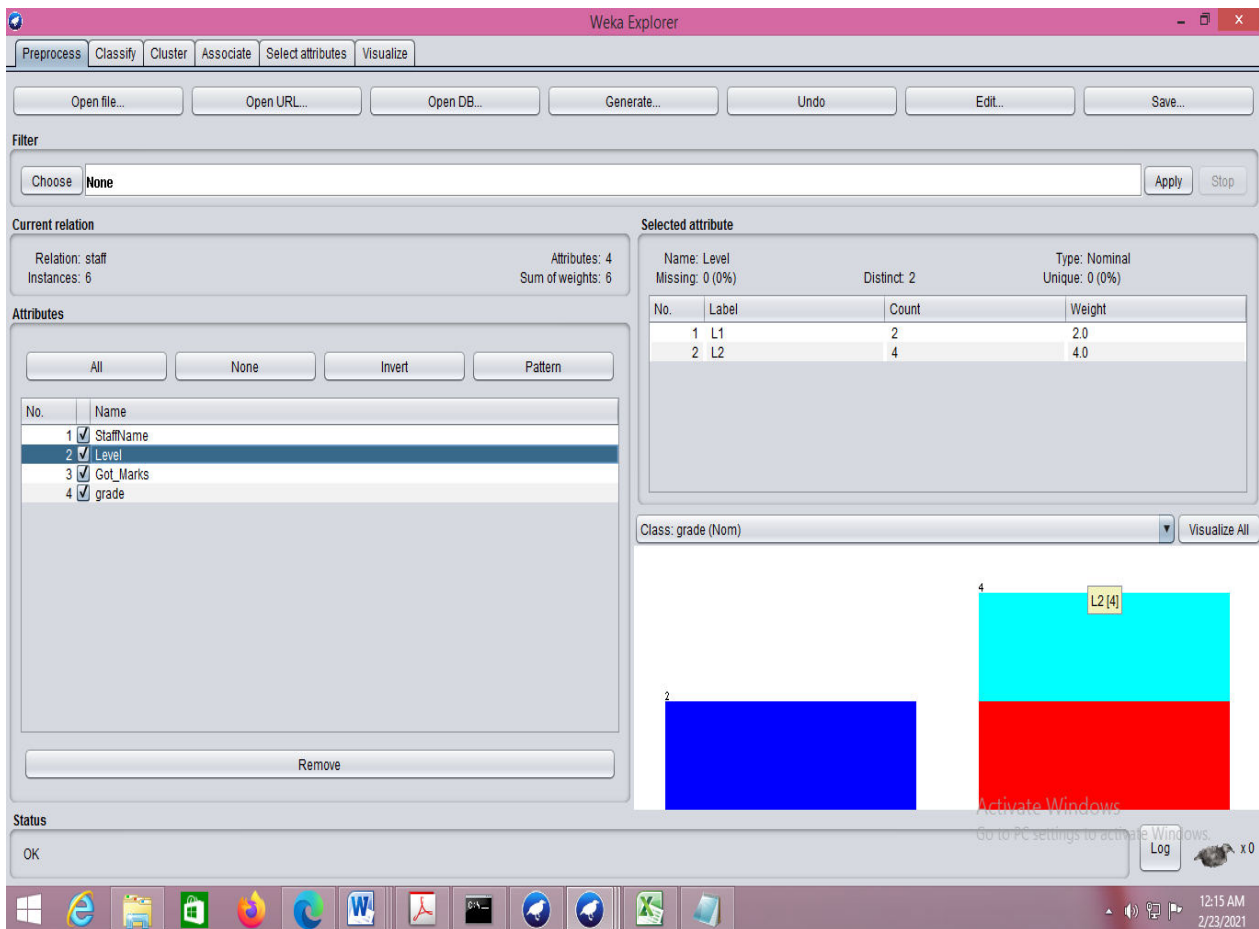
VII. CONCLUSION AND FUTURE SCOPE

In this paper, classification techniques are used for prediction on the dataset of 150 faculties in institution, to predict and analyze their performance as well skill among them. In this study, a model was developed based on some selected faculty related input variables collected from institutions.In future, Integration of data mining techniques with Weka tool and Digital Learning techniques is combined together on different datasets to find accuracy and predictions of desired results. The problem also should be identified if the attribute value is huge, whether it provide correct result or not, This datasets will applied on recent classification algorithms to evaluate the performance analysis in future. Also with new classification tool will be applied to predict the faculties performance by adding more data sets collected from institutions.

REFERENCES

1. Jantan et al. "Towards applying Data Mining Techniques for Talent Management", IPCSIT vol.2, IACSIT Press, Singapore, 2011.
2. Jantan, H., Hamdan, A.R. and Othman, Z.A.(2010b). "Human Talent Prediction in HRM using C4.5 Classification Algorithm", International Journal on Computer Science and Engineering, 2(08-2010), pp. 2526- 2534.
3. Sivaram, Ramar 2010."Applicability of Clustering and Classification Algorithms for Recruitment Data Mining," International Journal of Computer Applications, vol 4, pp.23-28, 2010.
4. V. Kalaivani, Mr. M. Elamparithi. An Efficient Classification Algorithms for Employee Performance Prediction. International Journal of Research in Advent Technology, September 2014.2(9). E.ISSN: 2321-9637.

5. HamidahJantan, Norazmah Mat Yusoff and MohamadRozuan Noh.Towards Applying Support Vector Machine Algorithm in Employee Achievement Classification.

6. Raorane, A, A., Kulkarni, R, V., "Data Mining: An effective tool for yield estimation in the agricultural sector", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume.1, Issue.2, August 2012, pp: 75-79.

7. Nasira, G, M., Hemageetha, N., "Vegetable Price Prediction Using Data Mining Classification Technique", Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering, March 2012, pp: 100-102.

8. Mehta, S, T., Kathiriya, R, D., "Survey of Data Mining Techniques in Precision Agriculture", IJSR - INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH, Volume.4, Issue72, July 2015, pp: 363-364 [4] .

INTERNATIONAL JOURNAL
OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING