INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 10, Issue 3 - V10I3-1184)
Available online at: https://www.ijariit.com

Applying Semantic Analysis and Machine Learning to Develop Recommendation System

Daniyal Ganiuly
222122@astanait.edu.kz
Astana IT University, Kazakhstan

ABSTRACT

Completing elective courses at university has an important role in the career path of future specialists. Making the decision to study courses that cover certain skills and technologies directly influences receiving a job offer after graduation. Despite the high level of importance of this topic, there is an issue when students make wrong decisions about elective courses that impact negatively their academic performance. This research observes a solution to the actual problem by creating software to analyze data and give recommendations about appropriate elective courses for university students.

Keywords: *Semantic analysis, fuzzy logic, recommendation system, SVM*

1. INTRODUCTION

Courses covered at university might come in handy at professional work and that's why it is crucial to select courses as accurately as possible. Students' experience demonstrates that having only information about elective courses in the form of a table with title and description does not fully explain how the discipline is taught and is represented. When students have several options to choose from the list of courses, they intend to select the most optimal courses that might involve most aspects that are in demand in the job market. Since not all students have a clear comprehension of listed courses, there is a chance for learners to select courses that will not meet expectations. Therefore, there is a strong point in creating a solution through advanced guidance between education and professional career.

It was observed that in order to solve the trouble related to inaccurate elective selection the best option would be to give recommendations about courses. The development of a recommendation system is convenient for this type of issue because it is sufficiently flexible to provide advice and leave the learner with his right to make decisions. Analyzing the meaning of words and sentences by applying semantic analysis and machine learning was consistent in integrating with web technologies.

2. PREVIEW STUDIES

Numerous studies have examined the application of basic data mining methods, including support vector machines, decision trees, and neural networks, in educational data mining. One of the previous research projects examined two approaches—Decision Tree

and Bayesian Network—for forecasting bachelor, master, and doctoral students' academic achievement at two distinct academic institutions [1]. Another scientific study utilized association rules in its smart course recommendation system to suggest courses to students based on shared rules, although this approach was not customized for every pupil [2]. Some RS methods that may be employed to forecast learners' academic achievement were proposed by researchers in order to solve the issue of forecasting student progress via RS [3]. They haven't yet developed a specialized tool for course recommendations, though.

The educational system does not always fully utilize all of the technological capabilities to satisfy consumer demands. Learners' interest was not very high since most students chose their courses without doing any research beforehand [4]. Due to inconsistent course descriptions or a lack of knowledge about various areas, students had difficulties choosing optional courses. Fuzzy logic implementation and transcript analysis of finished courses in the past were two potential solutions [8]. The study that was done indicates that another explanation for this kind of problem is the absence of educational and career counseling [6].

In the study conducted in the past, the group filtering technique predicted the item that a student would be passionate about by comparing the user's data in a system with the user and suggesting an item to the user. In accordance with research methodology, the author presents a technique to suggest courses to a learner by examining their educational history and comparing it with that of other pupils in order to determine similarities [5].

3. MATERIALS AND METHODS

The research included data collection, observation, analysis, and development of a web platform. One of the problems that colleges frequently encounter is the recommendation of students for course enrollment. If pupils are able to study what they enjoy or are engaged in, it not only helps them choose what to study but also maximizes their academic achievement. Rule mining techniques, group filtering, and filtering based on content are the techniques that are frequently used for recommendations [10].

3.1 Cosine similarity

Regardless of the scale of difference between two vectors, cosine similarity is a statistic used to assess how similar they are. It is especially helpful for comparing text texts or word embeddings in the context of semantic analysis.

The cosine of the angle formed by two non-zero vectors in an inner product space is measured by cosine similarity. Below is an illustration of the cosine similarity formula between two vectors, x and y.

$$k(x, y) = \frac{xy^\top}{\|x\|\|y\|}$$

3.2 Document Similarity

Each document is represented as a vector in a high-dimensional space for document similarity, with each dimension denoting a distinct corpus term. Term frequencies or TF-IDF scores can be found among the values in the vector. By calculating the cosine of the angle between each document's vector, cosine similarity then gives an indication of how similar two documents are to each other based on their content [9].

3.3 Support Vector Machine

The fundamental model of the SVM, a two-class classification model, is specified in the characteristic space of the greatest linear classifier. Maximizing the interval, that can be converted into a convex quadratic programming issue, is the SVM's training technique [9]. Reducing the standardized hinge loss function is also the same as it.

3.4 Fuzzy clustering

The growing number of training data sets and the escalating intricacy of data collection have led to a growing interest in clustering methods. The aim of the research was to apply fuzzy clustering to huge data sets and samples with great spatial complexity. In order to ensure classification accuracy, fuzzy clustering reduces both the sample's spatial complexity and the quantity of training examples [9].

As a result, multiple kernel learning's effectiveness is increased. Fuzzy C-means (FCM) is the most popular and effective fuzzy clustering technique out of all of them [9]. By maximizing the objective function, it determines the class of sample points and

accomplishes the goal of automatically categorizing the sample information [9]. This allows every sample center to become an element of every class center.

3.5 Model training

Training machines using the sci-kit-learn Python library involves a sequence of essential steps. The process starts with collecting and preparing data, which includes cleaning, transforming, and splitting the data into training and test sets. Next, feature engineering is performed to scale, encode, and select relevant features. An appropriate machine learning model is then selected from the sci-kit-learn extensive library, such as linear models, decision trees, or support vector machines. The model is trained by optimizing parameters to minimize errors. Following training, the model's performance is evaluated using metrics like Mean Squared Error for regression or accuracy and F1-score for classification [10].

3.6 Mean Squared Error

Mean Squared Error (MSE) is a common metric used to evaluate the accuracy of a predictive model, particularly in regression analysis. It measures the average of the squares of the errors, which are the differences between the predicted values and the actual values [7]. Mean Squared Error is calculated by taking the average of the squared differences between each actual value and its corresponding predicted value.

$$\mathrm{MSE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} (y_i - \hat{y}_i)^2.$$

4. WEB PLATFORM

Figure 1 illustrates the example of recommended courses for a job position as “Senior Java Developer”. Text from the course description was analyzed to recommend these elective courses for particular job positions. Each time the update button is clicked, new information from job announcements is analyzed with existing courses. The analysis is conducted every time the button is clicked in order to display actual information since the labor market and trends related to jobs in IT can change quickly.

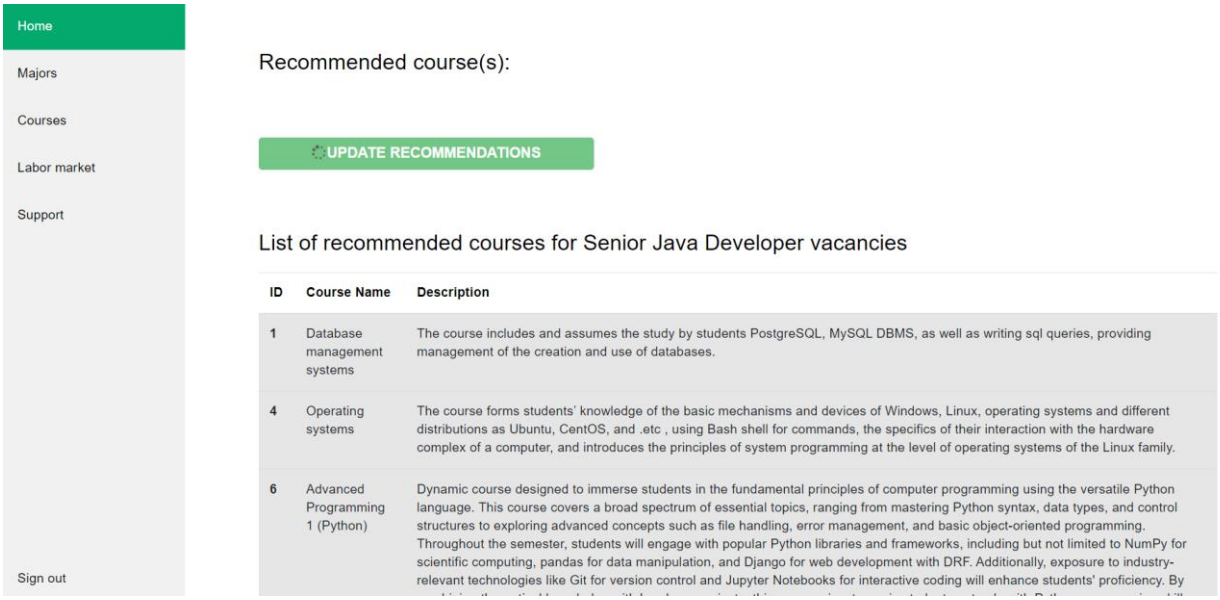

Figure 1: List of recommended courses

The approach of analyzing data about courses and required qualifications and skills from the labor market was used in order to define in-demand technologies that can be included in potential new courses. As shown in Figure 2, the list includes technologies that are not among existing courses at the university. Therefore, there is a recommendation to create elective disciplines that cover these technologies.

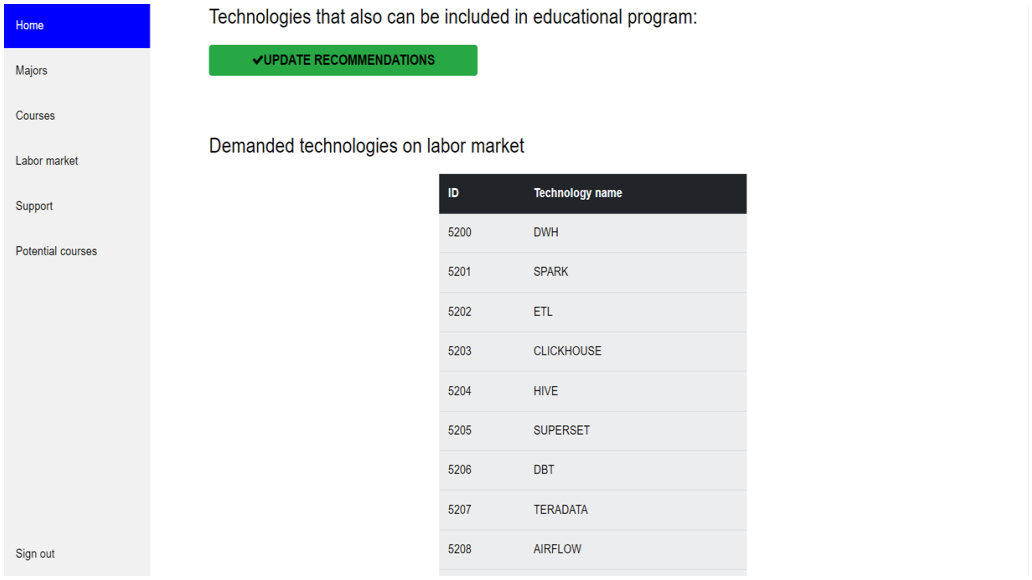

Figure 2: List of potential new courses

5. RESULTS AND DISCUSSION

The project aimed to address the challenge of selecting university elective courses that align with professional work requirements by developing a recommendation system. The primary observation was that students often struggle to understand the full scope and practical implications of courses based solely on titles and descriptions. This can lead to suboptimal course selections that do not meet their expectations or professional needs.

5.1 Key results

1. The semantic analysis component effectively extracted and analyzed key topics and skills from course descriptions, offering a more comprehensive understanding of each course.
2. The machine learning model, trained on historical data of course selections and career outcomes, provided personalized course suggestions that matched students' professional aspirations.

5.2 Discussion

The implementation of a recommendation system proved to be a practical solution to the problem of inaccurate elective course selection. By incorporating semantic analysis, the system could interpret the nuanced content of course descriptions, which traditional tables fail to convey. The use of machine learning allowed the system to adapt to the evolving job market and provide up-to-date recommendations.

The project highlights the importance of bridging the gap between educational offerings and professional requirements. It demonstrates that advanced technologies can play a significant role in enhancing educational guidance and ensuring that students make informed decisions that align with their career goals. Future enhancements could include real-time feedback from employers about in-demand skills and more sophisticated algorithms to further refine recommendations.

6. CONCLUSION

Analyzing information about courses and job announcements was beneficial to give recommendations about elective courses accurately for students. Semantic analysis was applied to find the similarity of meaning of information about courses and job vacancies. The practical implementation of the project included the development of a web platform with computational operations on the backstage with semantic analysis and machine learning. The problem related to unconscious decisions that influence negatively academic performance due to loss of interest and other factors was solved with creative solutions through a web platform. Overall, the recommendations about educational disciplines increase the accuracy of selecting elective courses.

7. ACKNOWLEDGEMENT

I would like to thank Assel Smaiyl, the research supervisor, for his guidance and suggestions on how to make this scientific research better.

8. REFERENCES

[1] Nguyen T, Janecek P, Haddawy P. A comparative analysis of techniques for predicting academic performance. In: Proceedings of the 37th ASEE/IEEE Frontiers in Education (FIE 2007). 2007.
[2] Al-Sarem M. Building a Decision Tree Model for Academic Advising Affairs Based on the Algorithm C4.5. IJCSI International Journal of Computer Science Issues. 2015 Sep;12(5). ISSN (Print): 1694-0814.
[3] Nguyen T, Horváth T, Schmidt-Thieme L. Factorization Models for Forecasting Student Performance. In: Pechenizkiy M, Calders T, Conati C, Ventura S, Romero C, Stamper J, editors. Proceedings of the 4th International Conference on Educational Data Mining (EDM 2011). 2011.
[4] Chen Z, Liu X, Shang L. Improved course recommendation algorithm based on collaborative filtering. In: 2020 International Conference on Big Data and Informatization Education (ICBDIE). Zhangjiajie, China; 2020. p. 466-469.
[5] Adak M, Yumusak N, Taskin H. An elective course suggestion system developed in computer engineering department using fuzzy logic. In: 2016 International Conference on Industrial Informatics and Computer Systems (CIICS). Sharjah, United Arab Emirates; 2016. p. 1-5.
[6] Froyd J, Borrego M, Cutler S, Henderson C, Prince MJ. Estimates of Use of Research-Based Instructional Strategies in Core Electrical or Computer Engineering Courses. IEEE Trans Educ. 2013;56(4):393-399.
[7] Bhumichitr K, Channarukul S, Saejiem N, Jiamthapthaksin R, Nongpong K. Recommender Systems for university elective course recommendation. In: 2017 14th International Joint Conference on Computer Science and Software Engineering (JCSSE). Nakhon Si Thammarat, Thailand; 2017. p. 1-5.
[8] Bydžovská H. Course Enrollment Recommender System. In: Proceedings of the 9th International Conference on Educational Data Mining. 2016. p. 312-317.
[9] Alodadi M, Janeja VP. Similarity in Patient Support Forums Using TF-IDF and Cosine Similarity Metrics. In: 2015 International Conference on Healthcare Informatics. Dallas, TX, USA; 2015. p. 521-522. [10] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011.