CMPE 256: Large Scale Analytics

Summer2019

Individual Project Report

Individual Project Name:
Course Recommendation System

Advisor:
Professor Shih Yu Chang

Submission By:

Name	Email	SJSU ID
Premal Dattatray Samale	**premaldattatray.samale@sjsu.edu**	**012566333**

Table of Contents

1. Introduction:

Nowadays recommendation system is widely used in commercial and educational domain. In commercial industry recommendation systems are used to increase the sale of items. For example, Amazon.com, Netflix etc. All these recommendation systems are also helpful for users in choosing the right product.

In today' s world of digital media, many on-line or in class training courses are available considering needs of students. Many students find hard to select the right course, when they have little information about the course and lack of guidance. In the educational industry, incorporation of effective course recommendation system in university will help students in selecting courses and ultimately improving student's success rate and reducing stress.

Therefore, in this paper, solution is purposed to build intelligent course recommendation system using collaborative filtering methodology along with student's grade predictions. The main reason for selecting this methodology is to provide quality recommendation based on user's interest. Collaborative filtering recommendation algorithm suggests courses based on student interest and major. Additionally, similarity between students is evaluated based on historical preferences of student, using pearson correlation.

2. Literature Survey

In today's world of digitization recommendation systems are getting popular. Through various research papers I understood various techniques and limitations of recommendation systems. For course recommendation, latest technologies are content based filtering, collaborative filtering, hybrid-based recommendation.

Amer Al-Badarenah et al. proposed a method using new collaborative filtering for elective course recommendation for university students [1]. The authors provided new collaborative filtering using association rule algorithm. Experimental results have shown that association rule tool provides better result for giving recommendation to target students.

There are many experiments are performed to advise students about courses during course enrollment. One of such experiments is performed using collaborative filtering along with Ordered Weighted Averaging (OWA) operator. The authors Alican Bozyiit et.al. proposed solution considering the case of repeating a course and students' grades in the course for each repetition [2]. OWA is used to aggregate the score of each student. It is observed that OWA improved the quality of recommendation.

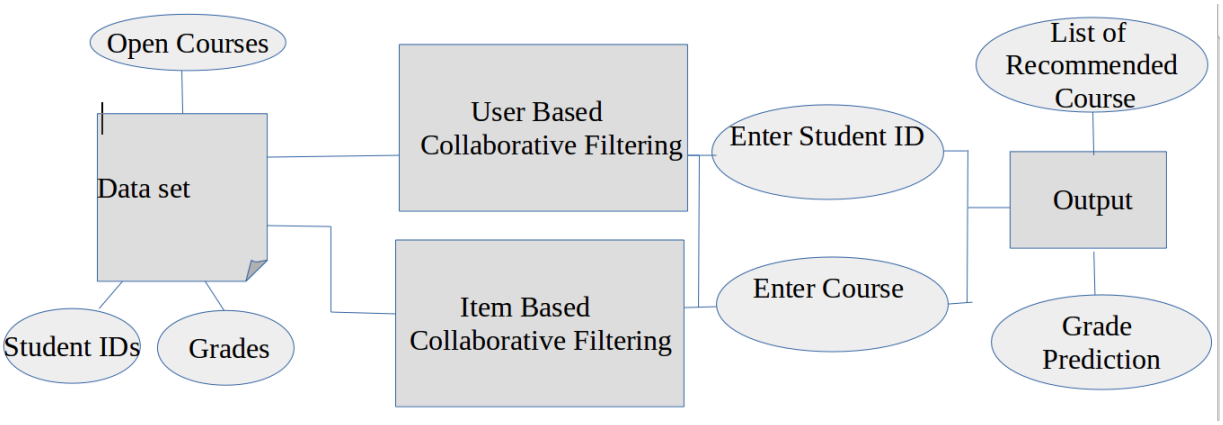
To improve user satisfaction of course recommendation, the authors Hualong Ma et. al. proposed a solution in year 2017 using semantic analysis. The authors used Each
course description as a document, and a built-in TFIDF,Word2Vec and Doc2Vec pipeline to improve the quality of recommendation.

The authors Huang, C. Y et.al. claim solution for course recommendation using ontology. In this student and experts are users of system. This solution provide adaptive learning recommendations and expert build program ontology according to their experience.

In conclusion, each technique has its own limitation and can be used based on specific requirement. In this paper, I am using collaborative filtering recommendation because it is best to predict grade for the course. Additionally, I have done a comparative study of item based and user based collaborative filtering.

3. System Design

There are multiple ways of implementation using collaborative filtering. In this paper user based and item based collaborative filtering approach is used to recommend courses. In user-based filtering algorithm, computed similarity between two students based on their earlier choices of courses. Then for suggesting new course, algorithms check into courses which user similar to target user preferred previously. For example, consider the real-life scenario where junior students take guidance of senior students of the same major to select courses. In item based filtering algorithm similarity between two courses are computed. And based on similarity score similar courses are recommended to students.



4.Method-Collaborative Filtering (CF)

In Collaborative filtering it is assumed that two similar user has similar interests.
In this method, two elements are present which are list of m students S= {S1 .... Sm} and list of n courses C= {C1 ...Cn}. Additionally, there is user-item matrix R [i, j] where i is students and j are courses. Collaborative Filtering method predicts based on the user-item matrix. Below is the user-item matrix [2].

$$R = \begin{bmatrix} r_{1,1} & \cdots & r_{1,n} \\ \vdots & \ddots & \vdots \\ r_{m,1} & \cdots & r_{m,n} \end{bmatrix}$$

4.1 User based collaborative filtering

The basic idea in user based collaborative filtering is that students who have similar past preferences, courses and majors are most likely make similar choices in the future [2].
This technique compares all students' behaviors and find out most similar students and make predictions of courses for the selected students. The first step is to calculate similarity score between selected student their using past choices.

There are many algorithms to calculate similarity between two students but Pearson -r correlation coefficient performs the best. Using this we calculated sim(u, v) i.e. similar value of user u with v [10].

$$Sim_{u,v} = \frac{\sum_{i \in I}(r_{u,i} - \bar{r_u})(r_{v,i} - \bar{r_v})}{\sqrt{\sum_{i \in I}(r_{u,i} - \bar{r_u})^2}\sqrt{\sum_{i \in I}(r_{v,i} - \bar{r_v})^2}}$$

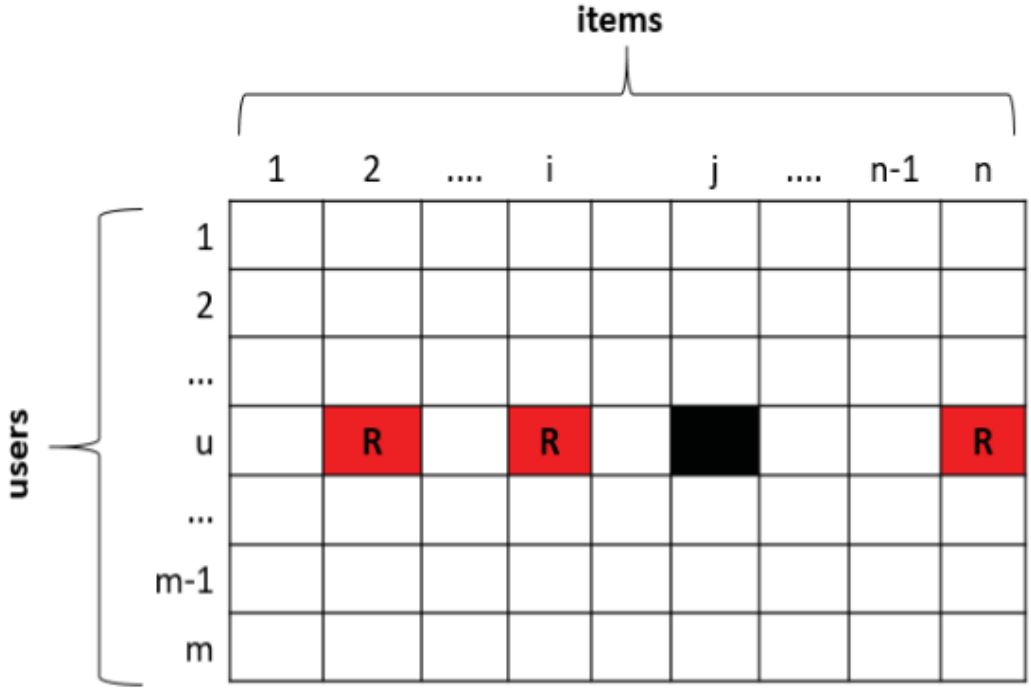
Here r u, i denotes the rating of user u for item i, and r u is the average rating given by user u calculated over all items rated by u. Similarly, r v denotes the rating of user v for item i, and r v is the average rating given by user v calculated over all items rated by v [10].
Now for the last step, grades of target student are calculated using an adjusted weight-age sum formula, to consider the fact that different students have different grades distributions.

$$P_{u,j} = \bar{r_u} + \frac{\sum_{v \in V} Sim_{u,v}(r_{v,i} - \bar{r_v})}{\sum_{v \in V} |Sim_{u,v}|}$$

4.2 Item based Filtering

The main idea behind item-based filtering is that it finds out similar courses. To predict the score for target course A by student B, the first step is to find a set of courses C that are most similar to Course A. The score in set C by student A are used to predict the score of course A.

To calculate similarity between the item i and item j, below figure shows the method. Color red boxed indicates the items which are previously rated by user u. The black boxes show score will be predicted [2].



Let the set of students who have scores of both courses i, and j be denoted by U, then similarity coefficient (Sim i, j) between them is calculated as [10]

$$Sim_{i,j} = \frac{\sum_{u \in U}(r_{u,i} - \bar{r_u})(r_{u,j} - \bar{r_u})}{\sqrt{\sum_{u \in U}(r_{u,i} - \bar{r_u})^2}\sqrt{\sum_{u \in U}(r_{v,i} - \bar{r_v})^2}}$$

Here r u, i denote the scores of student u for course i, and ru is the average scores scored by user u calculated over all courses previously taken by u. Similarly, r u, j denotes the rating of student u for course j.
To compute the predicted rating for a target item i for target user u, we use the following formula [10]

$$P_{u,j} = \frac{\sum_{j \in I} Sim_{i,j} * r_{u,j}}{\sum_{j \in I} |Sim_{u,v}|}$$

5. Data Collection:

First challenge in building intelligent course recommendation system is to collect proper data. Due to privacy reasons students' actual grades are not disclosed. Therefore, data is simulated to make data like real data. Initially implementation is performed on 30 courses and 50 students. Data is generated and stored in excel sheet and weight for each course is given based on random function and grades between 1 to 10 range.

StudentIDs	courseID1	courseID2	courseID3
1	4	3	2
2	10	2	5

6. Experimental Study:

In the experimental study, user based CF and item based CF are demonstrated for suggesting courses to students. Here, students are referred as users and courses are items to be predicted in these CF techniques.

Experiment 1: User based filtering

- First compute the similarity between student A and other students using pearson
- correlation.
- Then based on similarity score, select list of students who have high similarity score.
- After that, compute grade of student A for course selected by student A.
- Predicted grade is weighted average grades of other similar students.

Experiment 2: Item based filtering

- In item-based collaborative filtering, first, similarities between the courses are calculated.
- Then from the set of courses previously taken (score) by the target user, k courses most similar to the target course are selected.
- For computing the prediction for the target course, weighted average is taken of the target student's scores on the k similar courses earlier selected.

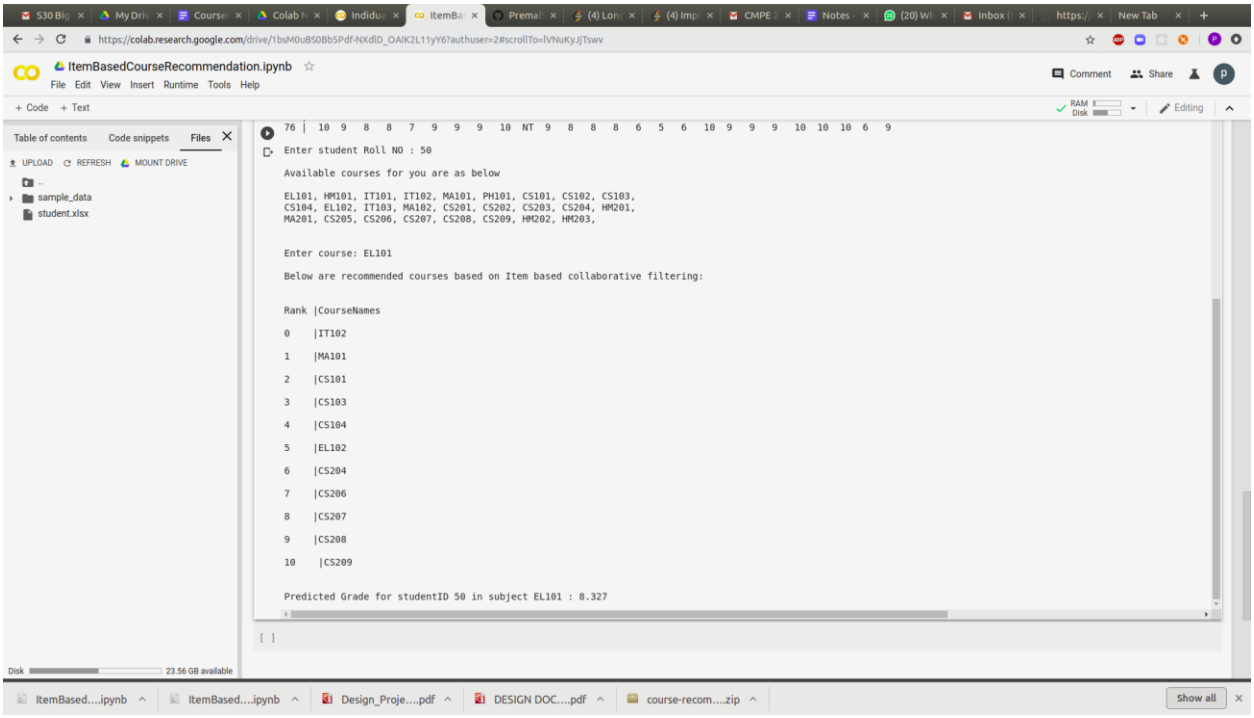
7. Instruction to run the code:

- Download the code from GitHub repository.
  https://github.com/PremalSamale/Course-Recommendation-System
- Open below two files in google colab.
  ItemBasedCourseRecommendation_PremalSamale.ipynb
  UserBasedCourseRecommendation_PremalDattatraySamale.ipynb
- Upload dataset students.xlsx file.
- Run the code.
- System will prompt as below
  Enter studentID:
  enter any id ex. 50 enter any no between 2 to 70.
- Then the system will display all available courses.
- System will prompt again:
  Enter course:
  Enter courses which are listed in the available courses
- System will show all recommended courses along with predicted score.

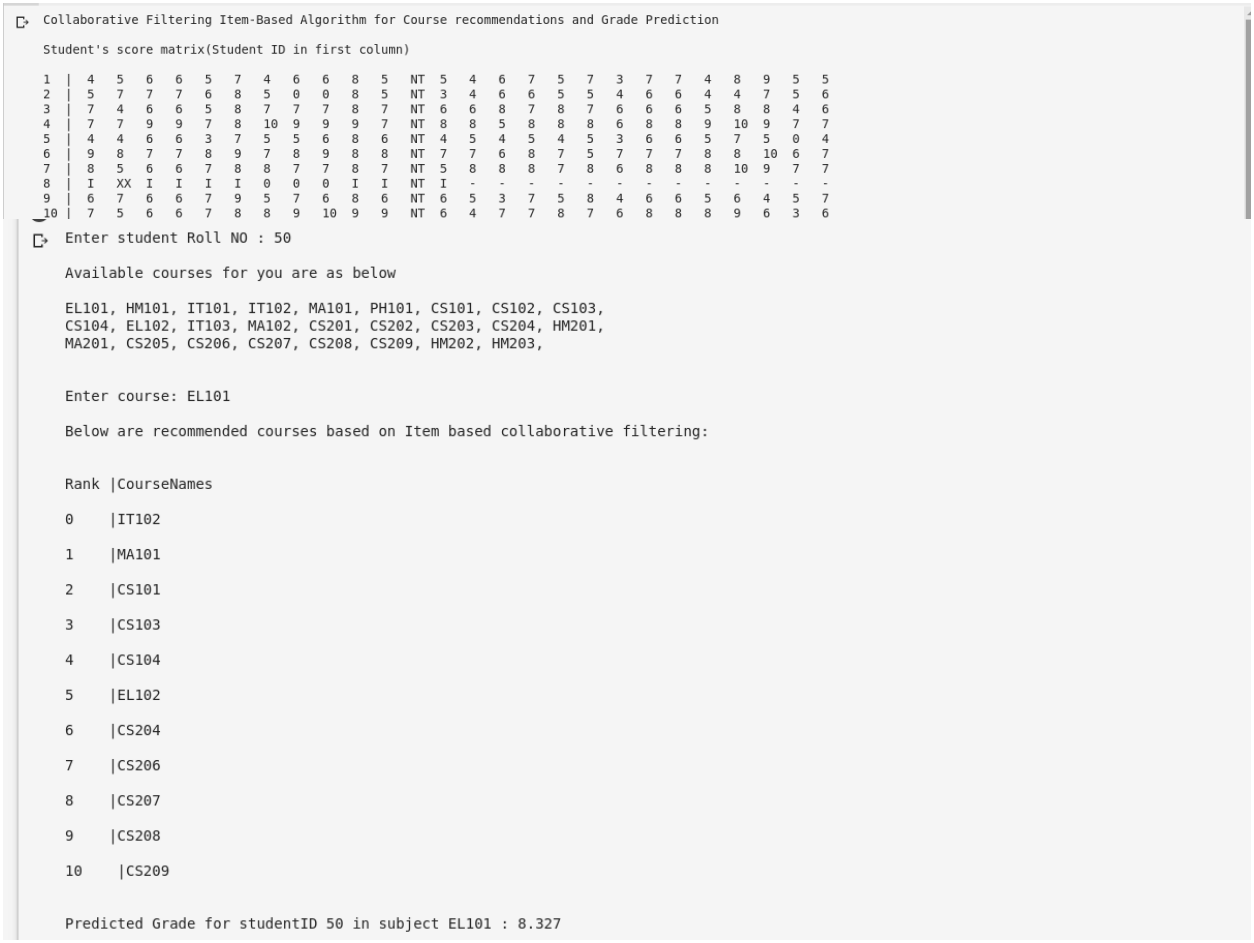
8.Results and Screenshots:

Below are the screenshots of outputs. System ask student to enter the roll no and then available courses are displayed to students. Then system ask student to enter course id to get to know grade prediction. Then system shows recommended courses and evaluate grade of selected course based on similar other students' grades.

Item based filtering: Screenshot 1:
Below is the result of item based collaborative filtering.
System is showing recommended courses and its rank for selected student and course.
System is also predicting the grades of course.

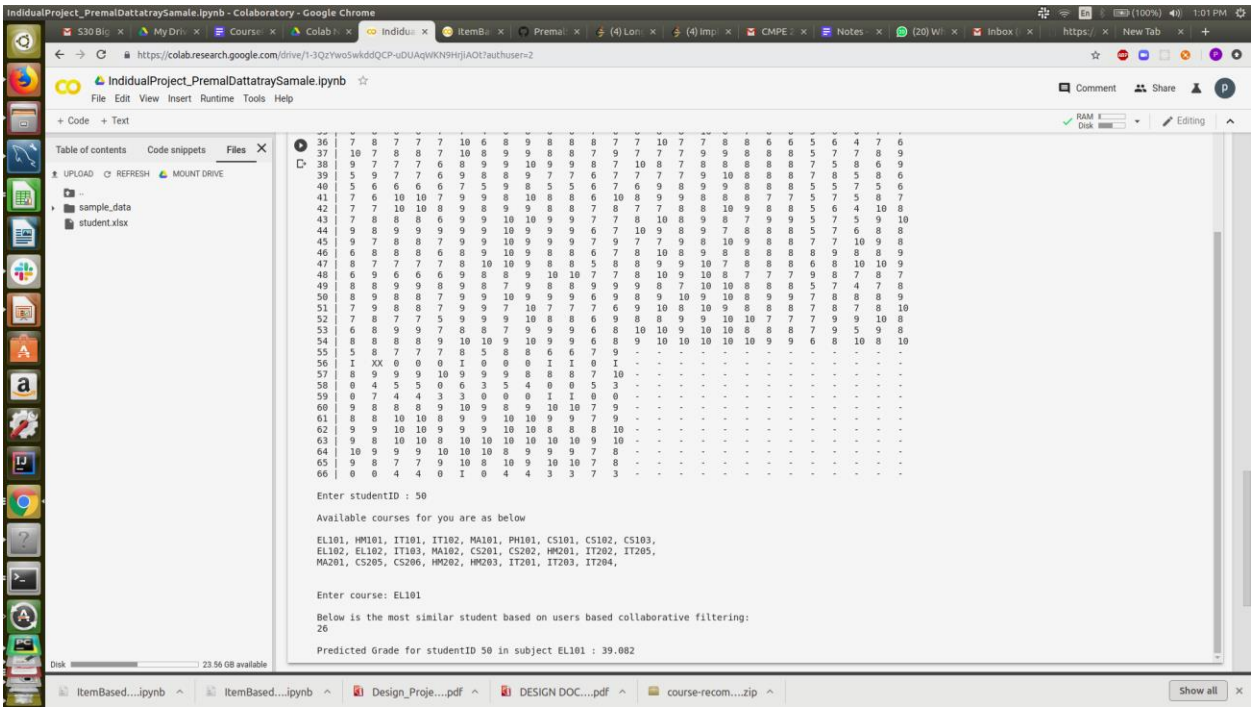


Item based filtering: Screenshot 2:



```
Collaborative Filtering Item-Based Algorithm for Course recommendations and Grade Prediction

Student's score matrix(Student ID in first column)

1  |  4   5   6   6   5   7   4   6   6   8   5   NT  5   4   6   7   5   7   3   7   7   4   8   9   5   5
2  |  5   7   7   7   6   8   5   0   0   8   5   NT  3   4   6   6   5   5   4   6   6   4   4   7   5   6
3  |  7   4   6   6   5   8   7   7   7   8   7   NT  6   6   8   7   8   7   6   6   6   5   8   8   4   6
4  |  7   7   9   9   7   8   10  9   9   9   7   NT  8   8   5   8   8   8   6   8   8   9   10  9   7   7
5  |  4   4   6   6   3   7   5   5   6   8   6   NT  4   5   4   5   4   5   3   6   6   5   7   5   0   4
6  |  9   8   7   7   8   9   7   8   9   8   8   NT  7   7   6   8   7   5   7   7   7   8   8   10  6   7
7  |  8   5   6   6   7   8   8   7   7   8   7   NT  5   8   8   8   7   8   6   8   8   8   10  9   7   7
8  |  I   XX  I   I   I   I   0   0   0   I   I   NT  I   -   -   -   -   -   -   -   -   -   -   -   -   -
9  |  6   7   6   6   7   9   5   7   6   8   6   NT  6   5   3   7   5   8   4   6   6   5   6   4   5   7
10 |  7   5   6   6   7   8   8   9   10  9   9   NT  6   4   7   7   8   7   6   8   8   8   9   6   3   6

Enter student Roll NO : 50

Available courses for you are as below

EL101, HM101, IT101, IT102, MA101, PH101, CS101, CS102, CS103,
CS104, EL102, IT103, MA102, CS201, CS202, CS203, CS204, HM201,
MA201, CS205, CS206, CS207, CS208, CS209, HM202, HM203,


Enter course: EL101

Below are recommended courses based on Item based collaborative filtering:


Rank |CourseNames

0    |IT102

1    |MA101

2    |CS101

3    |CS103

4    |CS104

5    |EL102

6    |CS204

7    |CS206

8    |CS207

9    |CS208

10    |CS209


Predicted Grade for studentID 50 in subject EL101 : 8.327
```

User based filtering: Screenshot 3: Below is the result of user based collaborative filtering.



User based filtering: Screenshot 4:

```
Collaborative Filtering User-Based Algorithm for Grade Prediction
Student's score matrix(roll no in first column)

1  |  4   8   5   5   6   9   0   6   5   3   3   6   5   3   3   3   4   0   0   0   0   4   5   0   4   4
2  |  0   5   4   4   0   5   4   6   6   5   5   6   6   3   6   5   4   0   0   6   6   4   5   0   5   3
3  |  0   7   5   5   5   6   0   6   0   4   4   7   5   4   3   4   4   4   3   6   6   4   5   0   4   5
4  |  10  9   8   8   7   9   9   9   10  9   9   8   8   I   I   4   I   I   I   -   -   -   -   -   -   -
5  |  4   7   7   7   6   7   3   6   4   4   4   5   5   4   6   4   4   4   3   6   6   5   6   4   3   6
6  |  4   4   7   7   3   9   5   7   6   6   6   8   5   5   7   0   7   5   6   7   7   3   5   7   6   6
7  |  3   5   4   4   5   9   4   5   5   4   4   7   7   6   7   7   4   5   3   6   6   4   5   4   7   7
8  |  5   5   7   7   6   6   3   7   6   5   5   6   5   5   7   6   6   4   5   6   6   4   6   0   5   6
9  |  0   6   7   7   3   8   4   5   9   4   4   6   5   5   7   5   7   4   5   6   6   5   6   4   5   6
10 |  3   7   4   4   3   6   5   7   7   5   5   7   6   5   8   5   6   4   6   6   6   5   6   0   4   7
```

```
Enter studentID : 50

Available courses for you are as below

EL101, HM101, IT101, IT102, MA101, PH101, CS101, CS102, CS103,
EL102, EL102, IT103, MA102, CS201, CS202, HM201, IT202, IT205,
MA201, CS205, CS206, HM202, HM203, IT201, IT203, IT204,


Enter course: EL101

Below is the most similar student based on users based collaborative filtering:
26

Predicted Grade for studentID 50 in subject EL101 : 39.082
```

9.Conclusion

In conclusion, course recommendation and grade prediction are performed by using user based and item based collaborative filtering to suggest courses to students. According to experimental results user-based filtering gives better results than item-based filtering. It is also observed that collaborative filtering is best to predict the grades of courses. In future few other methodologies can be combine with collaborative to improve the accuracy of results.
Below is the link of GitHub repository of project which includes code and documents.

GitHub Repository: https://github.com/PremalSamale/Course-Recommendation-System

10. References:

[1] Al-Badarenah, A., & Alsakran, J. (2016). An automated recommender system for course selection. *International Journal of Advanced Computer Science and Applications*, *7*(3), 1166-1175.

[2] Bozyiğit, A., Bozyiğit, F., Kilinç, D., & Nasiboğlu, E. (2018, September). Collaborative Filtering based Course Recommender using OWA operators. In *2018 International Symposium on Computers in Education (SIIE)* (pp. 1-5). IEEE.

[3] Ma, H., Wang, X., Hou, J., & Lu, Y. (2017, August). Course recommendation based on semantic similarity analysis. In *2017 3rd IEEE International Conference on Control Science and Systems Engineering (ICCSSE)* (pp. 638-641). IEEE.

[4] Huang, C. Y., Chen, R. C., & Chen, L. S. (2013, July). Course-recommendation system based on ontology. In *2013 International Conference on Machine Learning and Cybernetics* (Vol. 3, pp. 1168-1173). IEEE.

[5] O'Mahony, M. P., & Smyth, B. (2007). A recommender system for on-line course enrolment: an initial study.

[6]https://www.researchgate.net/post/Can_someone_differentiate_between_Cosine_Adjusted_cosine_and_Pearson_correlation_similarity_measuring_techniques

[7] https://github.com/Elective-Course-Recommender-System/Course-Recommender-System-

[8] https://towardsdatascience.com/how-to-build-a-simple-recommender-system-in-python-375093c3fb7d

[9] https://brenocon.com/blog/2012/03/cosine-similarity-pearson-correlation-and-ols-coefficients/

[10] Textbook of "Recommender System" by Charu C. Aggarwal.
==========END==========