# Personalized College Recommender: A System for Graduate Students Based on Different Input Parameters Using Hybrid Model

Ashutosh C. Harkare, Nishita Pali, Nikita Khivasara, Ishita Jain, Ravi Murumkar
Department of Information Technology
Society for Computer Technology and Research's Pune Institute of Computer Technology Affiliated to Savitribai Phule Pune University (formerly known as Pune University) Survey no 27, Dhankawadi, Pune - 411043 , Maharashtra, India

**Abstract:- In recent times, it is seen that many graduate students are willing to learn in foreign universities. Many factors drive students and experienced people to apply for different colleges and universities such as better opportunities of research, post-graduation, PhD and wider exposure to grab work in plethora of jobs. This situation is predominant in students from Indian sub-continent and Asian countries. These students aim to get admissions in many top universities of USA. As per data, scores of exams like GRE, TOEFL, IELTS, recommendation letters like SOPs and LORs, GPA of UG play pivotal role in university admissions. We are aiming to build a recommendation web platform which will suggest users with top 3 recommended colleges based on their profiles and inputs. As students spend a lot on counseling for university recommendations, our system holds a complete cost affordable platform for accurate results and user preferences. Collaborative filtering and content-based filtering is used to form a hybrid model on various hidden attributes. In this paper we summarized the methodology of underlined algorithms and focused on different parameters which will affect the overall recommendations.**

***Keywords:-*** *Model based Collaborative Filtering, Content based Filtering, Pearson's Coefficient, Neural Network, Matrix Factorization, K-NN, Recommender Systems, TF_IDF Vectorization, Prediction, Preference.*

## I. INTRODUCTION

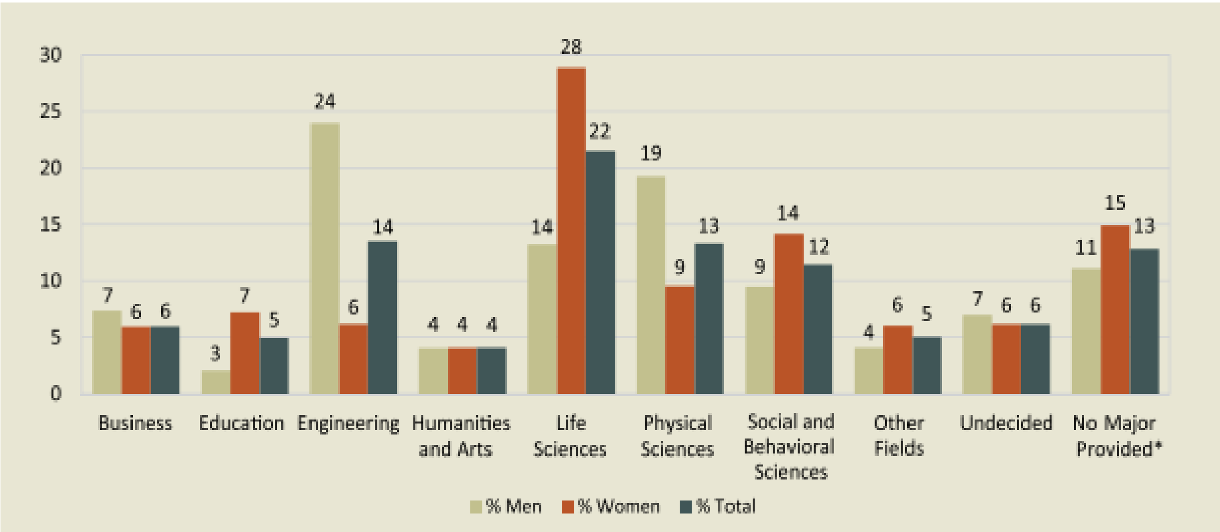


**Figure 1: Percentages of GRE General Test Examinees, by Intended Graduate Major Field and Gender**

According to a research analysis done by the official ETS, there are approximately 3.2L and 1L GRE test-takers in one-year time frame from USA and India respectively. Statistics show that 44% of the total test-takers are under the age of 23 i.e. they plan for further studies after their graduation, whereas the remaining 56% go abroad after getting some work experience. The survey also segregated the test-takers based on their age, gender, desired major - as shown in the graphs.

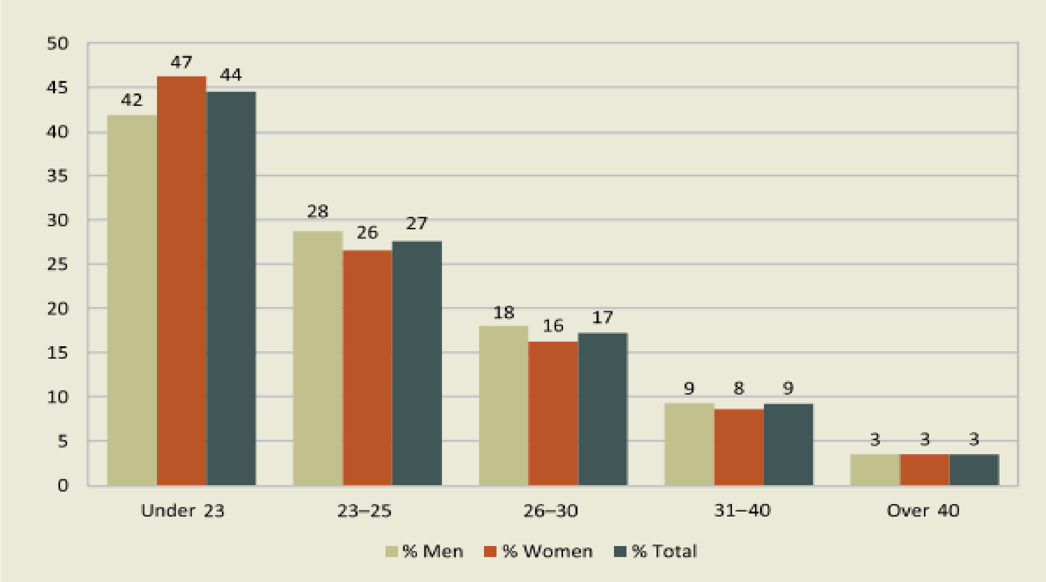


**Figure 2: Percentages of GRE Test Examinees, by Age Group and Gender**

Owing to this data, it is evident that the large numbers of students take GRE for their future studies. With the surge in the number of students pursuing higher studies nowadays, and many universities available, the students often find themselves confused with the plethora of options to choose from.

Factors like the GRE score, TOEFL score, SOP, LOR, education fees, cost of living, university ranking, etc need to be considered. There are many firms available these days that offer a counselling plan to the students and help in choosing the ideal universities. However, these counselling firms cost a lot money that would have otherwise been saved by the students and rather invest that amount in their university applications. Consequently, our model comes into picture and helps the user to decide the top-n universities based on the user's profile whether it is LinkedIn or other input forms.

In the dataset we will have user's GRE score, TOEFL/IELTS score, number of projects, internships, job experience, cost of living, major, earnings, etc. We will get user's data from their LinkedIn profile, resumes or through the input form at the home page of our web portal. To predict the list of universities best for the user, Hybrid Filtering which is the combination of collaborative filtering and content-based filtering will be used. The user sees top-n universities on screen as output after the working of recommendation engine.

## II. APPROACHES

1. ***Content based filtering*** – In this approach of recommendation systems, the selected user preference or features play an important role. The recommendation is based on the similarity of features selected by related or matched users of the system. For example, if two users of the same interest are using the book finder and one user selects the science fiction books as its genre, then other user might also get science fiction as the recommendation.
2. ***Collaborative filtering*** –In this approach similarity between the users or the items is taken into consideration. For example, if two users of the book review system have rated the two books with similar rating, then the non-similar book item recommendation of one user might be based on the ratings or interest shown by another user.

Table below shows the subtypes of above approaches and the algorithms which can be used to implement.

| Recommendation Approaches | | Algorithms | Applied Techniques |
|---|---|---|---|
| Content-based RS | Heuristic-based CBF | - *K* Nearest Neighbor<br>- Clustering | |
| | Model-based CBF | - Bayesian<br>- Clustering<br>- Neural networks | |
| Collaborative Filtering RS | Memory-based CF | - User-Based<br>- Item-Based<br>- RA<br>- UOS<br>- MLCF<br>- ULPE<br>- IGPE | - *K*-Nearest Neighbor<br>- Graph Theory<br>- Decision tree<br>- Web mining<br>- Support vector machine<br>- Bayesian models<br>- Clustering<br>- Association mining<br>-Principal Component Analysis (PCA)<br>-Bayesian networks<br>- Clustering<br>- Latent semantic analysis<br>- Neural networks<br>- Linear regression<br>-Probabilistic models<br>-Maximum entropy |
| | Model-based CF | - Slope 1<br>- Weighted Slope 1<br>- RSVD<br>- NMF<br>- PMF<br>- BPMF<br>- NPMF<br>- RSNMF<br>- ANLF<br>- INLF<br>- ESLF | |
| Hybrid RS | | - Feature combination<br>- Recommendation results combination<br>Other | - Bayesian<br>- Clustering<br>- Linear combination<br>- Probabilistic models<br>- Maximum entropy |

**Table 1: Algorithms for the recommendation approaches**

## III. LITERATURE SURVEY

[1] Explains different algorithms and approaches used for recommender systems. It has classified approached into three types – Content-based filtering, Collaborative and hybrid model. It also provides summary of different algorithms to be used.

[2] Has a course recommender system for students in different colleges. It uses a novel user-based collaborative filtering algorithm implemented. Results of this algorithm are compared with different existing approaches.

[3] Provides Bayesian model which is good in accuracy and the results of this model can also be explained effectively. Compares results with different existing systems like matrix factorization.

[4] Has implemented model of educational services recommendations. It explains about the approach of, Multi-Layer-Perceptron (MLP) which takes the N-dimensional and non-linear features to implement the algorithm.

[5] Explains TF_IDF Vectorization technique for parsing user LinkedIn profiles and through their resumes. It also provides the detailed explanation about university recommendation system. It takes the data from user profiles and recommends to 10 colleges. It uses cosine similarity for recommendation engine.

[6] Explains in detail about cosine similarity and Pearson's coefficient for collaborative filtering approach in recommendation engine. It uses these algorithms to predict graduate schools for students. Implementation is also done on USA graduate schools' dataset with accuracy results provided.

[7] Has detailed explanation about Multi criteria Collaborative Filtering. It also proves it by taking a survey on one dataset. For hard, intricate and huge datasets, Multi criteria Collaborative Filtering (MC-CF) gives better accuracy as well as performance and top-quality recommendations for users considering all varied features of items and users. CF algorithms often need continuous updating because of a constant increase in load of information, ways of retrieve that information, scalability and sparseness in rating matrix. Dimensionality Reduction techniques like: Matrix Factorization and Tensor Factorization techniques have been evaluated.

[8] Consists of comparison with existing models like cosine similarity, Pearson's coefficient, Jaccard, mean squared difference for product recommendation using collaborative filtering. A novel approach of triangle similarity is discussed and elaborated with the help of six commonly used datasets.

[9] Has explained opinions about the existing algorithms kike collaborative filtering and mass diffusion. The authors provided a novel Cov-covariance recommendation method based on correlation coefficients. This approach also provides precision and accuracy of results. This method first expresses the positive and negative correlation among random samples without knowing the distribution of items in the dataset. Then it sorts and segregates popular items in recommendation list. Usability of these approaches was also shown by implementing those in movie recommendations.

[10] This paper shows innovative approach and lists popular algorithms to be used in recommendation systems. It also uses ANN with complex neural networks for improving the precision of results on various attributes of datasets. It uses student course and stream selection-based recommendation engine to reduce the dropout rate. It showed the use of random forest, k-means, multilevel perceptron, support vector machines to evaluate the dataset and including every attribute in the prediction. Encoding technique was also used to combine two to three attributes and form a score for all three combined.

## IV. PROPOSED METHODOLOGY

Our implementation aims to build a web application which will recommend top 3 universities based on user profile and preferences. In this process, we are aiming to use hybrid recommendation algorithm which will consist of collaborative filtering followed by content-based filtering. We have shown the detailed flow through the diagram.

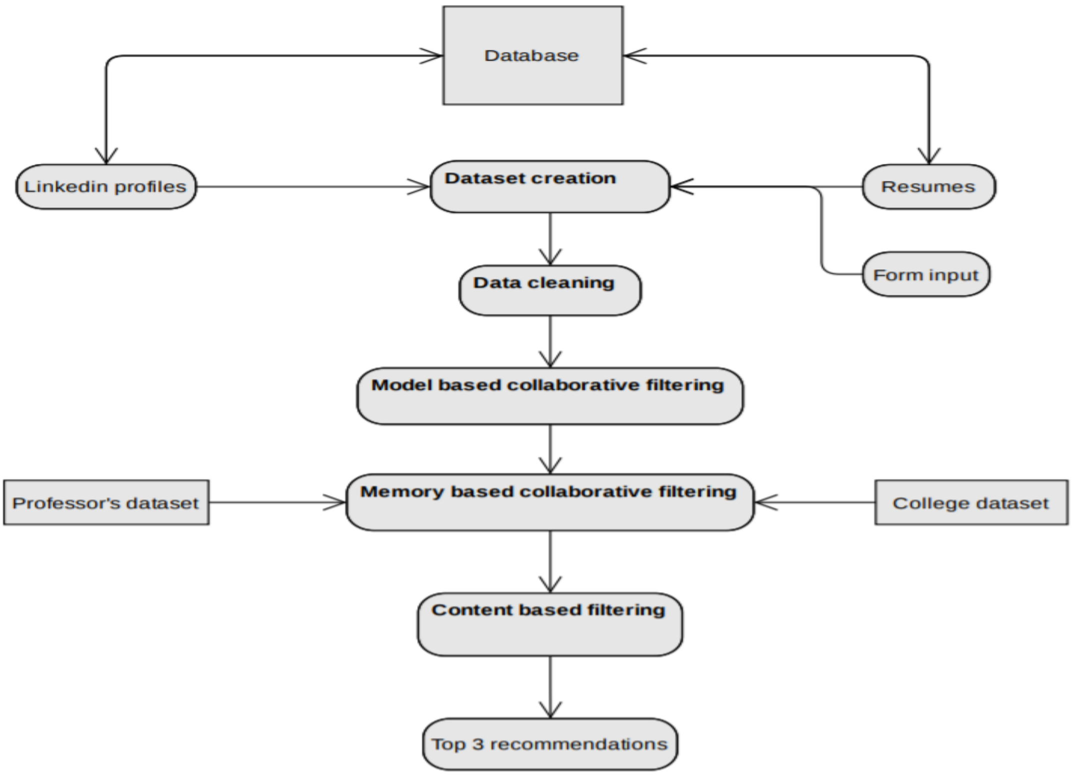


**Figure 3: Flowchart of the hybrid recommendation system**

Database is connected to the LinkedIn profiles and the resumes submitted by the students. Different algorithms can be used to parse the data and extract relevant information from the profiles. Database will fetch and store the data from these profiles. The database can also provide data to the data cleaning phase. Users can also provide the information through a form which will be provided after the successful login of student to our web portal.

The data from the user is then transmitted to the data cleaning phase. The output of this phase will prepare the data for model training. According to data parsed in phase 1, model based collaborative filtering algorithm will work to reduce the data weight.

Considering different parameters of the user data, the memory based collaborative filtering algorithm will match this data to the data of professors and university data. It will narrow down the recommendation options. This is then provided as an input to the content-based filtering algorithm.

According to the data fields in all the datasets, the content-based filtering algorithm will match the best option according to the items in the datasets. The output of this phase will give the user top 3 recommendations of their preference and qualifications.

## V. CONCLUSION

In this survey paper, we introduced a vast variety of implementations for recommender systems using different algorithms of collaborative filtering and content-based filtering. Different algorithm subtypes of collaborative filtering are also discussed. We also provided the methodology for our implementation. Various data sources for our recommendation engine are also discussed. It includes LinkedIn profiles, resumes of students, input forms in our website. Dataset of universities and professors are also taken from varied sources. We tried to summarize the process of recommendation engine for top 3 recommendations with different algorithms in this survey paper.

## REFERENCES

[1]. MahdiJalili,,SajadAhmadian, MalihehIzad (2018) "Evaluating Collaborative Filtering Recommender Algorithms": A Survey. In: IEEE Access 2018 date of publication November 28, 2018. Volume 6, 2018.

[2]. Ling Huang, Chang-Dong Wand, Hong-Yang Chao, Jian-Huang Lai. "A Score Prediction Approach for Optional Course Recommendation via Cross-User-Domain Collaborative Filtering". In. IEEE Access 2019 date of publication February 7, 2019. Volume 7, 2019.

[3]. Priscila Valdiviezo-Diaz, Fernando Ortega, Eduardo Cobos, and Raúl Lara-cabrera "A Collaborative Filtering Approach Based on Naïve Bayes Classifier". In. IEEE Access date of publication August 5, 2019. Volume 7, 2019.

[4]. Farhan Ullah, Bofeng Zhang, Rehan Ullah Khan, Tae-Sun Chung, Muhammad Attique, Khalil Khan, Salim El Khediri, And Sadeeq J "Deep Edu: A Deep Neural Collaborative Filtering for Educational Services Recommendation". In. IEEE Access 2020 date of publication June 15, 2020. Volume 8, 2020.

[5]. Vidish Sharma, TarunTrehan, Rahul Chanana, Suma Dawn "StudieMe: College Recommendation System". In 2019 3rd International Conference on Recent Developments in Control, Automation & Power Engineering (RDCAPE). 978-1-7281-2068-3/19/$31.00 ©2019 IEEE.

[6]. Madhav Iyengar, Ayanava Sarkar, Shikhar Singh "A Collaborative Filtering based Model for Recommending Graduate Schools". In IEEE Conference 2017. 978-1-5090-5454-1/17/$31.00 ©2017 IEEE.

[7]. Dheeraj kumarBokde, Dheeraj kumarBokde, Debajyoti Mukhopadhyay "An Approach to A University Recommendation by Multi-Criteria Collaborative Filtering and Dimensionality Reduction Techniques". In 2015 IEEE International Symposium on Nanoelectronic and Information Systems. 978-1-4673-9692-9/15 $31.00 © 2015 IEEE.

[8]. Arta Iftikhar1, Mustansar Ali Ghazanfar1, MubbashirAyub, Zahid Mehmood And Muazzam Maqsood "An Improved Product Recommendation Method for Collaborative Filtering". In IEEE Access 2020 date of publication June 30, 2020. Volume 8, 2020.

[9]. Biao Cai, Yusheng Huang "Personalised recommendation algorithm based on covariance". In. The 3rd Asian Conference on Artificial Intelligence Technology (ACAIT 2019). J. Eng., 2020, Vol. 2020 Iss. 13, pp. 577-583.

[10]. Antonio Jesús Fernández-García, Roberto Rodríguez-Echeverría, Juan Carlos Preciado, José María Conejero Manzano, And Fernando Sánchez-Figueroa "Creating a Recommender System to Support Higher Education Students in the Subject Enrollment Decision". In. IEEE Access 2020 dateofpublicationOctober16,2020. Volume 8, 2020.

[11]. NacimYanes, Ayman Mohamed Mostafa, (Member, IEEE), Mohamed Ezz,, And Saleh NaifAlmuayqil "A Machine Learning-Based Recommender System for Improving Students Learning Experiences". In IEEE Access 2020 date of publication November 5, 2020. Volume 8, 2020.