



International Journal of Control Theory and Applications

ISSN : 0974-5572

© International Science Press

Volume 10 • Number 18 • 2017

Big Data Analysis for Teacher Recommendation using Data Mining Techniques

Atta-ur-Rahman¹ and Sujata Dash²

¹ College of Computer Science & Information Technology (CCSIT), University of Dammam, Saudi Arabia.

Email: aaurrahman@uod.edu.sa

² North Orissa University, Odisha, India. Email: Sujata238dash@gmail.com

Abstract: In this research, investigation of data mining techniques for institutional survey of teacher evaluation data is proposed. This includes data punching, data cleansing, data fusion, data dimensioning, analysis, visualization and prediction. Motivation behind this thesis is to create a decision support system for the guidance of administration of an institute. Administration will be able to find out strengths and weaknesses of faculty members. It will help admin to take necessary steps for the guidance/counseling of faculty and to appreciate those who perform well. Data has been punched into database files having proper schema. System will help administration to find out various results in no time without any problem. In addition to finding weaknesses of faculty members and identifying various issues, the system will also be able to identify trends and behavior of faculty members with ease, allowing counseling of faculty members and allocation of suitable courses to teachers. Apriori algorithm, weighted average prediction, Lagrangian method and Regression Models are used for trend analysis and prediction. Results demonstrate effectiveness of the scheme.

Keywords: Big data, Data mining, Teacher evaluation, Apriori Algorithm, Regression analysis, Prediction

1. INTRODUCTION

Nowadays educational data mining (EDM) governs, new data mining techniques and other innovative research approaches that understand how students learn the data. Proper e-learning tools and techniques have created opportunities to scrutinize and gather students' data, to determine pattern and trends in given data, and to apply assumptions and new discoveries about learning of students. Enhanced technology learning heavily depends on two systems provided. Course Management System (CMS) or Learning Management System (LMS). These management systems save every key pressed by individual user as logs of server automatically. Mining of these logs gives many patterns which can recognize late learners and adjust different strategies of teaching. Gathered data from learning system and education systems can be aggregated to explore different data mining techniques and algorithms for large number of students. Shahzadi (2011) [1] proposed a semantic network for Holy Quran, in which the entire repository can be searched for a specific concept. Prior to the creation of semantic network the data was preprocessed like noise word removal, concept network building and addition of new concepts with

expert approval etc. After successful completion of the Semantic Word Net, an algorithm was proposed to make the search optimum. Shahazadi (2012) [2] proposed and designed a classifier for the verses of the Holy Quran using Semantic Networks, in which any verse from the Holy Quran can be categorized into one of the predefined categories like pillars of islam, prophets and day of judgment etc. In this work, firstly a semantic network is designed by taking into account many translations of Holy Quran, Ahadees and Bible and then the verses were categorized. Atta-ur-Rahman (2013) [3] proposed a teacher assessment and profiling system (TAPS) using a fuzzy rule based system (FRBS) and Apriori Algorithm. In this paper, FRBS was used to assess the user while Apriori Algorithm was used as profiling technique. Significance of the proposed scheme was shown through examples. System can help organization to sort out weaknesses of faculty members and to increase the quality of education provided to the students. Association and relationship among data items has been found using Apriori algorithm.

Atta-ur-Rahman (2013) [4] proposed a fuzzy rule based system (FRBS) for network user's behavior classification. The proposed scheme makes use of the users' various logs like web, machine and network logs to classify him/her behaviorally. Simulation results are presented to signify the effectiveness of proposed scheme. Web, databases, hardware and other application logs provide information of user to classify. System helps in securing privacy of system/user and also can secure network from un-authorized and criminal mind users. Proposes work will help organization to make precautionary measures to secure the system. Bhardwaj and Pal (2011) [5] studied performance of student of different colleges and universities. 300 students were selected for this purpose. After applying classification method of Bayesian classification, different factors such as grade of student in senior secondary exam, teaching medium, location of student, qualification of mother, family income and some other factors highly effect performance of student. It was concluded that mother's low education affects the students' performance. Otherwise students have good grades. Performance of students also effects on living area and family factor as well. Home tuition is also an important factor considered in student's good performance. In this research, investigation of data mining techniques for teachers' evaluation data is proposed for sake of recommendation, important data visualization, profiling, trend analysis and prediction. Rest of the paper is organized as follows. Section 2 contains the introduction to data mining techniques, section 3 explains the proposed approach, section 4 shows the results while section 5 concludes the paper.

2. DATA MINING TECHNIQUES

Mining on education data uses different methods such as Apriori Algorithm [6], Decision Trees [7], FP Tree Algorithm [8], Naïve Bayes, K- Nearest neighbor (KNN) [9], ID-3 algorithm [11], Neural Networks [12] and many others. By these mining methods, data can be revealed such as classifications, clustering and associating rules. Knowledge received is useful for predicting data about student's enrollment in a particular course, finding of illegal techniques, finding of grades that are abnormal, predicting student's behavior and performance.

3. PROPOSED EDUCATIONAL DATA MINING (EDM)

Before applying the data mining techniques, the data must in a proper format. That requires following steps. Fig-1 shows the proposed approach for educational data mining.

- Data processing (Formatting, cleansing, removal of unnecessary data)
- Completion of missing data (if some examples are missing in the data set, they must be provided)
- Data mining

3.1. Data format

The data to be analyzed is in the form of teachers' evaluation forms containing students' feedback. This feedback is obtained once from the students about the teacher for each course they are taking in the semester. Following are the questions in feedback form.

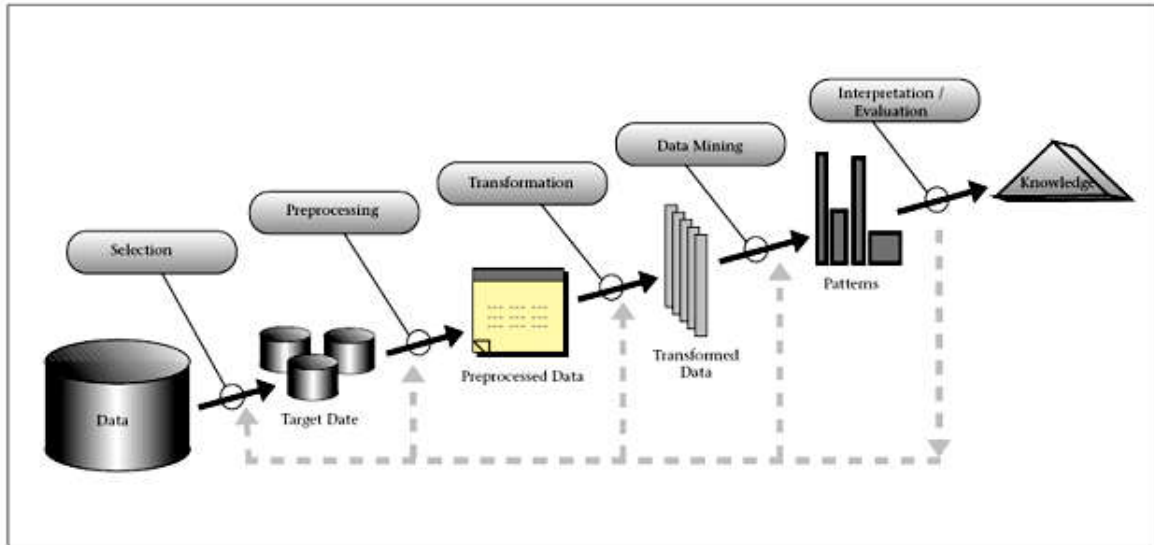


Figure 1: Data Mining Flow Chart

Question No. 1: Knowledge of teacher about the subject?

Question No. 2: Ability of this teacher to explain clearly?

Question No. 3: Is the teacher fair and impartial in treatment of all students?

Question No. 4: Ability of this teacher to maintain good discipline?

Question No. 5: Is the teacher friendly, considerate, patient and helpful?

Question No. 6: Is the teacher audible to the whole class, especially to the last row?

Question No. 7: The ability this teacher has to make classes interesting?

Question No. 8: The teacher uses board extensively to write definitions, draw diagrams, show and solve problems?

Question No. 9: Does the teacher involve the students during lecture?

Question No. 10: Has the teacher completed the whole course, as per course outline?

Question No. 11: Does the teacher provide additional material apart from the textbook?

Question No. 12: Does the teacher arrive on time?

Question No. 13: Does the teacher return graded papers and quizzes in a reasonable amount of time?

Question No. 14: Was the teacher available and helpful during the specified office hours and after class consultations?

Question No. 15: The general (all-round) teaching ability of this teacher?

Answer for each question ranges from 1 to 5, representing Poor, Fair, Average, Good and Excellent respectively. Schema for database has been designed adequately so that analysis must provide useful results to improve quality education provided to students. Data of each form has been punched into Microsoft Excel sheet manually. Punching data manually is quite a tough job. Afterwards excel sheet data has been exported to SQL server database file. Data conversion to SQL server was required so that maximum use of tool would help the system work more efficiently and accurately. Especially security and confidentiality of data is more important. SQL server offers more security as compared to Microsoft Excel. Fig-2 shows the format of excel sheet that contains the filled form data for a particular class and teacher.

Tname	Course	Session	Sno	Class	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15
Farukh	FA	FALL2013	1		5	5	5	5	5	5	5	5	4	5	5	4	4	5	5
Farukh	FA	FALL2013	2		4	5	3	1	5	5	4	4	4	5	4	4	3	4	4
Farukh	FA	FALL2013	3		4	5	3	4	4	5	4	3	4	5	4	5	3	4	5
Farukh	FA	FALL2013	4		5	5	3	2	4	4	5	4	5	3	5	5	5	5	5
Farukh	FA	FALL2013	5		4	4	5	4	3	5	5	5	3	4	5	5	3	5	4
Farukh	FA	FALL2013	6		4	3	1	3	3	3	3	3	3	3	3	5	3	3	3
Farukh	FA	FALL2013	7		5	4	1	3	1	5	5	3	1	5	3	5	5	2	5
Farukh	FA	FALL2013	8		5	4	5	3	5	5	3	4	3	3	5	5	3	5	5
Farukh	FA	FALL2013	9		5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Farukh	FA	FALL2013	10		5	4	5	5	3	4	2	4	3	5	4	5	4	4	5
Farukh	FA	FALL2013	11		4	3	2	1	5	5	5	2	4	4	2	5	1	4	3
Farukh	FA	FALL2013	12		3	3	2	1	3	2	4	3	2	3	4	3	3	3	2
Farukh	FA	FALL2013	13		3	2	4	4	4	4	3	2	5	3	3	4	4	2	5

Figure 2: Sample excel data sheet

After that this sheet is imported into Microsoft SQL Server 2010. Once data has been imported into SQL server, a dashboard for result generation has been designed in ASP .net MVC (modeled-view-controller) platform. Number of views has been designed to get a wider perspective of data. Aggregate results have been generated using Structured Query Language (SQL). Description of the views is presented below

- Evaluation of tutor in each question.
- Evaluation of tutor in his/her specific course.
- Evaluation of tutor in his/her specific course for specific session.
- Overall best 3 tutor of the semester.
- Comparison between different tutors.
- Comparison between different tutors on the basis of specific question.
- Evaluation of different tutors on the basis of a specific course.
- Best tutor in a course.
- Best tutor in a course in a specific question.

These queries are handled by the data mining techniques. Data mining is a process to find out hidden knowledge from big data sets. This process is also known as Knowledge Discovery in Databases (KDD).

4. RESULTS

Analysis of teacher's evaluations has been conducted using aggregates. Number of views has been created to show behavior of teachers in particular area. Bar charts have been created which displays the analysis results. An average function has been used to calculate overall evaluation of different aspects. Followings view has been generated on the basis of data punched.

4.1. Data Visualization

4.1.1. Teachers Evaluation Aggregate in specific course in fall 2013

Fig-3 and fig-4 show the aggregate of teacher evaluation against each question described above for two teachers respectively. From the given controls, one can choose the teacher, then the courses being taught by selected teacher will automatically populated in next dropdown in a specific semester. These values can be seen in table-1.

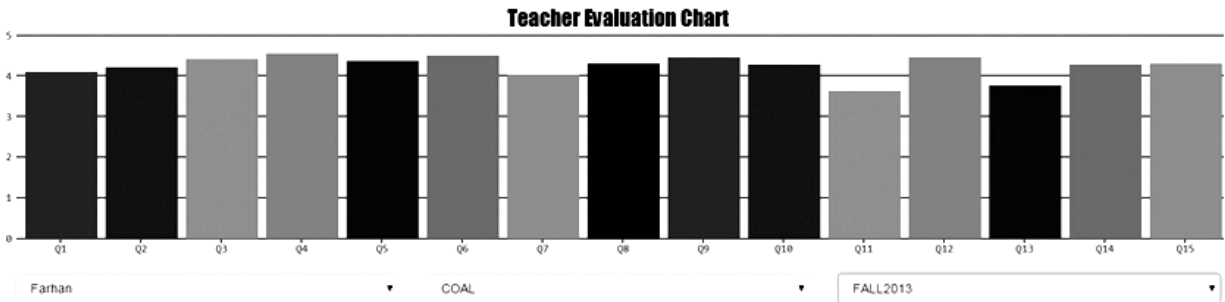


Figure 3: Teachers Evaluation Aggregate 1

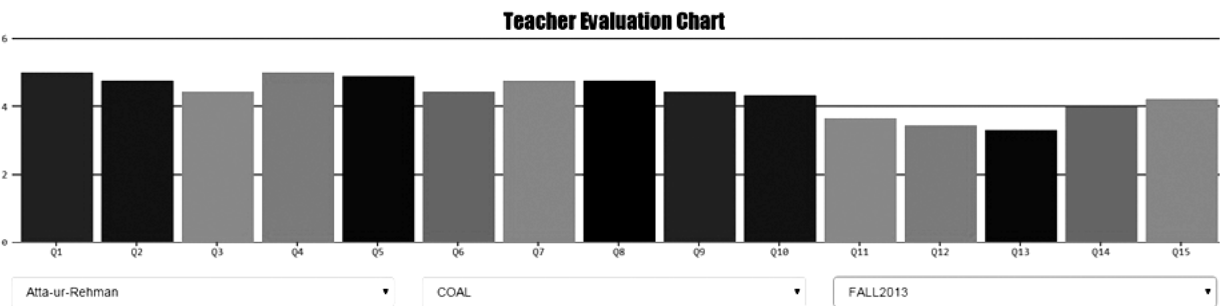


Figure 4: Teachers Evaluation Aggregate 2

Table 1
Teachers Comparison

Question Number	Teacher1 (Atta)	Teacher2 (Farhan)
1	5	4.11
2	4.7777	4.2234
3	4.4444	4.4301
4	5	4.5425
5	4.8888	4.3723
6	4.4444	4.5106
7	4.7777	4.0319
8	4.7777	4.2978
9	4.4444	4.4565
10	4.3333	4.2765
11	3.6666	3.6382
12	3.4444	4.4680
13	3.3333	3.7765
14	4.0	4.2872
15	4.2	4.3085

Table-1 is comparing data of two teachers of Computer Organization and Assembly Language (COAL) subject. This data is useful for organization to sort out weaknesses and strengths of a faculty member and for proper counseling. Courses can be assigned to teachers on the basis of their evaluations so that quality education can be provided to students. This data is useful for organization. Courses can be assigned to teachers on the basis of their evaluations so that quality education can be provided to students.

4.1.2. All Teachers Evaluation Aggregate according to specific question.

Fig-5 and fig-6 show the aggregate score of all the teachers against Q2 and Q15 respectively and the numerical values can be seen in table-2. This view provides the strengths and weaknesses of all the teachers in a particular question. If administration wants to see them in this way, they can take the decision whether to council a teacher etc.

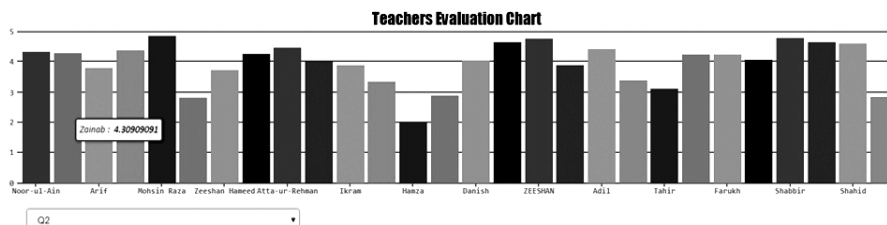


Figure 5: Aggregate According to a Specific Question (Q2)

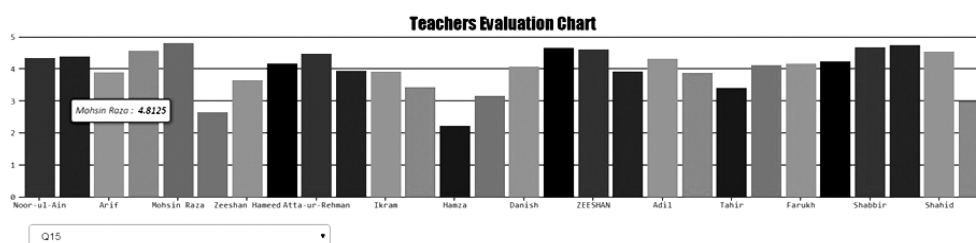


Figure 6: Aggregate According to a Specific Question (Q15)

Table 2
All Teachers Evaluation in a particular question

Teacher Name	Aggregate Result
Noor-ul-Ain	4.3474
Zainab	4.3909
Arif	3.9024
Amir	4.5800
Mohsin Raza	4.8125
Bilal	2.6666
Zeeshan Hameed	3.6666
Khalid	4.1739
Atta-ur-Rehman	4.4873
Nauman	3.9562
Ikram	3.9346
Shafique	3.4339
Hamza	2.2424
Naseer	3.16
Danish	4.0838
Neelam	4.6635
Zeeshan	4.6153
Mohsin Hassan	3.9248
Adil	4.3203
Sidra	3.875
Tahir	3.4045
Munir	4.1338
Farukh	4.1686

4.1.3. Top 3 teachers of all sessions

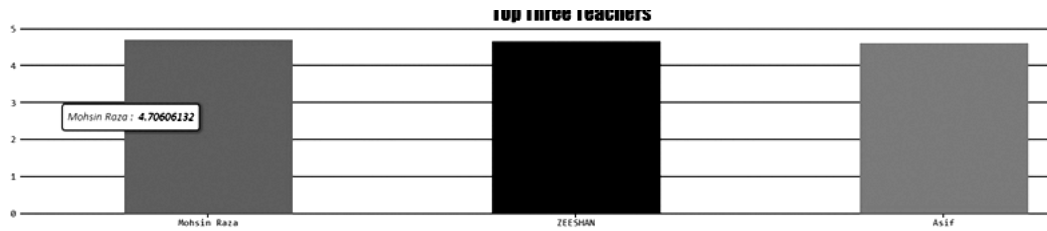


Figure 7: Top three teachers of all sessions

Table 3
Top 3 Teachers

Teacher Name	Over-all Aggregate
Mohsin Raza	4.7060
Zeeshan Hameed	4.6666
Asif	4.6265

Fig-7 and table-3 are showing the top three faculty members of overall evaluation in all semesters and in all courses. Best teacher award ceremony is held in BIIT for top three for best top three faculty members in all courses.

4.1.4. Teacher’s Individual Evaluation Analysis for All Questions

Fig-8 and table-4 shows the teacher’s individual evaluation analysis for all the questions. This can help highlighting the individual teacher’s strong and weak areas according to students evaluation. Evaluation of Mr. Azhar is given as an example.

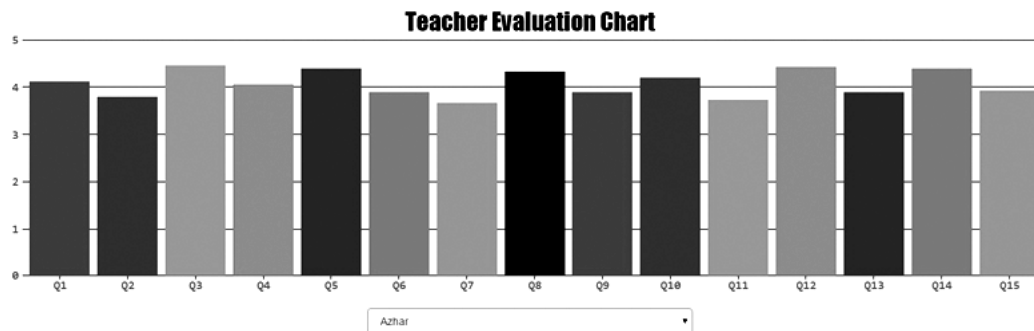


Figure 8: Teacher’s Individual Evaluation Analysis for All Questions

Table 4
Teacher’s Over-All Evaluation

Question Number	Aggregate Result
1	4.1333
2	3.8
3	4.4666
4	4.0666

(contd...Table 4)

Question Number	Aggregate Result
5	4.4
6	3.9
7	3.6666
8	4.3333
9	3.9
10	4.2068
11	3.7333
12	4.4333
13	3.9
14	4.4
15	3.9333

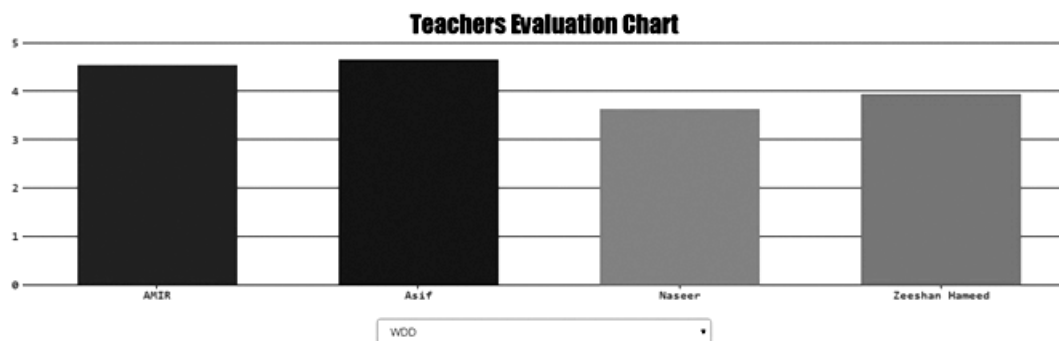


Figure 9: Ranking according to a specific course

4.1.5. Teachers Evaluation According to Specific Course

Fig-9 and table-5 show the comparative analysis of more than one teacher, teaching the same course (Web design and development). This can be of great help to see which among many teachers is teaching the subject effectively.

Table 5
All Teachers evaluation against a shared course

Teacher Name	Over-all Aggregate
Amir	4.5500
Asif	4.6590
Zeeshan Hameed	3.9507
Naseer	3.6364

4.1.6. Teacher's Comparison

In the above dashboard (fig-10), user can select multiple faculty members from the given list to see the comparison based on their overall score. User can select as many teachers as he wants.

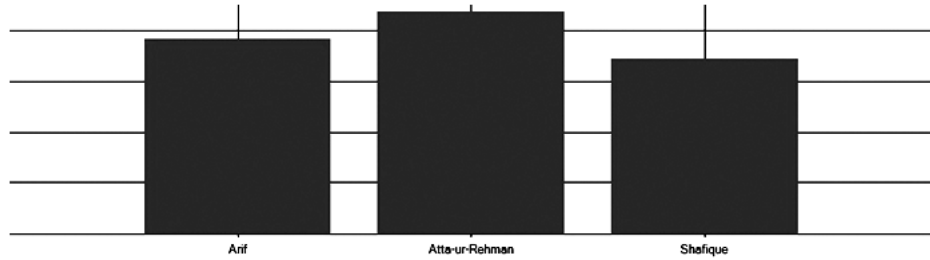


Figure 10: Comparison of different teachers

4.1.7. Future Prediction by Weighted Average and Lagrange Interpolation Formula

In fig-11 and 12 the weighted average formula and Lagrange algorithm are used to predict the next value from the given historic values of teacher evaluation. This algorithm will predict the evaluation of any teacher of next session selected from the given list box. This will help administration to forecast the performance of a teacher (Mr. Munir).

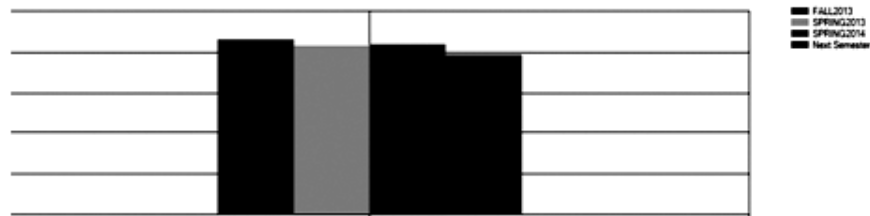


Figure 11: Weighted Average Prediction

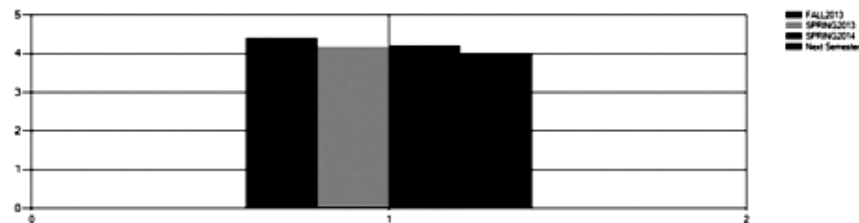


Figure 12: Lagrange Interpolation Formula

4.1.8. Evaluation of Teachers According to Specific Profile

Below dashboard (fig-13) is designed to view overall trends in evaluation according to last university attended in his/her last degree by faculty members.

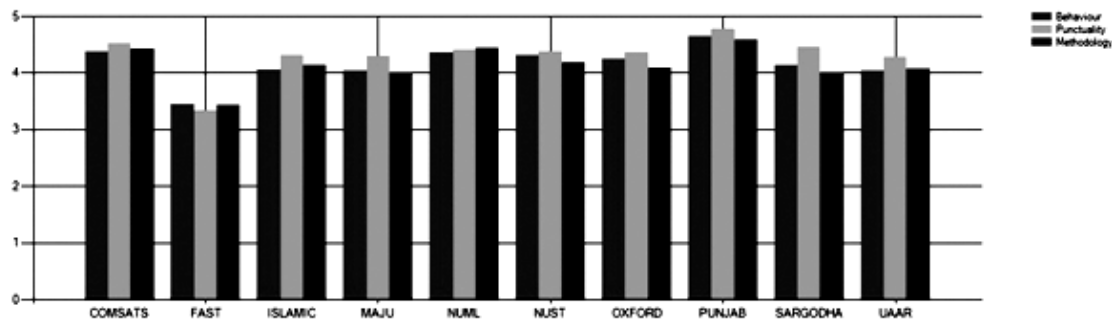


Figure 13: Evaluation According to Last Degree

4.1.9. Evaluation of Teachers According to Age Group

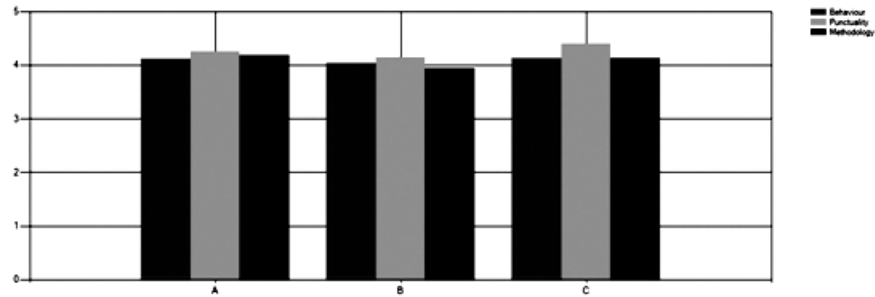


Figure 14: Evaluation According to Age Factor

Age of faculty members is grouped in three groups.

- less than 25
- between 25 and 28 and more than 28

On basis of these groups, evaluation has been displayed in the form of graph, which will help finding the effect of age factor on teacher’s evaluation.

4.1.10. Evaluation of Teachers According to his/her Mother tongue

A dashboard (fig-15) is included to find the effect of mother tongue factor on teacher’s evaluation. This will help institute to hire proper teachers for the betterment of institute.

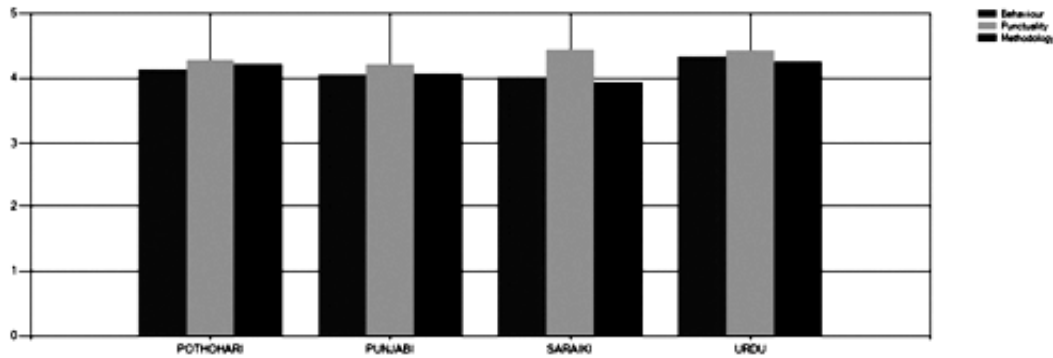


Figure 15: Evaluation According to Mother tongue Factor

5. CONCLUSION

The main purpose of this research was to develop a decision support system for the institutions to analyze teachers’ evaluation submitted by students. Teachers’ evaluation is a procedure held by administration in each semester. Students are provided evaluation form for each course they have enrolled in that particular semester. It is quite a tough job to analyze thousands of forms. It takes a lot of time and effort to analyze evaluation of teachers individually as well as comparison of different teachers. Evaluation form consists of fifteen questions. Each question is designed by a team of highly professional members so that teacher’s evaluation must be up to the mark. Data of previous evaluations has been gathered and punched into soft form. System has been developed which read data of faculty members and courses they have taught throughout their carrier. By proper analysis,

different views have been created which display aggregate results of teacher's performance throughout the semester. Following views has been designed which are helpful for organization make decisions about course allocation in coming semesters, to encourage hardworking faculty members and to eliminate weak zones of teachers if any.

- Evaluation of tutor in each question.
- Evaluation of tutor in his/her specific course for specific session.
- Overall best 3 tutor of the semester.
- Comparison between different tutors on the basis of specific question.
- Evaluation of different tutors on the basis of a specific course.
- Best tutor in a course.
- Best tutor in a course in a specific question.

Various data mining techniques were investigated in this regard. The uniqueness of the research is that there is no current evaluation process which can analyze results. All work was manually done by organization which was time consuming and did not cover all aspects of analysis.

REFERENCES

- [1] Atta-ur-rahman, "Teacher Assessment and Profiling using Fuzzy Rule based System and Apriori Algorithm", *International Journal of Computer Applications (IJCA)*, Vol. 65, No. 5, pp. 22-28, 2013.
- [2] Atta-ur-Rahman, D.N. Zaidi, M.H. Salam and S. Jamil, "User Behavior Classification using Fuzzy Rule Based System", *13th International Conference on Hybrid Intelligent Systems (HIS'13)*, pp. 118-123, December 04-06, 2013, Tunisia.
- [3] Shahzadi, N., Atta-ur-Rahman and A. Shaheen, "Semantic Network based Semantic Search of Religious Repository", *International Journal of Computer Applications (IJCA)*, Vol. 36, No. 9, pp. 1-5, December, 2011.
- [4] Shahzadi, N., Atta-ur-rahman and M.J. Sawar, "Semantic Network based Classifier of Holy Quran", *International Journal of Computer Applications (IJCA)*, Vol. 39, No. 5: pp. 43-47, February, 2012.
- [5] B.K. Bharadwaj and S. Pal, "Mining Educational Data to Analyze Students' Performance" *International Journal of Advance Computer Science and Applications (IJACSA)*, Vol. 2, No. 6, pp. 63-69, 2011.
- [6] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules", *Proceedings of the 20th VLDB Conference*, Santiago, Chile, 1994.
- [7] J. Jha and L. Ragha, "Educational Data Mining using Improved Apriori Algorithm" *International Journal of Information and Computation Technology*. Vol. 3, No. 5, pp. 411-418, 2013.
- [8] J. Han, J. Pei, Y. Yin and R. Mao, "Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach", *Data Mining and Knowledge Discovery*, Volume 8, Issue 1, pp 53-87, January 2004.
- [9] A. Likas, N. Vlassis and J. Verbeek, "The Global K-Means Clustering Algorithm", *Pattern Recognition*, vol. 36, pp. 451-461, 2003.
- [10] J. R.Quinlan, "Induction of Decision Trees." *Mach. Learn.* 1, 1, 81-106, March 1986.
- [11] Mitchell, Tom M. *Machine Learning*. McGraw-Hill, 1997. pp. 55-58.
- [12] Atta-ur-Rahman, Qureshi I.M., Malik A.N., Naseem M.T. "A Real Time Adaptive Resource Allocation Scheme for OFDM Systems using GRBF-Neural Net. & Fuzzy Rule Base System", *International Arab Journal of Information Technology*, vol. 11 (6), pp. 590-598 2014.