UniAssist- Implementation of Machine Learning Based Higher Education University Recommendation System

[1]Ruchi Sharma, Sarvesh Joshi, Abhishek Agrawal, Vaibhav Kulkarni

[1]Assistant Professor, Department of Information Technology, MPSTME , NMIMS University, Mumbai, India

[2]Department of Information Technology, MPSTME ,NMIMS University, Mumbai, India

abhishekragrawal@gmail.com, sarvesh.manu8@gmail.com

Abstract - **UniAssist project is implemented to help students who have completed their Bachelorette degree and are looking forward to study abroad to pursue their higher education such as Masters. Machine Learning would help identify appropriate Universities for such students and suggest them accordingly. UniAssist would help such individuals by recommending those Universities according to their preference of course, country and considering their grades, work experience and qualifications. There is a need for students hoping to pursue higher education outside India to get to know about proper universities. Data collected is then converted into relevant information that is currently not easily available such as courses offered by their dream universities, the avg. tuition fee and even the avg. expense of living near the chosen university on single mobile app based software platform. This is the first phase of the admission process for every student. The machine-learning algorithm used is Collaborative filtering memory-based approach using KNN calculated using cosine similarity. A mobile-based software application is implemented in order to help and guide students for their higher education.**

Keywords — *Collaborative filtering, Machine Learning, Recommendation system, mobile application, Higher education*

I. INTRODUCTION

There is a need for students planning to pursue higher education outside India to get to know about proper universities. Students would get to know about useful knowledge not easily available such as courses offered by their dream universities, the avg. tuition fee and even the avg. expense of living near the chosen university.

The primary objective of this project is to help them identify appropriate countries to study as well as universities to be admitted into.

Following are the features in the software application implemented:

1. University Recommendation based on student profile and course preference.

2. Guide for applying to universities.

3. Providing students with appropriate information about universities.

4. Systematic guide for admission to a university.

5. Connecting students with appropriate alumni based on the alumni admitted courses and countries - Alumni Connect.

Students would have a software application containing all their student information, university information, recommended university information, alumni information and the shortlisted university information.

The data, utilized to train the datasets for the ML algorithm, is the actual data of the students who have pursued or are currently pursuing their higher studies.

II. BACKGROUND

Currently students who are eager to go abroad for further education generally approach counsellors for shortlisting universities and knowing the process of admission due to limited guidance. If the above process can be implemented through automated software application, it would help students get appropriate information and be time efficient at reduced cost. Current approach is requiring manual intervention. The purpose of this project is to automate the initial manual process and further aid in counselling sessions of students that would provide better results than

the existing approach. We propose to solve multiple problems in the existing system through our technology solution at various stages using various tools, converging into one integrated system.

The reason this particular topic is chosen is due to the amount of students going abroad for higher studies. The ratio of such students has increased over years. This study is depicted in Figure (1). This graph is being created with the help of student data collected from various universities and counsellors.
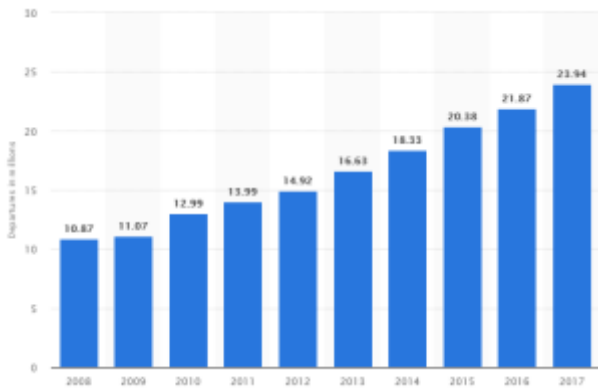

Figure 1. Analysis showing the no. of student going abroad to study (in millions)

The reason the four countries - USA, Canada, Australia and New Zealand are selected particularly is seen from the graph indicating the graph in Figure (2).

According to several studies, we researched that USA, Australia, UK, Canada and New Zealand are the major five countries where students mostly pursue higher education.

III. RELATED WORK

There has always been a trend and eagerness of students wanting to go abroad to study for their further education. However, it is a tough task for them due to lack of knowledge and misinformation. This application makes this task easier for students to do so. Yocket.com website [2] offers online support to students by providing them with appropriate information about the universities and the courses it offers, but it lacks the feature of match making by filtration of student grades and credentials as well as academic achievements and results. It displays the generic information regarding the tuition fees and avg. living costs. Our application provides customized filtering according to the preferred course or country of the student.

In collaborative filtering algorithm, the similarity in the data is computed. The data from the data sets can be stored either in a csv file, json file or a normal xls file as well. This data can be stored in percentage form making the calculation of similarity easier [9]. When calculating the similarity between two items it is better if we obtain a ratio of users who have a particular amount of score. This proposed method can increase the accuracy of the algorithm [5]. Recommendations are based on the data between the

numbers of users. It collects the data from the user, compares them and generates the output. For e.g. for a student the data can be collected by conducting a vote for the preference of the universities. [6]. The recommendation is calculated based on association rule mining. Using association rule mining can help discover relations between the variables where there is a large quantity of data involved.

Rather than recommend items because they are similar to items a user has liked in the past, we recommend items other similar users have liked [11].

The association rule mining generates rules based on frequently occurring patterns, associations that are observed in the data [7]. Rating matrix would contain the admit of the alumni's and the universities. The rows in the rating matrix are the universities and the columns are be the alumni's. Based on the admits the matrix can be constructed [4].

In [10], authors have authors have stated that contextual features that could account for situations where otherwise good suggestions could be rejected should be of interest.

However, this massive amount of data is available across several heterogeneous platform making it a challenging task especially to comprehend useful information and effectively. Machine learning and data mining techniques are often used to extract data from different sources [8].

UniAssist uses Collaborative based recommendation algorithm predicts unknown outcomes by taking inputs from the users and creates a matrix of choices or preferences, which in turn provides the exact results.

IV. STATISTICAL ANALYSIS

A. *Statistical Analysis for The Data Collected Based On Universities in Respective Countries*

This is the statistical analysis for the percentage of universities in every country across the entire data collected. This data includes universities from four countries namely USA, Canada, Australia and New Zealand.
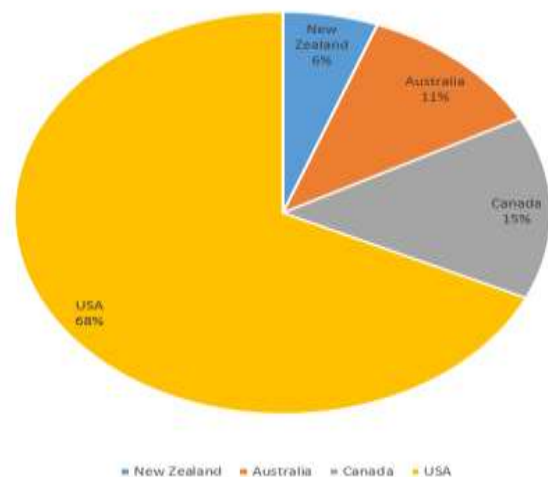

Figure 2. Statistics for Universities based on Countries.

We can see in Figure (2). from the actual data acquired from the students across universities, 68% of student's study in the USA, 15% of student's study in Canada, 11% of student's study in Australia and 6% of student's study in New Zealand. This gives us a general idea of the countries students study in and the students going in the future may wish to study in.

B. Statistical Analysis for The Data Collected Based On Courses pursued

We have researched and explored that there are generally seven major courses in the computer field where students look to pursue their Masters in which include Computer Science (CS), Cyber Security, Data Science (DS), Machine Learning (ML), Software Engineering (SE), Management Information System (MIS) and Artificial Intelligence (AI).
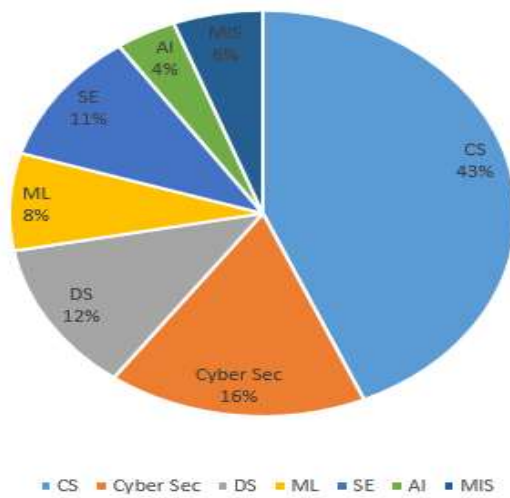

Figure 3. Statistics of Students based on Courses.

As we can see in the Figure (4) from the real time data collected, we came across the statistical fact that 43% of the students pursue MS in Computer Science (CS), 16% students pursue MS in Cyber Security, 12% students pursue MS in Data Science (DS), 11% students pursue MS in Software Engineering (SE), 8% students pursue MS in Machine Learning (ML), 6% students pursue MS in Management Information System (MIS) and 4% students pursue MS in Artificial Intelligence (AI).

V. MACHINE LEARNING MODEL

The Machine Learning algorithm used is Collaborative filtering memory-based approach using KNN calculated using cosine similarity. This algorithm predicts unknown outcomes by taking inputs from the users and creates a matrix of choices or preferences, which in turn provides the exact results. Data is collected from the students having academic details and other credentials of the students. Based on the details the algorithm makes a predictive analytics and display the universities that are applicable for them where they can apply to those specific universities that the algorithm has predicted. This system is being implemented based on the student's profile.

Three distance calculation algorithms were implemented for this project, namely - Cosine Similarity, Pearson Correlation and Euclidean Distance Calculation. But amongst these three algorithms Cosine

Similarity had the maximum accuracy in the recommendations. We have considered the following parameters for the recommendations is based on the data sets:

1. GRE score
2. IELTS/TOEFL score
3. Grad score
4. Course preference
5. Country preference

In Figure (4), we can see an example portraying the use of the Collaborative Filtering Algorithm. The following example is with respect to the application proposed. The example shows two people, one person say A in purple and the other man in blue. Considering A is an old alumnus of the application and has, her data enrolled in the database that is helpful for further recommendations. Both have similar grades and both have similar course requirements as well.

A has already received admit from a few universities or is already admitted to a university. The man say B who we consider as a new user has similar grades as the lady with similar course preferences as well.

So B is recommended universities based on such real time historical data of the mentioned lady and many more entries as hers. That is the reason it is simply said that 'A's Admits, B's Recommendations'.
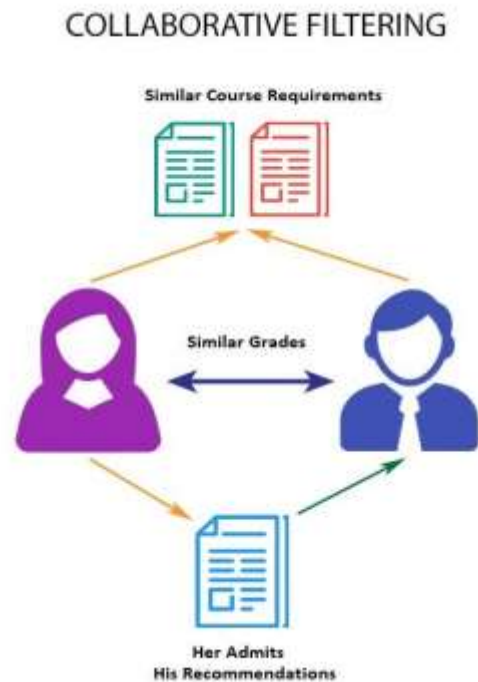

Figure 4. Use of Collaborative Algorithm in UniAssist

VI. PROPOSED METHODOLOGY

Before the coding phase begins, data is from the students that is necessary for our application. A survey form is created where the students were asked about their academic details for data collection. Based on the response, a historic dataset is created containing details about student admit in different universities. This historic data is important for the application as the datasets is trained for recommending the universities. Making of the historic data is the first priority.

Here the coding phase begins. First login and registration page are implemented for the students that are linked to the MySQL database in MS Azure server. After creating the login/registration page, the UI design is converted into actual code. After implementing the design, a machine-learning algorithm that is Collaborative-filtering memory based approach using KNN calculated using cosine similarity is used to predict and analyse the results. It is best suited for the application because it makes automatic prediction based on the historical datasets collected. Therefore, the application is recommending the universities based on the data given by the students that is historic data. Comparative analysis is done between the existing student data and the university data collected, this would help the algorithm predict and recommend universities more accurately.

After training the datasets using filtering approach, the database is integrated with the application. This database includes data about the universities, alumni and the students.

The technology used in this application is the use of software like Android Studio, MySQL database and Python for machine learning algorithm coding. The application and the data is hosted on the server using MS Azure cloud services.

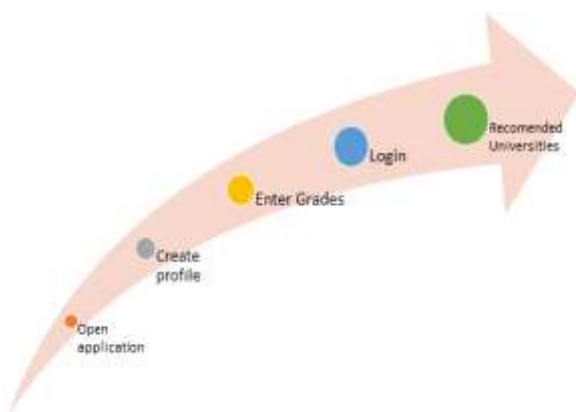The flow of the recommendation of universities is in the following manner portrayed in the Figure 5 below: -

Figure 5. Flow of the application for recommendation of universities

The flow of the application for recommendation of universities can be seen in Figure (5). It consists of five basic steps:

a) Open Application

b) Create Profile

c) Enter Grades while creating profile

d) Login

e) Recommended Universities Displayed

In step a), the user will firstly open the application and then opt for signup. The user must then enter their basic details and their scores, including their GRE, TOEFL/IELTS, CGPA scores. They must select their preferred Course and preferred Country where they wish to pursue their post-graduation. Based on the scores and preferences the user enters the recommendation is calculated. As the scores are unique for every user hence the recommendation that is calculated is unique for every user, therefore there are customized recommendation for each student user. Once the user has successfully registered themselves, they will need to login in the system with the credentials they set in the previous step. After logging in successfully the user is presented with the recommended universities, using ML based algorithm. The recommendations are in a horizontal slider format that arranges the universities according to their country rank. The user can then browse through the universities and select one according to their choice Client is redirected directly to the university page where they can find additional information about the university.

VII.   FUTURE WORK

The developer who wishes to expand this subject of the project can include more courses, more degrees (e.g. Certifications, MBA, under-grad programs, etc.), more countries and a live chat feature with the alumni as well. Currently at this stage, it only focuses on four countries (USA, Australia, and Canada, New Zealand), more countries can be targeted and the database can adapt to these new countries and the new students. It will help develop the data sets increasing the recommendation accuracy.

The application developed as a web application as well as on the iOS platform.

Updating student information in the database that acts as data input for the machine learning algorithm, updating university lists, updating university requirements data and even updating the list of alumni over time is a continuous process. This data must be updated every year with the change in requirements from the university end.

A data update cycle can be set for every quarter in a year as it acts as one semester in the tri-semester pattern followed in the majority of the countries for the majority of the courses.

VIII. CONCLUSION

There is high demand of students going abroad to pursue higher education, thus making it clear that there is a need for the application as it would help such students get appropriate guidance and recommendation from universities based on their grades. Students get an idea based on the recommendation that they receive and get a clear idea about getting admission in the respective universities. Connection with alumni would safeguard their interest and clear the doubts in the mind. The use of a ML algorithm for memory based recommendation approach using KNN calculated using cosine similarity increases the accuracy of these recommendations.

REFERENCES

[1] Abhaya Kumar Sahoo, Chittaranjan Pradhan, Rabindra Kumar Barik ,” Machine Learning Based Health Recommender System Using Collaborative Filtering”,Computation MDPI, 2019.

[2] Yocket.in. (2019). Yocket | MS & Bachelors in US, Canada, Germany & more. [online] Available at: https://yocket.in/ [Accessed 10 Aug. 2019].

[3] Makkar, S., “ Cutting Short Post-Study Period: No US Dream For Indian Youth In Trump Era?” [online] Business-standard.com., 2020, Available at: <https://www.business-standard.com/article/economy-policy/cutting-short-post-study-period-no-us-dream-for-indian-youth-in-trump-era-117022200059_1.html> [Accessed 12 March 2020].

[4] D. K. Bokde, S. Girase and D. Mukhopadhyay, "An Approach to a University Recommendation by Multi-criteria Collaborative Filtering and Dimensionality Reduction Techniques," IEEE International Symposium on Nanoelectronic and Information Systems, Indore, 2015, pp. 231-236.

[5] Jinbo Zhang, Zhiqing Lin, Bo Xiao and Chuang Zhang, "An optimized item-based collaborative filtering recommendation algorithm," IEEE International Conference on Network Infrastructure and Digital Content, Beijing, 2009, pp. 414-418.

[6] A. S. Tewari, A. Kumar and A. G. Barman, "Book recommendation system based on combine features of content based filtering, collaborative filtering and association rule mining," IEEE International Advance Computing Conference (IACC), Gurgaon, 2014, pp. 500-503.

[7] Khanam Z., Alkhaldi S. ,”An Intelligent Recommendation Engine for Selecting the University for Graduate Courses in KSA: SARS Student Admission Recommender System “, In: Smys S., Bestak R., Rocha Á. (eds) Inventive Computation Technologies, ICICIT 2019. Lecture Notes in Networks and Systems, vol 98. Springer, Cham

[8] Ruchi Sharma, Pravin Srinath. ,”Business Intelligence using Machine Learning and Data mining techniques-An Analysis”, IEEE conference ICECA, 2018.

[9] Mohd, Abdul & Hameed, Mohd & al jadaan, Omar & Sirandas, Ramachandram. , “Collaborative Filtering Based Recommendation System: A survey”, International Journal on Computer Science and Engineering. 4, 2012.

[10] Eoin Hurrell, Alan F. Smeaton, , “Context Ontologies for Recommending from the Social Web”, CaRR’13, February 5, 2013, Rome, Italy., ACM 978-1-4503-1847,2013

[11] Balabanovic M., Shoham Y. ,” Fab: content-based, collaborative recommendation”, Commun. ACM 40(3): 66–72,1997