IJMLNCE JOURNAL

International Journal of Machine Learning and Networked Collaborative Engineering

Journal Homepage: http://www.mlnce.net/home/index.html

DOI : https://doi.org/10.30991/IJMLNCE.2019v03i03.002

Recommender System for Educational & Corporate Sector In Prediction of Domain Recommendations & Analysis using Machine Learning

Praneet Amul Akash Cherukuri

Dept. of Computer Science & Engineering, CMR Institute of Technology, Hyderabad, India

chamul91@gmail.com, https://orcid.org/0000-0003-1408-6784

Abstract

These days Career and Domain options have always been a very big ambiguous decision-making process for many prospective aspirants. Many aspirants make substantial domain changes very late in their career which may result in drastic effects on their career as well as their financial status. Many reports suggested that companies have suffered huge losses because of making wrong choices regarding the domain and employee interest. Hence providing a common platform early in the education sector for both the aspirants as well as companies that would provide appropriate domain suggestions for aspirants as well as right employee choices for companies would be highly beneficial that could help in generating better results when compared to the traditional ways of career choices and employment. In this research, we are proposing a recommender system based model that would bridge the gap and help in formulating their future needs.

Keywords

Machine Learning,

Recommender Systems,

Big Data,

Bayesian Classifiers

1. Introduction

Over the last few years, many technologies have taken over the IT Industry making several changes that are permanently renounced and are still having a huge impact in the present day scenario. In this research paper, we are proposing a model that uses some of these very important and highly applied technologies such as Machine Learning, Recommender Systems, Big Data, Bayesian Classifiers', Regression models and Natural Language processing. Machine Learning is a branch of Artificial intelligence, it is a method of Data analysis in which it builds a model that performs huge amounts' of data analytics and learns from that data by itself, identifies patterns and makes decisions on its own with the minimal human intervention [1]. Recommender Systems is an application of machine learning in which the model learns from the given data and hence proposes a new recommendation. The recommendation system gives the option to comprehend an individual's taste and find the new, appropriate choice for them consequently. However individuals preferences shift, they do-follow patterns. Individuals will in-general like things which are comparable social individual preferences. Most of the times these sorts of examples can be identified with the importance of things[2].

Big Data refers to the advancement in innovation clubbed with an assortment of data, strategies that

have now furnished the instrument to manage a wide range of issues that show up during the procedure of data assortment and during working with enormous volume, assortment, and speed of data. The main purpose behind Big Data is the application of information tools to pave the way for data analysis and retrieve appropriate information for better estimation, planning, and judgment in any business process [3]. A thorough machine learning analysis of the student's sentiments does wonders in the student enrolment process. Big Data solutions, therefore, serve as an insight analysis podium [4].

In this paper, we use collaborative filtering because of the main drawbacks that the proposed model consists which are, most users do not provide evaluation and hence user matrix can result in great sparsity. Another such issue is for a given metric there is no previous data then that particular metric could result in not being recommended to any user hence by using collaborative and content-based filtering we can oversee these issues. Bayesian classification will be utilized to arrange items to bring down the components of the user-produced matrix. Likewise, by executing a hereditary calculation on the ordered items the clients can be placed into groups. At that point, by input on significance, it is conceivable to make a profile for new clients, consequently taking care of the first rater issue [5]. Using these technologies and algorithm we build the overall model that can input and output data. On the other hand, a user-interactive platform built using Web Design & Development technologies is very important for the user experience and user interaction that could help in the improvement of data generated daily.

1.1. Prevalent System and Need for Change

In the present day scenario, many educational institutions have their learning management systems that are a majority of temporary or very privately held dealing only concerning a particular educational institution. The disadvantage of these LMS is that they are very inefficient in means of data collection and the scope of doing it. Also, another huge disadvantage is all the data that is being generated concerning that particular institution is not stored well or is made private and hence there is no possibility of doing data analysis in this case that creates a huge impact in the adaption process for the collection and assembling of data. Hence these prevalent systems are mostly data collection systems but not most of them have been in the process of analyzing and proposing a better analysis to the users or the aspirants, rather they have been just made to collect data and provide temporary solutions.

The need for a change of these systems is although all the sectors have been implementing and experiencing a huge change in their activities the education sector has fallen far behind following the same old traditional methodologies. Hence the proposed model could help the education sector in identifying a different perspective of education to help students, users, organizations to benefit their careers in an optimized direction

2. Proposed Model

The proposed model is a machine learning model that generates recommendations' using recommender systems concept of machine learning by analyzing the huge amounts of data assembled using big data technologies. The model would be divided into three major phases of development. The three models clearly define the process of data collection, analysis, classification & recommendation. The model analyses the metrics in a very curated way and could suggest the user or an organization with a recommendation that could be early advice and could help in making better decisions at a given point of time. Also, it would result in making the required changes to achieve the goal even before starting the process so that there are no errors in the early phases of the process.

The model proposed creates a bridge between educational institutions and organizations where data particular to a similar user and their proposed metric are evaluated and established. Since the data is ever recurring hence the probability of using big data to analyze and hold that data is mandatory. The data that was already collected by the present-day learning management systems and other useful softwares both from educational institutions and other companies are evaluated, cleaned and then set the stage for the upcoming

platform to generate new data and then combine it. The model starts by accumulating all the data that it has been collecting using various learning management softwares and then are stored using the big data methodologies so that at a later point of time they can be retrieved for analysis. Then the data is divided concerning two databases are User data and the company-specific data which is evaluated and then the states are formulated. Int eh next phase the data which is collected is undergone content and collaborative filtering in which the data is compared to many similar metrics and hence is classified according to its similarities. The data is then prone to the regression models where it is filtered and then is sent to the database with filtered metrics. In the next phase when we retrieve the information that we analyzed in the previous phase and stored in the database is now displayed back according to the need of the user or the company.

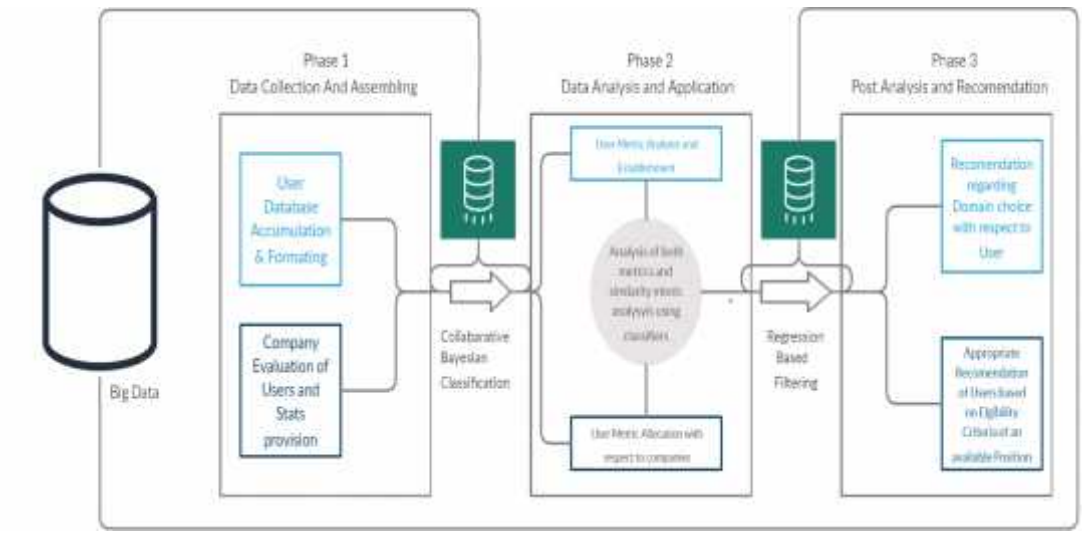

Figure 1. Representation of The Proposed Model

2.1. Phase 1: Data Collection and Assembling

This phase includes the pre-phase of analysis, where the data that is required for the analysis is recorded using the platform which is provided by the model that is being implemented in the everyday scenario of the user/student's academics from the very beginning of his/her academic year recording all the required metrics which are stated in Figure 2., which can be used to evaluate the overall metrics. The data being generated will be stored using big data technologies such as NoSQL, Hadoop, and Apache Spark. In the phase of data collection, the data collected is usually concerning the preferred metric and it is received from the platform which is then evaluated to yes or no from the data. Outliers for some metrics are removed such for example A user does not have any concerning specific data related to a metric then that data will become an outlier and would not be just removed because that would become a problem in getting that certain metric being analyzed in a similar user profile. Hence even outliers in this are properly specified and cleaned.

One of the areas that volume, variety, and velocity exist together in the information is higher education. A lot of educational information is caught and created regularly from various sources and in various configurations in the advanced education system. The educational information change from those delivered from students' utilization and communication with learning the board frameworks (LMSs) and stages to learning activities and courses information [6].

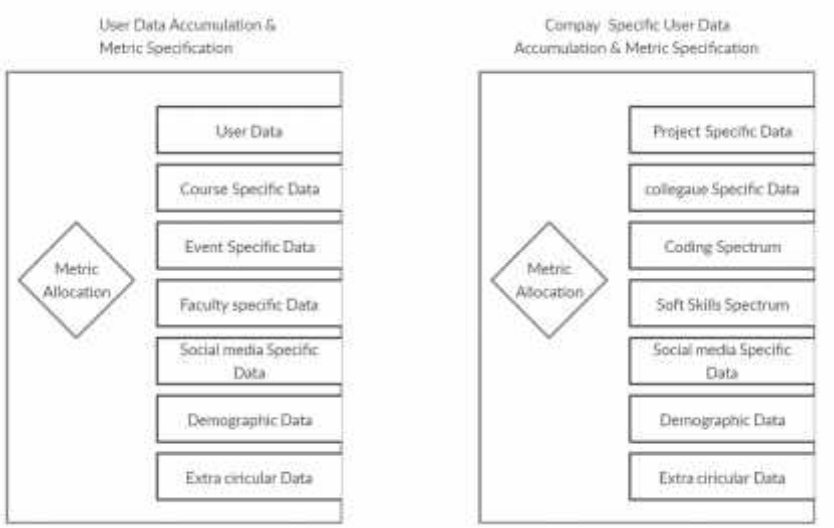

Figure 2. Representation of Phase 1

The constrained exploitation of large educational data and the size and sort of these data inside the setting of advanced education connotes the requirement for unique procedures to be applied to find new advantageous knowledge that at present is hidden up inside data [7]. Since the data generated is huge and recurring data, big data is used to make the process smoother rather than using a traditional SQL database. The data from companies resulting in the evaluation of different user's metrics and evaluation in the particular domain are also generated and records for it to be analyzed in the further phases of the model where the analysis takes place.

2.2. Phase 2: Data Analysis and Application

The second phase of the model deals with the main analysis and the application part. Here the analysis part refers to the metric allocation to all the data that has been generated and then cleaned form the first phase. Metrics then allocated to the data are then sent to the application phase. The application phase is the phase where the different metrics shall be processed into a Collaborative and Content Bayesian classification in which they are classified using the item based classification: that is the probability of similarity taking place and then the features are classified against the 'yes' or 'no' metrics. When we take into account the item based classifier that takes the inputs as 'yes' or 'no' then it will be as follows [8].

The analysis is taken place as follows when a user X with metrics A, C, G, H is given to the database then the data is undergone through the Classification in which the users' metrics are then matched with a yes or no feature under all the various profile in the database and then a compilation of all the similar user profiles is made. Then this compilation is then compared to extract the most similar and redundant career choice metric that has been made between these similar profile sand it analyses them to find a pattern of success rates in them. After finding the most successful career choice using the metrics then the model would run on a database of the company-specific evaluation for that career choice concerning the profile and the user metrics and, hence a proper evaluation that is both concerning the previous user data as well as the feedback from companies for users with such metrics, career success rates have been evaluated for the User X and are stored on to the database. On the other hand, the same evaluation from these users is also evaluated to know the eligibility of the certain user for a job role that has been posted by the company by using the evaluated metrics and can, therefore, understand which user would be highly preferable for a certain job role with concerning metrics.

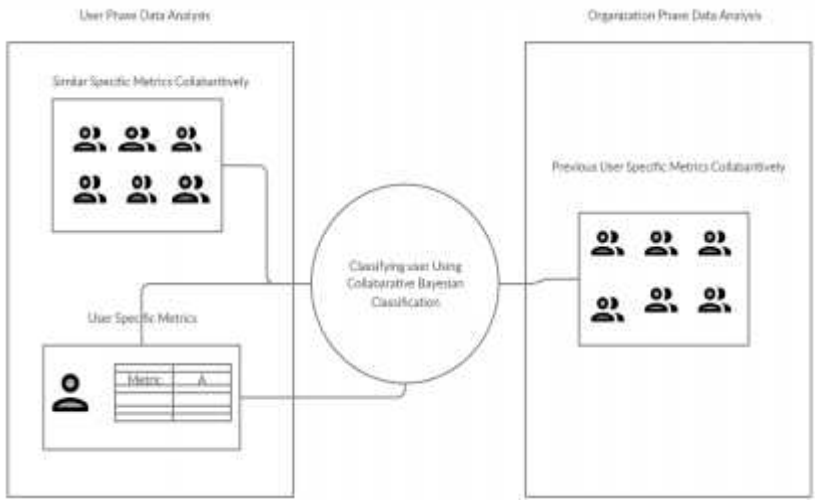

Figure 3. Representation of Phase 2

$$\frac{P_{feature}(yes\,|user\_metric)}{P_{feature}(no\,|user\_metric)}$$

A 'yes' metric can be defined as :

$$P_{feature}(yes\,|user\_metric) = \frac{P_{feature}(user\_metric|yes) \times P_{feature}(yes)}{P_{feature}(user\_metric)}$$

A 'no' metric can be defined as :

$$P_{feature}(no\,|user\_metric) = \frac{P_{feature}(user\_metric|no) \times P_{feature}(no)}{P_{feature}(user\_metric)}$$

When combined to take the classification:

$$\frac{P_{feature}(yes|user\_metric)}{P_{feature}(no|user\_metric)} = \frac{P_{feature}(user\_metric|yes) \times P_{feature}(yes)}{P_{feature}(user\_metric|no) \times P_{feature}(no)}$$

The classified data is then directed to the regression algorithm where they are filtered and hence the recommendation is made based on which the data and the metrics are defined. To recognize comparative users, we should gauge the closeness between two users. There are two most mainstream strategies: cosine-based closeness and correlation-based similarity[9]. Most recommender systems prefer the correlation similarity since it outperforms the cosine-based similarity in most problems [10]. Then these metrics that have been filtered and reset to the particular user features are again stored in the database for the future use of users as well as organizations. Also, this data can be retrieved when they are being used for the comparison studies of other similar user models.

2.3. Phase 3: Post Analysis & Recommendation

This phase is the last step and the post-analysis phase that deals with the data retrieval from the database in which classification and filtration of the featured metrics are done and displayed onto the portal. This phase displays the appropriate recommendation of a particular domain or career choice listing out all the possibilities of success rates in percentile for that particular field of domain and comparing it with the profile of the user concerning a similar database. The recommendations made by the model are concerning the data that has been already recorded and are real-time examples that are analyzed to find a pattern.

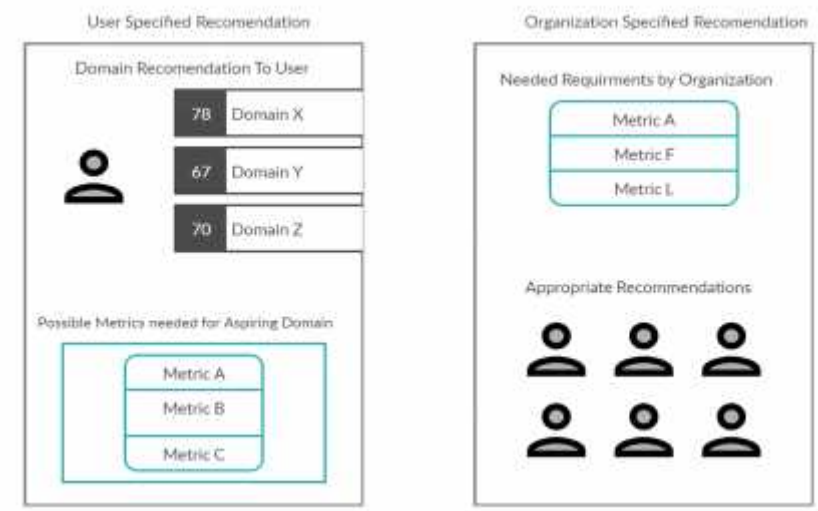

Figure 7. Representation of Phase 3

Also, it intakes the aspired domain choice of the user and provides the outputs and their possibilities of success in that particular domain along with the metrics he/she must improve to be able to sustain in their aspired domain choices. On the other hand, companies get a properly sorted list of the best recommendation for a particular role that had been requested and the proposed success rates of him/her being employed for that particular career choice. Also, the company could list down the particular success rates for a person of their choice. These recommendations are retrieved to the user after undergoing the process of filtration using the regression techniques and hence more accurate results and recommendations are generated to the user that could increase the success rate for him in a particular carer or domain choice. The results generated are accurate and could be improved on the training done to the model that could help the model to generate many accurate recommendations. The next phases of the model are outlined in the future scope of the paper that would help in the improvement of the model from the data generated.

3. Future Scope and Prospective Advantages

The proposed model would aim to help users build a better career by recommending them the best career options that are appropriate to their profile evaluations a well as everyday behavioral data that has been generated by the platform. The model would also help him/her by giving the user an appropriate success rate for the recommended domain and career options by defining their success rates while classifying them using the real-time company profile evaluation data that could help the user in understanding his/her success rates in a defined domain. Another noted advantage would be to organizations where the model could be used to get a recommendation for their recruitment as the proposed model hosts a million bytes of data related to thousands of users along with their evaluated metrics and could help the organization to evaluate the best candidate for a proposed role.

Since this is a real-time model the data is very large and hence causes a lot of disturbances. The future scope of this model has no limits since the collection of data here is only limited to educational institutions of higher level, if the same model is being implemented from the very early stage of the education phases, the data and analysis could become more refined due to the increase in the quality and the quantity of the data and hence the model can become more accurate as far as possible. Also, the model can be used in the overall analysis of job aspirants by linking his social media profiles and then understanding his/her behavioral data could help the model in making more appropriate choices. Also, other classification and regression algorithms could be used that could help in the prostatic increase in the accuracy of recommendations. The metrics that are being observed in the above model are very outlined and can sometimes fail to understand the deep patterns hence the collection of some intuitive metrics that actually defer the user behavior and also his ability to perform a task could also be recorded to provide much more accuracy. Hence adding metrics that are highly particular in detail and analyzing them with an optimal picture would help to unfold various hidden details that could help in the outcome of much more accurate and highly beneficial data. Moreover, results that have generated from the model can be reused for the self-improvement of the model by rolling out the feedback of a particular recommendation that has been previously made by the model. This data could help the model in rectifying wrong recommendations and provide highly accurate decisions in the next recommendations

4. Conclusion

Therefore the approach of this model has a huge impact on the educational institutions and the corporate sector of today's highly competitive world. The model proposes a simple and cross-sector solution to both the corporate and educational sectors that could result in the huge increase of employability solving the problem of wrong decision making of job aspirants as well as mistakes made by the organization whereby suffering losses from those decisions. Hence it could benefit every academician in evaluating his/her students as well as their academic performance in a more sophisticated and a single independent platform that has analysis related to current world trends and scenarios.

The advantage of getting an analysis that is evaluated and compared with data from all over the world could help a user to understand well enough about his/her decision makings and personal evaluation. The only hurdle this model can face is acceptance from educational institutions accepting it to evaluate and generate data from their colleges but if this problem is overseen, then a huge variety of data being generated every day can make and change the model by making it much more accurate system over the days. This way it provides a platform for educational institutions to analyze the students' performance in real-time and his industry level aspects as well. Concluding, the model has a vast scope of improvement as well as can provide great accuracy with positive results in the future.

References

[1]. *Machine Learning* - https://www.sas.com/en_in/insights/analytics/machine-learning.html

[2]. Verma, Jaiprakash & Patel, Bankim & Patel, Atul. (2015). Big Data Analysis: Recommendation System with Hadoop Framework. Proceedings - 2015 *IEEE International Conference on Computational Intelligence and Communication Technology*, CICT 2015. 92-97. 10.1109/CICT.2015.86..

[3]. Michael Yao-Ping Peng, Sheng-Hwa tuan, Feng - Chi Liu “Establishment of Big Data Application for Education Industry”, 2017 *2nd International Conference on Image ,Vision and Computing.*

[4]. Nour Moahammed Almadhoun, P. Dhanapal Durai Dominic and Lai Fung Noon “Social Media as a Promotional Tool in Higher Education in Malaysia”, 978-1-4577-1884-7/11, *IEEE.*

[5]. Ko S., Lee J. (2002) User Preference Mining through Collaborative Filtering and Content based Filtering in Recommender System. In: Bauknecht K., Tjoa A.M., Quirchmayr G. (eds) E-Commerce and Web Technologies. EC-Web 2002. *Lecture Notes in Computer Science*, vol 2455. Springer, Berlin, HeidelbergS.M.R. Islam, D. Kwak, M.D.H. Kabir, M. Hossain, and K. Kwak, "The internet of things for health care: a comprehensive survey," IEEE Access 3, 2015, pp. 678-708.

[6]. Christos Vaitsis, Vasilis Hervatis and Nabil Zary (July 20th 2016). Introduction to Big Data in Education and Its Contribution to the Quality Improvement Processes, Big Data on Real-World Applications, Sebastian Ventura Soto, José M. Luna and Alberto Cano, *IntechOpen*, DOI: 10.5772/63896. Available from: https://www.intechopen.com/books/big-data-on-real-world-applications/introduction-to-big-data-in-education-and-its-contribution-to-the-quality-improvement-processes

[7]. Romero C, Ventura S. Educational data mining: a survey from 1995 to 2005. *Expert Systems with Applications.* 2007;33(1):135–46.

[8]. Introduction to Recommender Systems - https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada

[9]. Yu Chuan, Xu Jieping, and Du Xiaoyong. 2006. Recommendation algorithm combining the user based classified regression and the item-based filtering. *In Proceedings of the 8th international conference on Electronic commerce: The new e-commerce: innovations for conquering current barriers, obstacles and limitations to conducting successful business on the internet (ICEC '06)*. Association for Computing Machinery, New York, NY, USA, 574–578. DOI:https://doi.org/10.1145/1151454.1151463

[10]. Ailin Deng, Yangyong Zhu and Bole Shi, "A Collaborative Filtering Recommendation Algorithm Based on Item Rating F Prediction", 2003 *Journal of Software* , vol.14, No.9.

Author's Biography

Praneet Amul Akash Cherukuri is currently pursuing his Bachelor's in Technology – Computer Sciences & Engineering from CMR Institute of Technology, Hyderabad. He is the CEO of Kirg Technologies an IT startup. He was an International Delegate for the Harvard University's HPAIR Conference 2020, in Boston, USA. His research interests are Data Science, Natural Language Processing, Recommender Systems, Neural Networks, and Prediction Analysis.

How to Cite

Cherukuri, Praneet Amul Akash (2019). Recommender System for Educational & Corporate Sector In Prediction of Domain Recommendations & Analysis using Machine Learning. *International Journal of Machine Learning and Networked Collaborative Engineering*, 3(03) pp 135-142.

doi : https://doi.org/10.30991/IJMLNCE.2019v03i03.002.