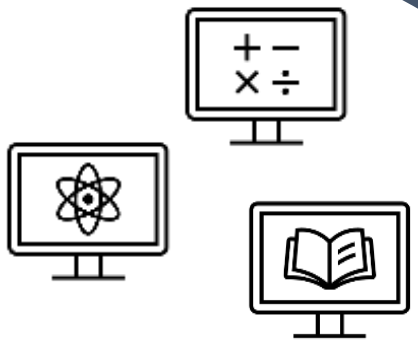
Embedding-based Clustering Analysis of User Reviews: A Case Study of Distance Learning in Saudi Arabia

By : Rahma Hassan

Agenda



- Overview
- Problem Statement
- Objectives
- Research Questions
- Proposed Solution
- Literature Review
- Methodology
- Expected Findings
- Evaluation
- Timeline
- Conclusion
- References

Overview

E-learning is the use of electronic tools and techniques for delivering information to students.

The prevalence of technology all over the world has made e-learning more popular recently.

E-learning has increasing importance day after day especially after corona-virus prevalence worldwide.

The National Center for E-learning and Distance Learning was established to create the E-learning systems in Saudi Arabia.

E-learning in Saudi Arabia started to be implemented years ago before Covid-19 in many institutions of higher education.

Although this, E-learning has been adopted much more since Covid-19 even in primary schools.

To the best of my knowledge, there are not many studies examining the experience of using E-learning in Saudi Arabia through Covid-19 time.

E-learning have several advantages and have disadvantages too.

Students face some obstacles and challenges.

Overview-Cont.

There is a need for evaluating, improving E-learning and motivating students for it.

More studies are needed to examine the effectiveness of E-learning and elaborate the difficulties teachers, students and their parents face.

Primary education is a critical stage in the educational life of students in which they learn the basic concepts in different subjects.

Therefore, there is need to evaluate E-learning to get useful insights about it and hence improve it.

The performance of online educational training platforms can be evaluated through **sentimental analysis by NLP and text mining techniques.**

Text mining deals with structured and unstructured data, data mining deal with structured data sets, so text mining deal with more complex data.

Therefore, we will do sentiment analysis for textual data obtained from students and their parent reviews.

Problem Statement

Evaluating, and examining E-learning is an important thing that will shed light on improving educational issues related to E-learning.

It will be more efficient to make analysis by advanced Artificial intelligence methods instead of doing it manually

Therefore, we will use NLP and data mining to perform this analysis in elementary School in ‘Saudi Arabia’ for students at sixth grade.

We chose ‘sixth grade’ because students at this age can express their opinions by text well.

We will apply our model to the suitable form of textual data we have. The universal Sentence Encoder (USE) NLP encoding technique will be used to encode data.

We might use different NLP techniques to make the sentiment analysis we want and compare them .

Research Questions

1.To what extent the students' and their parents' reviews are positive or negative or in-between regarding different sections of the E-learning process?

2. What are the levels of understanding of students of educational content through virtual environments?

3. What are the main problems that encounter sixth-grade school students and their parents in E-learning?

4. What are the experience levels of teachers in E-learning which will be inferred from students' points of view?

Objectives

- Explore the best models of NLP and text mining which can give us the best results in representing the sentiment.
- Make effective analysis to classify the main problems that are encountered by primary students.
- Obtain useful insights from the classification we will perform on textual reviews
- Get clusters on similar reviews to know the distribution of students' reviews.

Proposed Solution

Reviews of students and reviews of their parents will be taken, processed, and analyzed.

We will use the universal Sentence Encoder (USE) NLP encoding technique to encode data.

We might use Different NLP and text mining models to be applied to textual reviews of students to get their important sentiments of them.
We can compare these different NLP models

Implement classification to classify different sentiments of reviews and get the right rating label for review.

Cluster similar reviews together (cluster numerical sentiments vectors) using clustering ML algorithms.

Literature

Paper name	Author	Source	Data	Findings
Analysis of Blended Learning Application Using Text Mining Method.	(Wang, L., Huang, Y., & Omar, M. 2021).	International Journal of Emerging Technologies in Learning (IJET).	The directory summarized the basic data of 675 schools in 17 countries. The number of students participating was 150,000.	The most important obstacles facing the effective development of blended learning after applying Text mining which is the lack effective training of teachers, and effective preparing of students' abilities.
Student Evaluation of Teaching in Business Education: Discovering Student Sentiments Using Text Mining Techniques.	Baddam, S., Bingi, P., & Shuva, S. (2019).	E-Journal of Business Education and Scholarship of Teaching.	A total of 939 usable comments are retrieved across 110 business courses. The data is obtained from the business school in the United States.	In this study, the analysis showed significant differences in sentiments among students. The results also showed that the nature of the course also affects the sentiments of students, especially in the hybrid online courses.

Literature–cont.

Paper name	Author	Source	Data	Findings
Identifying child abuse through text mining and machine learning.	Amrit, C., Paauw, T., Aly, R., & Lavric, M. (2017).	Expert systems with applications.	The data consisted of (partly medical) files on 13.170 children born in the Amsterdam region.	The resulting of this paper confirmed that depending on text mining and machine learning scored highly in classifying potential cases of child abuse.
Evaluating a bilingual text-mining system with taxonomy of key words and hierarchical visualization for understanding learner-generated text.	(Kong et al., 2018)	Journal of Educational Computing Research	A class of 27 in-service teachers studied a course “e-Learning in primary mathematics” was asked to reflect “what is e-Learning” before and after the course.	It is possible to use text mining techniques as a supportive tool for teachers to gain a deeper understanding of the texts that learners generate to support more effective educational decision-making.

Literature-cont.

Paper name	Author	Source	Data	Findings
Evaluation of an educational training platform using text mining.	(Spatiotis et al., 2018)	In Proceedings of the 10th Hellenic Conference on Artificial Intelligence	collected from the questionnaires from over 2600 of teachers who participated and expressed their opinions by answering multiple choice and open-ended questions.	The results of this research confirmed that the performance of the education system can be improved by extracting and processing educational data using text mining and sentiment analysis techniques.
A Text Mining Analysis of Academic Libraries' Tweets.	(Al-Daihani and Abrahams, 2016).	The Journal of Academic Librarianship	Here the researchers applied text mining to a large dataset of tweets nearly 23,707 tweets	The results of the research is to understand the social data of academic libraries to aid in decision-making and strategic planning for patron outreach and marketing of services.

Literature-cont.

Paper name	Author	Source	Data	Findings
Mining social media data for understanding students' learning experiences.	(Patil & Kulkarni, 2018)	IEEE Transactions on Learning Technologies	Depending on the results of the mining of Tweets classification on approximately 35,000 Tweets that were posted from within the geographic location of Purdue University.	The results confirmed that through this research, the concerned authorities at the university can know more about the experiences and issues of students in order to take care of it and solve it.
Mining Student-Generated Textual Data in MOOCS and Quantifying Their Effects on Student Performance and Learning Outcomes.	(Tucker et al., 2014)	ASEE Computers in Education Journal	The sample of single course enrolment in MOOCs can range between 10,000 to 200,000 students, hereby providing a potentially rich venue for large scale digital data.	The researchers concluded that the textual content of discussions expressed by students within the online discussion forums in MOOCs can be used as indicators of student achievement.

Methodology

Data Acquisition

Qualitative Analysis

Quantitative Analysis

Overview on Methodology

Data:
1.Data Collecting

The objective is to collect at least 3000 tweets.

Structured Data and Unstructured Data.

Data will be collected from Two sources:

1-Tweets of students and their parents on ‘Twitter’

2-Data from educational surveys.

A data sample like:

2.Data Cleaning and Preprocessing

Like (not limited to):

Removal of URLs, special characters, and punctuations.

Tweets that have less than 3 words will be discarded.

All numbers will be replaced by the tag <number>.

If English, all letters will be converted to lower-case.

Quantitative Analysis

Sentiment Analysis Model

The machine learning model utilizes sentence embeddings in building a sentiment classifier.

We will use USE to encode data and transform sentiments to vector representations.

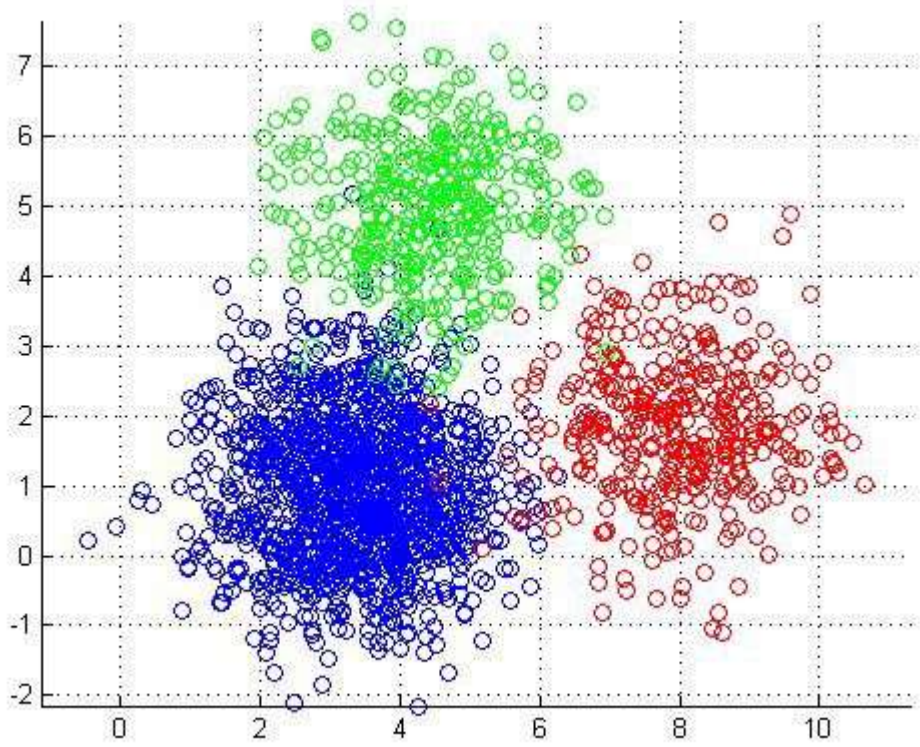
Ensured that all texts in the dataset have **sentiment labels.**

We will use classification to classify different sentiments whether they are Excellent ,moderate , bad**.**

Quantitative Analysis

Clustering Analysis

- Cluster similar reviews together.
- Use the numerical vectors of user reviews.
- **For example,** using **MeanShift** or **DBSCAN**.
- Make a dimensionality reduction of the numerical vectors using
- **For example,** using **PCA, t-SNE,** or **UMAP**.



Qualitative Analysis

Visualization Analysis:

- The distribution of sentiment in all text as a bar or pie chart.
- A map color-coded with the average sentiment in each region
- A bar chart of average sentiment per age-group.
- A bar chart of average sentiment per gender.

Evaluation

The evaluation component is the **sentiment classification** machine learning model.

- **Comparing the predicted sentiment of the users with the actual rating scores/labels they provided in the dataset.**

- **Another survey would be constructed to enquire reviewers and decision maker about their agreement with the findings and recommendations.**

Expected Findings:

We expect to get insight about different **obstacles** students face in E-learning system.

We expect to get **satisfaction scores** to reflect the general sentiment of users towards the subject.

Moreover, we expect to find **directions for improvements** from users' reviews.

Conclusion

E-learning has growing importance in era of technology and time of Covid-19 epidemics.

Evaluate different aspects of E-learning in primary education at ‘Saudi Arabia’ school for sixth grade students.

Make sentiment analysis of reviews of E-learning users using AI technology and will use NLP methodologies.

Extract insights about the main problems facing the students.

Obtain satisfaction levels of users about different aspects of E-learning like Ease of use of technology, understanding lesson, etc.

Insights and recommendations can be obtained from this analysis and will be useful for educational stakeholders.

References

[**50]** Meji, A., Dennison, M. S., Mobisha, M., & Umar, M. M. (2021). The Challenges in E-Learning Among Tamilnadu Students in the Covid-19 Induced Crisis. I-Manager's Journal on Humanities & Social Sciences, 1(February), 33–40.

[52] Mikolov, T., K. Chen, G. Corrado, and J. Dean, 'Efficient Estimation of Word Representations in Vector Space', arXiv:1301.3781 [cs], Sep. 2013, Accessed: Feb. 17, 2022.

[53] Moubayed, A., Injadat, M., Nassif, A. B., Lutfiyya, H., & Shami, A. (2018). E-Learning: Challenges and Research Opportunities Using Machine Learning Data Analytics. IEEE Access, 6, 39117–39138.

[26] Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., & Romero, C. (2019). Text mining in education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(6).

[94] Xing, W., Lee, H. S., & Shibani, A. (2020). Identifying patterns in students' scientific argumentation: content analysis through text mining using Latent Dirichlet Allocation. Educational Technology Research and Development, 68(5), 2185–2214.

[73] Salloum, S. A., Alshurideh, M., Elnagar, A., & Shaalan, K. (2020). Mining in educational data: review and future directions. Joint European-US Workshop on Applications of Invariance in Computer Vision. Springer International Publishing.

[82] S. Thavareesan and S. Mahesan, 'Sentiment Analysis in Tamil Texts: A Study on Machine Learning Techniques and Feature Representation', in 2019 14th Conference on Industrial and Information Systems (ICIIS), 2019, pp. 320–325. doi: 10.1109/ICIIS47346.2019.9063341.

[46] K. Pipalia, R. Bhadja, and M. Shukla, 'Comparative Analysis of Different Transformer Based Architectures Used in Sentiment Analysis', in 2020 9th International Conference System Modeling and Advancement in Research Trends (SMART), 2020, pp. 411–415. doi: 10.1109/SMART50582.2020.9337081.

[39] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding', arXiv:1810.04805 [cs], May 2019, Accessed: Mar. 22, 2022.

[65] R. Al-Rfou, D. Choe, N. Constant, M. Guo, and L. Jones, 'Character-level language modeling with deeper self-attention', in Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, Honolulu, Hawaii, USA, 2019, pp. 3159–3166. doi: 10.1609/aaai.v33i01.33013159.

[67] Rashed, A., Kutlu, M., Darwish, K., Elsayed, T., & Bayrak, C. (2020). Embeddings-based clustering for target specific stances: The case of a polarized Turkey. Association for the Advancement of Artificial Intelligence.

[51] M. K. Dahouda and I. Joe, 'A Deep-Learned Embedding Technique for Categorical Features Encoding', IEEE Access, vol. 9, pp. 114381–114391, 2021, doi: 10.1109/ACCESS.2021.3104357.

[79] Soliman, A. B., K. Eissa, and S. R. El-Beltagy, 'AraVec: A set of Arabic Word Embedding Models for use in Arabic NLP', Procedia Computer Science, vol. 117, pp. 256–265, 2017, doi: 10.1016/j.procs.2017.10.117.