A Student Performance Prediction and Course Recommendation System

Tupe Supriya Deepak

Masters in Computer Engineering,
SPPU, R. H. Sapat College of Engineering Management Studies and Research Name, Nasik, India

***Abstract*: Automatic detection of student's degree performance to help them cope up with failures and dropouts has become a very essential need. Using past performances of students in previous semesters, a student of a degree course is categorized best or lowest into 3 categories Theory, Testing and practical So that he/ she would focus on those particular categories in the next semesters. Also, Students are recommended as to which elective subject he/she should opt for based on those results for better scores.**

***Index Terms:* Clustering, Collaborative Filtering, Ensemble learning, Fuzzy rules, Matrix Factorization.**

I. INTRODUCTION

Higher Education has become the most important and essential investment to build the future of students. Parents with huge education loans are worried about their child's studies and their marks. The huge college expenses or the college fees have sky reached from a few couples of years making it hard for low-income families to pay these fees. Parents taking educational loans are worried whether their child would do something fruitful in life. In spite of paying huge fees students who don't do well in exams results in low grades even a year gap because of the subjects, they have failed. This low performance or year gaps will naturally affect their future. This fact motivates to design a Student Performance Prediction System and in addition, also Recommend Elective subjects to those students for better outcomes as Students often are confused for selecting Elective Subjects.

Students move to higher courses, the amount of fees keeps increasing with their courses. So to save students from dropouts and paying the fees, again and again, it has become important to know the performances of students beforehand.

Thus, the student will come to know in which specific subject he needs more learning so that he/she could concentrate on those specific subjects for better scores.

II. RELATED WORK

Jie Xu, et al. [1] proposed a method that uses a machine learning technique to predict the performance of students. The performance of students depends upon the past performances of the student that is the academic scores of the student. The bi-layered structure and the data-driven approach are the two most basic features used in this system. Next to the System also uses the base prediction and the Ensemble prediction layer as the most important concepts for prediction.

Yifan Hou et al. [5] implemented a method for Massive Open Online Courses for handling huge data sets using Big Data techniques. For better performance of students also courses were recommended to them.

A prediction system was proposed by Christopher G. Brinton, et al. [3] using the CFA prediction algorithm to find whether an answer is correct or not in Massive Open Online Courses (MOOC). Support Vector Machines are used to classify the features more precisely. Matrix Factorization concepts are used to lower the dimensionality of the data, in addition, KNN (K Nearest Neighbor) algorithm to predict the accuracy. Pearson correlating coefficient is calculated to find the similarity between 2 variables.

A Student prediction system was designed by Carlos Mrquez Vera, et al. [4] to find out student failure in exams and students who are dropouts from colleges.

Data Mining approaches were used by Yannic Meier, et al. [5] to find out the final grades of students in college programs. The students who tend to be weak are predicted in this system in Massive Open Online Courses (MOOC) so that appropriate safety measures could be taken before he/she drop out.

Jie Xuet, et al. [8] implemented a methodology using a forward search backward induction algorithm. It gives an arrangement of suggestions. The system uses the forward search backward induction algorithm to find the accurate and correct GPA (Grade Point Average) of the students. Multi-armed bandit theory is adopted in the system.

Cem Tekin, et al. [7] proposed a framework to train both the base predictor offline and ensemble predictors online using the Hedge Bandits to predict the performance of the students.

Man Ching Yeun, et al. [9] proposed a method of task suggestion and allocation performed by laborers. Legitimate work or undertaking is allocated to the best and proper laborer so that the organization will benefit from it and would get higher benefits. Matrix Factorization is the concept used for task allocation to get higher benefit. Matrix Factorization is the procedure utilized for suggesting proper undertakings in publicly supporting frameworks. Task recommendation will help labors to complete their part of work with more dedication.

Adaptive learning techniques were introduced by YA huei Wang, et al. [10] to analyze student's performance using vast techniques of data mining. ANN (Artificial Neural Network) is further used to improve the prediction of Performance.

In the system proposed by Ruslan S, et al. [11] handling of the huge dataset is done. The system can easily handle the data. Probability Matrix Factorization was used.

Yueshen Xu, at al. [14] used a recommender system using Collaborative Filtering methods to rate the recommendations of the prediction. This will help the system to be more accurate.
Andrew E. Waters, et al. [15] used a bi-convex algorithm to solve the Sparse Factor Analysis Problem. The Sparse Factor Analysis Problem. This methodology can help diminish educators' outstanding burden.

III. SYSTEM OVERIEW

A. Problem Statement:

To Predict the Performance of Undergraduate students of ongoing courses and recommend them the Elective subject that should be taken.

B. Contribution:

As a contribution, a gene-fuzzy model was developed based on" The survival of the fittest" theory optimizing the final results using a Genetic algorithm.

The information on Real-World data information is often flawed and missing in specific practice and is at high risk which may contain huge errors. After uploading a dataset, data is validated as of whether the data is appropriate with none of the missing values or is there in duplicate entries in the data. Such data is of no use to the system and may conclude to wrong results or outcomes. This imperfect data should be resolved and made error-free for further processing.

Student Grade is calculated using the formulae [1] and [2]. A Matrix of dimensionality m*n is constructed consisting only of the required attributes of the data that are the marks of students in each subject of the semesters. A Probabilistic Matrix Factorization model is built to perform course clustering. Fuzzy Rules of each student are generated is categorized into Expert, Average and Low performance in subjects.

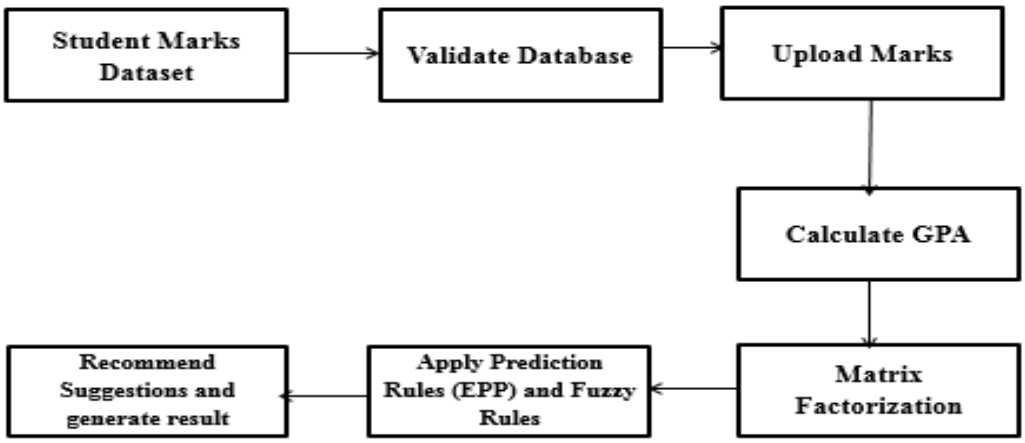

Fig1: System Architecture

The Students Semester Marks of Each subject is taken as an input to the predictors for training. Subject Recommendation uses the past choices of the clients and utilizing those choices to predict the choices of other clients. The concepts used in the system are:

A. Matrix Factorization: This is a Collaborative Filtering Method used to decompose items in lower dimensionality space. It is essentially utilized for the computation of complex framework activity.
B. Probabilistic Matrix Factorization: is a method to group similar clusters together using the probability measures.
C. Collaborative Filtering: Learns Records of previous Students.
D. Gene Fuzzy Model: Rules are predefined to obtain more accurate results than the previous model.
E. Clustering: Groups Students using k means clustering algorithm to 3 groups based on the better performance in theory, testing and practical.

IV. ALGORITHM

A. The Algorithm proceeds with the following abbreviations initialized as:
s: Student, c:course, C: total number of courses, gpa: Grade point average, cr: Credit Points

Step1: Initialize Parameters s, c, C, gpa, total_ grade, cr, cc;
Step2: Observe the subject marks of each student
Step3: For every subject course c=1 to C observe total marks and assign Grades to find total_grade.

total_grade = (cc * cr) (1)

Step4: Using the grades, calculate gpa

$$gpa = \frac{total\_grade}{cc} \quad (2)$$

Step5: Compute Matrix Factor of lower dimensionality

Step6: Build Probabilistic Matrix Factorization model

Step7: Generate fuzzy rules for each student s

Step8: Cluster the students best in specific categories

Step9: For each student s, display best and lowest performance subjects.

Step10: Recommend courses using Collaborative filtering

B. Fuzzy Rules Generation

The following Rules are predefined and generated for each student in the system for more accurate results.

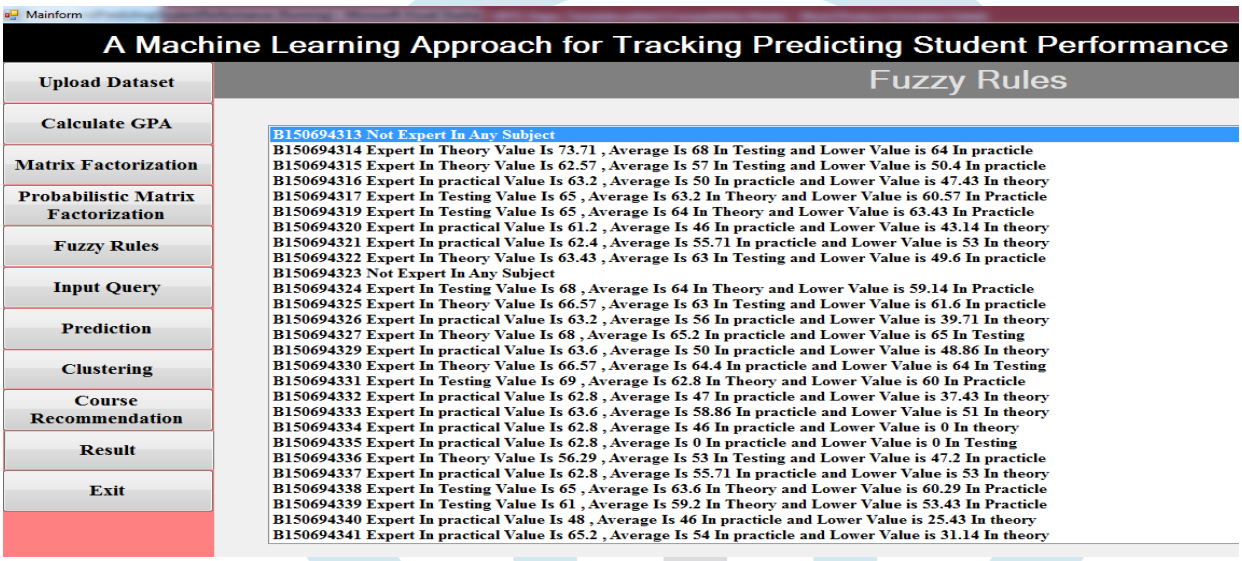

Fig 2: Generation of Fuzzy rules

V. PERFORMANCE MEASURES

1. Mean Square Error Rate: For a perfect model, the mean square error rate is always zero. There is no negative error rate for any model. This measure calculates the average prediction error rate of the system.
2. Precision: The measure used to find how good the system is with the same results and same inputs at different times.

$$Precision = \frac{True\ positive}{True\ positive + False\ Positive}$$

3. Recall: A measure that the relevant results are successfully retrieved.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

4. F1 Measure: used to test the accuracy of the system. It considers both precision p and recall r.

$$F1\ Measure = 2 * \frac{Precision * Recall}{Precision + Recall}$$

VI. DATA SET

Data Set	Engineering Student Dataset (R.H SAPAT College of Engineering)
Students Year of passing	2018-2019
No. of attributes	16
Year of Engineering	TE, BE COMPUTER
Size Of Data	75,776 bytes

VII. RESULT AND IMPLEMENTATION

Fig 3. Shows that Performance Prediction best and lowest of a student of R. H. Sapat College of B.E Computer Engineering and Course or Subject recommended for him in next semesters:

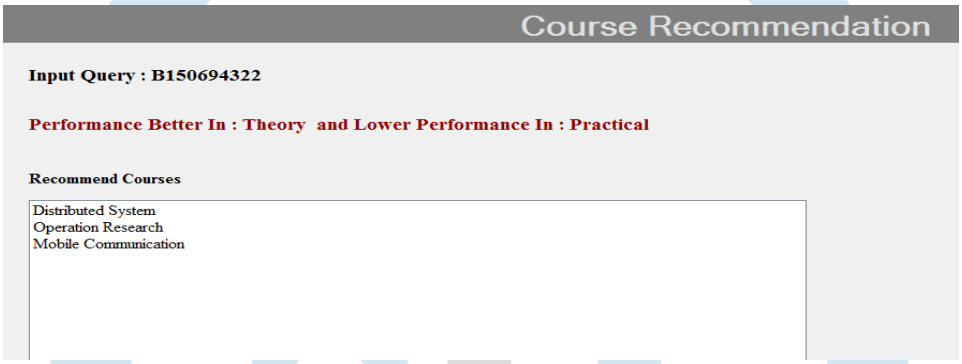

Fig 3: Course Recommendation

Fig 4 shows the Mean square error rate, precision, recall, F1 Measure calculated by the overall system.

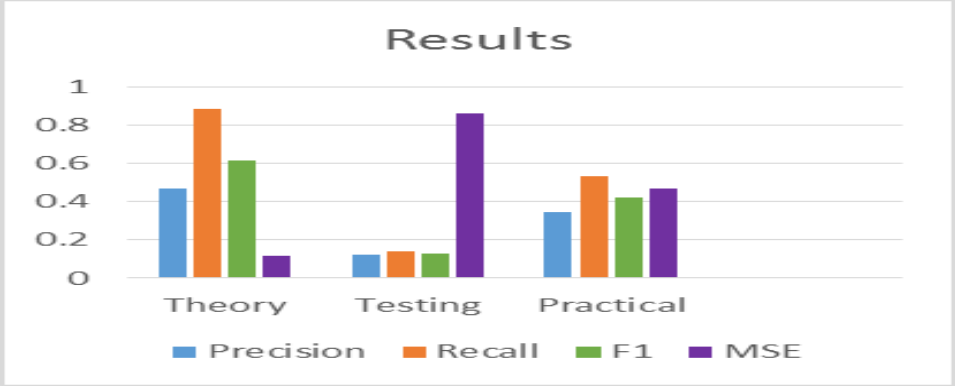

Fig 4: Result

VIII. CONCLUSION

A Student Performance Prediction System is proposed based on the students Previous semester marks Theory, Testing and Practical as best or low in which of the following 3 categories In addition the students who lack in specific categories will be recommended or suggested to opt for the elective subjects to be taken for further improvement in next semester so that the final score will be better than the previous score. Fuzzy rules are generated for more accurate results.

IX. ACKNOWLEDGMENT

I am truly indebted and thankful to the Guide for their valuable guidance and encouragement. I would also like to thank Staff members of the Computer Department of GESRHSCOE.

REFERENCES

[1] Jie Xu, A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs, IEEE Journal of Selected Topics in Signal Processing,2017

[2] Tupe Supriya Deepak, "A Student Performance Prediction and Course Recommendation System: A Survey" , International Journal of Scientific Research in Computer ScienceApplications and Management Studies (IJSCRSAMS), Vol 7, Issue 5 (2018).

[3] C. G. Brinton and M. Chiang, Mooc performance prediction via clickstream data and social learning networks, in 2015 IEEE Conference on Computer Communications (INFOCOM). IEEE, 2015, pp. 2299 2307.

[4] C. Marquez-Vera, C. Romero, and S. Ventura, "Predicting school failure using data mining," in Educational Data Mining 2011.

[5] Y. Meier, J. Xu, O. Atan, and M. van der Schaar, Personalized grade prediction: A data mining approach, in Data Mining (ICDM), 2015 IEEE International Conference on. IEEE, 2015, pp. 907912. (2002) The IEEE website. [Online]. Available: http://www.ieee.org/

[6] Yifan Hou, Course Recommendation of MOOC with Big Data Support: A Contextual Online Learning Approach, 2018 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS): KCN 2018: Knowledge Centric Networking.

[7] C. Tekin, J. Yoon, and M. van der Schaar, "Adaptive ensemble learning with confidence bounds," arXiv preprint ar X iv:1512.07446, 2015.

[8] J. Xu, T. Xing, and M. van der Schaar, Personalized course sequence recommendations, IEEE Transactions on Signal Processing, vol. 64, no. 20, pp. 53405352,Oct 2016.

[9] M.-C. Yuen, I. King, and K.-S. Leung, Task recommendation in crowd sourcing systems, in Proceedings of the First International Workshop on Crowdsourcing and Data Mining. ACM, 2012, pp. 2226..

[10] Y.-h. Wang and H.-C. Liao, Data mining for adaptive learning in a tesl-based elearning system, Expert Systems with Applications, vol. 38, no. 6, pp. 64806485,2011. and control," Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.

[11] R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization," in NIPS, vol. 20, 2011, pp. 1–8.

[12] T. Mandel, Y.-E. Liu, S. Levine, E. Brunskill, and Z. Popovic, "Offline policy evaluation across representations with applications to educational games," in Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems, 2014, pp. 1077–1084.

[13] N. Thai-Nghe, L. Drumond, T. Horvath, L. Schmidt-Thieme, et al." Multi-relational factorization models for predicting student performance," in Proc. of the KDD Workshop on Knowledge Discovery in Educational Data. Citeseer, 2011.

[14] Y. Xu, Z. Chen, J. Yin, Z. Wu, and T. Yao, "Learning to recommend with user-generated content," in International Conference on Web-Age Information Management. Springer, 2015, pp. 221–232.

[15] A. S. Lan, A. E. Waters, C. Studer, and R. G. Baraniuk, "Sparse factor analysis for learning and content analytics." Journal of Machine Learning Research, vol. 15, no. 1, pp. 1959–2008, 2014Available: http://www.halcyon.com/pub/journals/21ps03-vidmar