Mining learners' Data to Perceive the Need for German for Academic Purposes

Subhasri Vijayakumar, Syed Ibrahim S P

Abstract: ***Learning a foreign language at the tertiary level opens up many opportunities for the learners and is indeed an essential requirement for those aspiring to continue their education abroad. With Germany becoming one of the most preferred destinations for higher studies, along with the general German language skills, the academic skills that would be needed is an aspect worth analyzing in the given context of growing demands. Educational data mining is an emerging field and analysis using data mining techniques in educational settings aid in better understanding of the learners, their learning environment and their learning needs. Through this study, by applying one of the data mining techniques called "Clustering", we explored the German learners' perception of which academic skill they deem important to be learned in German at the tertiary level. A significant difference in the perception of German learners when it comes to learning academic skills in German was found between the two clusters that were formed. This divide in perceptions among learners indicates the awareness of the learners about studying in German universities or the lack thereof. Educational data mining and its techniques aid in significant decision making that could enhance the teaching-learning process of German. These findings are discussed in this paper in light of augmenting the German curriculum at the tertiary sector.***

Keywords: ***Educational Data Mining, Foreign Language learning, K-Means Clustering, Needs Analyses, German for Academic Purposes***

I. INTRODUCTION

A famous quote by Johann Wolfgang von Goethe states, "Those who know nothing of foreign languages know nothing of their own". Learning another language opens the doors of possibilities and in this era of interconnectedness and borderless working, learning a foreign language gives that additional impetus to succeed. Be it for better employment opportunities or for facilitating higher studies abroad, learners have learned a foreign language and the trend continues. To meet the needs of the growing economy, academia has also started teaching different languages catering to the varied requirements of the stakeholders of education. While the USA and other English speaking countries might still be the most preferred destinations for higher studies abroad among the Indians, there has been a steady increase in the number of Indian learners in German Universities too.

According to DAAD - the German Academic Exchange Service's statistical report, the number of Indian students in Germany has grown by 13.14% over the last year and stands at 17,570 during the Winter Semester 2017-18 [1]. Germany is a safe country and many of the study programmes are now fully offered in English medium of instruction also. The courses combine theory and practice and most universities do not charge tuition fees. The flexible work permit policies are also a boon to international students trying to establish their career in Germany. Proficiency in German is certainly a big advantage as it helps in day-to-day conversations and integrating with the target community. Along with the general German language skills, the academic skills that would also be needed are an aspect worth analyzing in the given context of growing demands.

A. Objective of the study

The main objective of this study is to determine the academic skills in German that were deemed important by the learners of the language by applying data mining techniques. Although various statistical tools are constantly employed to analyze the data collected and make choice decisions on the findings, applying data mining techniques in educational settings have been a recent trend only. The goal of data mining is to find a novel pattern and extract useful knowledge from large amounts of raw data. It is widely used in Business and it has scarce applications in Education. Educational Data Mining (EDM) is an upcoming research field dealing with mining of educational data for knowledge extraction. EDM is being used to identify learning needs from different group of students, to assess/predict learners' performances, optimize subject curriculum renewal etc. EDM includes data mining, machine learning, psychometrics, computational modeling, statistics and information visualization [2]. According to Baker classification prediction, clustering, relationship mining, discovery with models and distillation of data for human judgment form a part of DM categories [3]. In the forthcoming paragraphs, we explain different data mining techniques and how it is used in educational settings and the techniques that we have applied in this study to analyze the learner data collected.

Revised Manuscript Received on July 10, 2019.
Subhasri vijayakumar, VIT Chennai, Vandalur-Kelambakkam Road, Chennai, Tamil Nadu, India
Syed Ibrahim S P, VIT Chennai, Vandalur-Kelambakkam Road, Chennai, Tamil Nadu, India

II. REVIEW OF LITERATURE

In general, educational data mining uses the same methods as those in the data mining field. As stated by Baker the frequently used methods of data mining are classification and regression, clustering, association rule mining, outlier detection, social network analysis, text mining, sequential pattern mining and visualization techniques [3]. According to Romero and Ventura, the frequently used data mining methods in educational data mining are regression, clustering, classification and association rule mining [4]. Zimmermann et al. studied the 'predictive power of undergraduate performance indicators and their aggregates using regression models. Through their study they analyzed graduate level performance using indicators of undergraduate-level performance and found that "undergraduate achievements are highly indicative of graduate-level success (54% explained variance) and that the third-year GPA is the most important explanatory variable" [5]. Waters et al. proposed a 'novel methodology for collaboration-type identification' in a Bayesian setting and inferred the likeliness of collaboration among learners in online courses and classified the type of collaboration employed. They also established the efficacy of this model on both synthetic and real world educational data [6]. Chauhan et al. applied K-Means clustering algorithm in an educational environment to analyze student's data. Using the cluster analysis they segmented students into groups according to the predominant characteristics. They predicted students' placement behaviour and such analysis would enhance the quality of the educational system [7]. Dutt et al. reviewed various journal papers that had used different clustering algorithms in various educational settings. From profiling students to improving the quality of learning in MOOC or online courses or LMS, various clustering methods like K-means, C-means, Apriori algorithms are used. Their work synthesized the disparate entities found in literature and collectively addressed their significance in the field of educational data mining [8]. Cocea and Weibelzahl in their study explored the impact of learners' motivation on learning in e-Learning environments. They investigated the log files data to determine learners' motivation using a decision tree. Although the results could not be generalized since the data analyzed was minimal, the study confirmed that "general indicator of the motivational level could be predicted from very basic data commonly recorded in log files" [9]. Bouchet et al. used clustering technique to profile the learners based on their interaction with an intelligent tutoring system. In their study they analyzed if student clusters can be established according to their performance and interaction with the tutoring system – MetaTutor and examined how the students relate to their use of self-regulated learning (SRL) process. They ascertained that the students' profiles differed in terms of performance between the three clusters and the "prompts they received by the system to perform SRL processes" [10]. Ashok and Kumar in their study blended heuristic and K-means algorithm to cluster "efficient students to predict placement opportunities". They clustered students based on Knowledge, Communication skill and Attitude and proposed an algorithm to accomplish the same [11]. Based on earlier works it can be reiterated that K-means clustering is a simple and a well-known clustering algorithm for performing non-hierarchal clustering to uncover relationships and associations between variables in the data. While techniques like classification and prediction analyze class labeled data objects, clustering analyzes the data without referring to the class labels. In clustering, the intra-class similarities are maximized while the inter-class similarities are minimized. Hence each cluster formed is a class of objects from which rules can be derived. This study also therefore employs k-means clustering to ascertain the academic skills in German that were deemed important by the learners of the language.

III. MATERIALS AND METHODS

The data used here in this study were collected as part of a larger study that was undertaken for doctoral research on German needs analysis [12]. The data from 330 learners of German at the tertiary sector and their responses to six academic skills related questions in the larger questionnaire were used in this current study. K-means clustering algorithms were used to group the learners based on how they deemed the below given six academic skills in German to be important in their learning of the language.

1. Listening to lectures/seminars/presentations/ group discussions
2. Writing essays/note taking
3. Attending job interviews
4. Writing Resumes, covering letters
5. Reading Course Handouts/ Lab Manuals/ Study notes
6. Reading technical Journals/Papers/Books

The system architecture is as shown in Figure 1 below.

Figure 1: System Architecture

A. K-Means Clustering

K-means is an unsupervised algorithm that takes a set of unlabeled data objects and a number 'k' that indicates the number of clusters.

The algorithm partitions the given 'n' objects into 'k' clusters so that the intra-cluster similarity is high and the inter-cluster similarity is low. Cluster similarity is measured using the mean value of the objects in a cluster, which can be viewed as the centroid or center of gravity of the clusters.

The steps in the k-means algorithm are as follows:

1. Read the number of clusters, 'k'
2. Select 'k' objects randomly from the input set of data objects. Initially, each object represents a cluster mean or center.
3. For each of the remaining objects in the input data set, compute Euclidean distance between that object and each of the clusters mean and assign the object to the cluster to which it is the most similar.
4. Compute new mean for each cluster.
5. Repeat Step 3 – 5, until the criterion function converges.

The findings of the study and the discussion of the results are as follows.

IV. RESULTS AND DISCUSSION

The 330 learners of German at tertiary sector came from four major engineering streams like Mechanical, Electrical/Electronics, Computer science and aeronautical. Figure 2 below shows the learners from each of these streams. 170 learners were from mechanical or mechanical related streams, 60 from electrical/electronic engineering, 48 from computer science and 52 from aeronautical engineering.

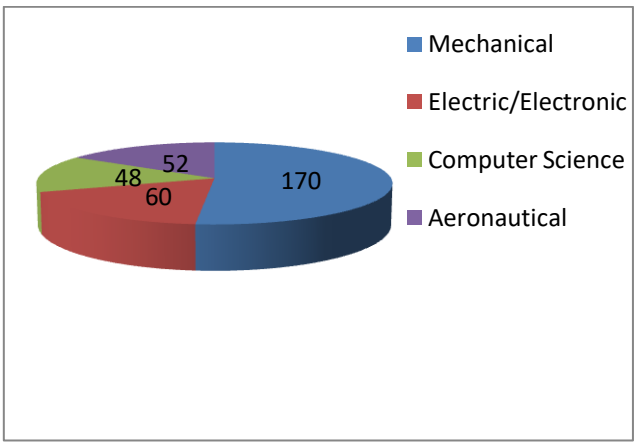

Fig 2: Total number of learners from different streams

After K-Means clustering was used on this data, two clusters were formed and the centroid of these clusters is as given in Table I below.

Table I: Centroid

Questions	Q1	Q2	Q3	Q4	Q5	Q6
Cluster 0	2.060	1.500	1.983	1.716	1.655	1.845
Cluster 1	2.575	2.033	2.860	2.743	2.266	2.355

The number of students under each of these clusters is as shown in Table II and III below. 117 students – 66 from Mechanical, 17 from Electrical/Electronic, 14 from Computer science and 20 from aeronautical form Cluster 0 as illustrated in Table II.

Table II: Total number of students in Cluster 0

Branch	mechanical	Electrical/ Electronic	Computer Science	Aeronautical
No. of Student	66	17	14	20

A total of 213 students – 105 from Mechanical, 41 from Electrical/Electronic, 34 from Computer Science and 33 from Aeronautical form Cluster 1 as illustrated in Table III.

Table III: Total Number of students in cluster 1

Branch	mechanical	Electrical/ Electronic	Computer Science	Aeronautical
No. of Student	105	41	34	33

Table I above shows the centroid and the two clusters formed based on the learners' response to the six academic skills. From Tables I - III, it is evident that there is a clear difference in the perception of German learners when it comes to learning academic skills in German. A total of 213 learners in Cluster 1 clearly deem learning of these academic skills in German as very important. They perceive German to be more important for attending job interviews (Q3), for writing resumes and cover letter (Q4), to listen to academic lectures/seminars/presentations etc. (Q1). It is also observed that Cluster 0, with 117 learners have deemed all the six academic skills to be not so important with only lectures/seminars etc. in German gaining an important status compared to the other skills. This divide in perception among learners indicates the awareness of the learners about studying in German universities or the lack thereof. Germany being a monolingual country had brought in various policy changes in its educational system, to meet the demands of globalization and internationalization of higher education. To cater to the international student community, English-taught programs at UG and PG level, improved post-study work rights for foreign students, flexible internships/part-time job opportunities while studying were introduced. These have a major impact on English being the language of academic communication.

Hence the use of German for listening to lectures/presentations/seminars (Q1) or for writing essays/note taking (Q2) or for reading lab manuals/course materials (Q5) is rare. With the advent of online translations and the Internet, many technical journals or articles are also mostly available in English. Moreover, to understand technical German which is completely different from technical English, the language courses offered in India also do not cover technical jargons. There is a wealth of scientific knowledge in German Journals/books but German language teaching has to include the necessary learning skills to enable their learners to effectively use this information. While job-related policies in Germany might be helpful for international students, the German job market is still very conservative and they do value German proficiency as major criteria for recruitment. While English might be spoken mostly within the walls of the academic institution, the general public in the country is by and large only monolingual. Part-time jobs or interning opportunities at small/medium sized industries or customer service industries definitely demand a good command in German and the ability to write a cover letter/resume or take an interview in German. While German universities might not charge tuition fees for many of its courses, language classes are expensive and might not be viable to international students also. From these analyses, we can hence deduce that the learners of German at tertiary level need to be made more aware of the academic skills required by them in German universities. The current curriculum of German should also reflect these requirements in teaching the language so that learners get better equipped to meet the demands there.

V. CONCLUSION

"EDM converts raw data coming from educational systems into useful information that could potentially have a greater impact on educational research and practice" [13]. As learners always look up at academia to help them learn what is relevant, useful and needed to meet the growing demands of the globalized world, this study through the application of data mining techniques aids in making the right decisions for the learners. K-Means clustering has been found to be computationally faster than hierarchical clustering. The tight clusters so produced enabled in better understanding of students' perceptions on academic skills in German. It also facilitates teachers of German and the academia to reflect on the skills taught in German and to bring in necessary changes in the curriculum or the teaching-learning process. Learning a foreign language is always contextual. Foreign language teaching has always emphasized what to learn and covers a plethora of general, professional and academic skills. Such techniques can be used on various learners' data to analyze their needs, learning, performances etc. and these can then be effectively used by the academia to fortify their language syllabus to meet the needs of all the stakeholders of tertiary education.

REFERENCES

1. DAAD, Available at https://www.daad.in/en/2018/10/04/indian-student-numbers-in-germany-touch-a-record-high/ , accessed January 2019.
2. A Algarni, Data Mining in Education, International Journal of Advanced Computer Science and Apllications. 2016; 7(6), 456-461.
3. R Baker et al, Data mining foe education, International encyclopedia for education. 2010; 7, 112-118.
4. R Cristobal and S Ventura, Data Mining In Education, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2012; 3(1), 12-27. Accessed December 2018.
5. J Zimmermann, K Brodersen, H Heinimann, J Buhmann, A Model-Based Approach to Predicting Graduate-Level Performance Using Indicators of Undergraduate-Level Performance, Journal of Educational Data Mining, 2015; 7(3), 151-176.
6. A Waters, C Studer, R Baraniuk, Collaboration-type identification in educational datasets, Journal of Educational Data Mining, 2014; 6(1), 28-52.
7. A Chauhan, V Chauhan, N Sharma, Application of K-means Clustering Algorithm for Mining Educational Data, International Journal of Computer Science and Technology, 2012; 3(4), 473-477.
8. A Dutt, S Aghabozrgi, M Ismail, H Mahroeian, Clustering Algorithms Applied in Educational Data Mining, International Journal of Information and Electronics Engineering, 2015; 5(2), 112-116.
9. M Cocea and S Weibelzahl, Can log files analysis estimate learners' level of motivation?, In: Lernen - Wissensentdeckung - Adaptivität, 9-11 October 2006, Hildesheim, http://trap.ncirl.ie/1033/1/Can_log_files_analysis_estimate_learners%27_level_of_motivation.pdf. Accessed December 2018
10. F Bouchet, J Harley, G Trevors, R Azevedo, Clustering and Profiling Students According to their Interactions with an Intelligent Tutoring System Fostering Self-Regulated Learning. JEDM | Journal of Educational Data Mining, 2013; 5(1), 104-146. Retrieved from https://jedm.educationaldatamining.org/index.php/JEDM/article/view/32
11. M V Ashok and G Hareeshkumar, Educational Data Mining: A blend of Heuristic and K-means algorithm to cluster students to predict placement chance, International Journal of Trend in Scientific Research and Development (IJTSRD), 2018; 2(6), 1401-1406.
12. S Vijayakumar. 2018, A Needs Based Approach to Designing German Curriculum for Engineering Students in Technical Universities / Institutions in Southern India: A Mixed Methods Inquiry, Ph.D.Dissertation. Vellore Institute of Technology, Chennai, India.
13. C Romero and S Ventura, Educational Data Mining: A Review of the state of the art, IEEE Transactions on Systems Man and Cybernetics Part C, 2010; 40(6), 601-618.

AUTHORS PROFILE

Dr. Subhasri Vijayakumar is a faculty of German with more than ten years of teaching and training experience in schools, colleges, universities and industries. Her research interest is in the field of German language teaching, Gamification, Curriculum Design, intercultural competency, technology aided language teaching and learning and comparative studies. She has done her majors in Sanskrit and is proficient in seven languages. She is a voracious reader and occasionally writes stories and poems in English and Tamil.

Dr. S.P. Syed Ibrahim has completed his Bachelors, Masters and Ph.D. in Computer Science and Engineering. He is working as a Professor for the past 6 years at VIT Chennai. He is a recipient of the Best Faculty Award from VIT for the year 2016 and EMC academic alliance award for the year 2013. He has published more than 60 research papers in journals and conferences. He has organized four short term courses on analytics for students from universities abroad and also organized DST sponsored national workshop on data science research.