iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VI **Month of publication:** June 2023

DOI: https://doi.org/10.22214/ijraset.2023.54508

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

MOOC Data Analytics through Text Mining-An Innovative Approach to Learning Improvement

Y.V.Ramesh[1], S.Shanmukh Rao[2]
[1, 2]*Faculty, CSE Department, GIST, Nellore*

Abstract: ***The COVID-19 pandemic has brought about significant changes in the perception of education, with Massive Open Online Course (MOOC) providers like Coursera witnessing a surge in millions of new user registrations on their platforms. However, despite the prevalence of online review systems in various industries, the MOOC ecosystem lacks a standardized or fully decentralized review system. We believe that there is an opportunity to utilize existing open MOOC reviews to create user-friendly and transparent reviewing systems, enabling learners to easily identify the top courses available. By leveraging the wealth of reviews already available in the MOOC ecosystem, we can create simpler and more transparent systems that empower users to make informed choices about the courses they enrol in. In our research, we conduct an analysis of reviews from the Coursera platform with the specific goal of determining the potential value of using NLP-driven sentiment analysis on textual reviews in providing valuable information to learners. By examining the sentiment expressed in the textual reviews, we aim to evaluate whether this approach can offer meaningful insights to learners in assessing the quality of MOOCs. The results of our research suggest that textual reviews may be a more advantageous choice compared to numeric ratings due to the disadvantages associated with numeric ratings, such as the potential for random or arbitrary selections. Our findings indicate that utilizing sentiment analysis on textual reviews could provide valuable information for learners in evaluating the quality of MOOCs. By relying on the rich and descriptive information conveyed through textual reviews, learners may be better equipped to make informed decisions when selecting courses on platforms like Coursera.***

Keywords: Text mining, Natural Language Processing, MOOC, Sentiment Analysis.

I. INTRODUCTION

MOOCs, or Massive Open Online Courses, have become a valuable tool in modern education due to their accessibility, flexibility, and diverse range of topics. They provide high-quality education to learners globally, overcoming geographical and financial barriers. MOOCs can be accessed online from anywhere, allowing learners to set their own pace and schedule, making them suitable for various individuals with different lifestyles. Moreover, MOOCs cover a wide range of subjects, from technical skills to humanities, arts, and business, catering to diverse interests and career goals.

However, selecting the right MOOC course can be overwhelming due to the abundance of options available. To choose the most suitable course, it is essential to consider several factors. Learners should evaluate their goals and ensure that the course aligns with them. They should review the syllabus, curriculum, and learning materials to determine their relevance and quality. Seeking recommendations from faculty members or friends, checking online reviews, and researching instructors' qualifications can provide valuable insights. It is also important to assess the course format, duration, time commitment, and associated costs to ensure they fit one's schedule, learning style, and budget.

While relying solely on number ratings and average ratings for course selection may seem convenient, it has potential drawbacks. Ratings can be subjective and influenced by bias, and a small sample size may not accurately reflect the course's quality. Additionally, fake or manipulated ratings can mislead learners. Therefore, it is crucial to consider multiple factors and cross-reference information from different sources when evaluating the suitability of a course.

Text mining techniques offer a solution to overcome the limitations of relying solely on ratings. By analyzing textual data such as learner reviews, course descriptions, and forum discussions, learners can extract valuable insights using natural language processing (NLP) techniques.

Text mining helps identify common themes, sentiment, and keywords related to course content, providing a more objective understanding of a course's relevance and quality. It also helps identify red flags or recurring issues that may not be apparent in ratings alone. Leveraging MOOC data analytics through text mining enables learners to make more informed decisions that align with their learning objectives and preferences.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

II. LITERATURE REVIEW

The paper "Exploiting sentiment analysis to track emotions in students' learning diaries" by M. Munezero, C.S. Montero, M. Mozgovoy, and E. Sutinen presents a system for automatically analyzing and visualizing student emotions expressed in their learning diaries. The authors recognized the challenge of manually analyzing a large number of diaries and developed a system that extracts emotions based on Plutchik's eight emotion categories. The system was tested on real course data and successfully provided valuable feedback to instructors about students' emotional well-being and engagement in the course. The authors conclude that their system can enhance communication between instructors and students and improve the learning experience.

The paper "A Bootstrapping Based Refinement Framework for Mining Opinion Words and Targets" by Hui Zhao, Peng Wang, Cheng Lv, and Conghui Zhang proposes a method for automatically extracting opinion words and targets from text. The authors use a bootstrapping-based framework that starts with a small set of opinion seed words and uses statistical word co-occurrence and dependency patterns to propagate between opinion words and targets. They also introduce an Automatic Rule Refinement process to improve the accuracy of the initial rules. The authors evaluated their approach and found that it achieved higher accuracy in identifying opinion words and targets compared to baseline methods. They suggest that their method can be applied to tasks such as product reviews and social media analysis to better understand consumer opinions.

The paper "Extracting Opinion Targets and Opinion Words from Online Review with Graph Co-ranking" by Linqing Liu, Xiaohua Xu, and Jun Zhao focuses on fine-grained opinion mining, specifically the extraction of opinion targets and opinion words from online reviews.

The authors propose a novel approach based on the partially-supervised alignment model, which uses a graph-based co-ranking algorithm to estimate the confidence of candidate opinion targets and words. Their model captures opinion relations more precisely, particularly for long-span relations, and achieves better precision compared to unsupervised alignment models. The authors evaluated their approach on different corpora and found that it outperformed state-of-the-art methods. Their approach has the potential to be useful in sentiment analysis and opinion mining applications.

III. METHODOLOGY

The proposed methodology goes in the below mention step wise manner.

A. Data Collection

Text reviews: Valuable feedback on course content, instructor quality, course difficulty, and user experience.
Ratings: Quantitative measure of user satisfaction with a course.
Other data: User demographics, enrolment numbers, completion rates, and course content usage data.

B. Data Pre-processing

Cleaning and transforming raw data.
Data normalization: Standardizing the data format.
Tokenization: Breaking down text into individual words.
Stop word removal: Removing commonly occurring words with little meaning.
Other techniques: Stemming, lemmatization, URL and email address removal.

C. Sentiment Analysis:

Using sentiment analysis to identify emotional content in text data.
Lexicon-based tools (e.g., VADER) or machine learning algorithms (e.g., Naive Bayes, SVM, Neural Networks).
Classifying reviews as positive, negative, or neutral.
Assigning sentiment scores to indicate the degree of positivity or negativity.

D. Feature Extraction:

Extracting specific aspects of the course mentioned in reviews.
Identifying relevant keywords and phrases.
Techniques: Frequency analysis, NER, POS tagging.
Feature-based sentiment analysis to analyse sentiment toward specific features.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

E. Analysis and Visualization:

Analysing sentiment and features of the courses.
Visualizing results with pie charts, bar graphs, word clouds, heat maps.
Reports and dashboards to provide comprehensive overviews.

F. Recommendations

Personalized course recommendations based on user's profile, course history, and sentiment/feature analysis.
Comparing user data with sentiment and feature data to identify relevant courses.
Machine learning algorithms or collaborative filtering for accurate recommendations.
Improving user experience, satisfaction, engagement, learning outcomes, and retention rates.

G. VADER

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is widely used in natural language processing (NLP) applications. It was developed by researchers at Georgia Tech and is designed to analyze the sentiment of text in a way that is sensitive to the nuances of human language.

One of the key features of VADER is its ability to handle complex linguistic features, such as sarcasm, irony, and emoticons. It uses a lexicon of words that are assigned scores based on their sentiment, with positive scores indicating positive sentiment and negative scores indicating negative sentiment. VADER also takes into account the intensity of the sentiment, as well as the presence of negation words and modifiers that can change the sentiment of a sentence. In addition to its lexicon-based approach, VADER also uses a set of rules to handle specific language patterns that can affect sentiment analysis. For example, it can identify and handle the difference between a statement that uses a comparative word (e.g., "better") versus a superlative word (e.g., "best"). VADER has been widely used in various NLP applications, such as social media analysis, customer feedback analysis, and market research. Its accuracy and robustness have been evaluated in several studies and it has been shown to outperform other sentiment analysis tools in many cases. Overall, VADER is a powerful and flexible tool for sentiment analysis that is well-suited for analysing text data that contains complex linguistic features. Its lexicon-based approach and rule- based processing make it a valuable tool for researchers, businesses, and other organizations that need to analyse the sentiment of large volumes of text data.

In this project, we propose an innovative approach to analyze MOOC datasets using text mining techniques. Our approach encompasses several key steps, including exploratory data analysis, sentiment analysis to provide a comprehensive evaluation of online courses and course recommendations. One of the main objectives of our proposed system is to assess the quality of online courses by analyzing learners' comments using VADER, which is a domain-specific lexicon-based sentiment analysis tool. By leveraging VADER, we can accurately measure the sentiment expressed in learners' comments, such as positive, negative, or neutral, and gain insights into their overall satisfaction with the course content, instructors, and learning experience. Furthermore, our approach involves text mining techniques to analyze MOOC data from the large volume of course review data, we aim to identify areas that need improvement, optimize course content and delivery, and enhance learning outcomes for students

Moreover, our proposed system aims to assist instructors in developing effective teaching strategies. By analyzing learners' comments and feedback, instructors can gain valuable insights into the strengths and weaknesses of their courses, identify areas for improvement, and make data-driven decisions to enhance the overall quality of their courses.

The proposed system in this project offers several advantages. Firstly, it provides personalized learning experiences to students by leveraging text mining techniques. Through analysis of learners' comments and feedback, the system can identify individual preferences, interests, and areas of improvement, and tailor course recommendations or interventions accordingly. This personalized approach can enhance student engagement, motivation, and satisfaction, ultimately leading to improved learning outcomes.

Secondly, the proposed system utilizes VADER, a domain-specific lexicon-based sentiment analysis tool, for accurate sentiment analysis of learners' comments. This enables a more reliable assessment of the quality of online courses and instructors' performance by identifying positive, negative, or neutral sentiments expressed in the feedback.

Thirdly, the proposed system facilitates data-driven decision making by employing exploratory data analysis and text mining techniques. By extracting meaningful patterns and trends from MOOC data, instructors and administrators can gain valuable insights for making informed decisions to optimize course content, delivery, and overall quality. This data-driven approach can lead to more effective course design and development, resulting in improved learning experiences for students. Furthermore, the proposed system provides support for instructors in developing effective teaching strategies.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

By analysing learners' comments and feedback, instructors can gain insights into the strengths and weaknesses of their courses, understand learners' needs and expectations, and make necessary adjustments to improve the quality of their teaching. This instructor-centric approach can enhance communication and engagement between instructors and students, fostering a more effective learning environment.

Lastly, the proposed system offers an enhanced course evaluation approach by incorporating sentiment analysis, exploratory data analysis, and text mining techniques. This comprehensive evaluation can provide a deeper understanding of the overall quality of courses, including strengths, weaknesses, and areas for improvement. Such insights can inform decision making for course design, development, and delivery, leading to continuous improvement of online courses.



Fig 1 Word Cloud

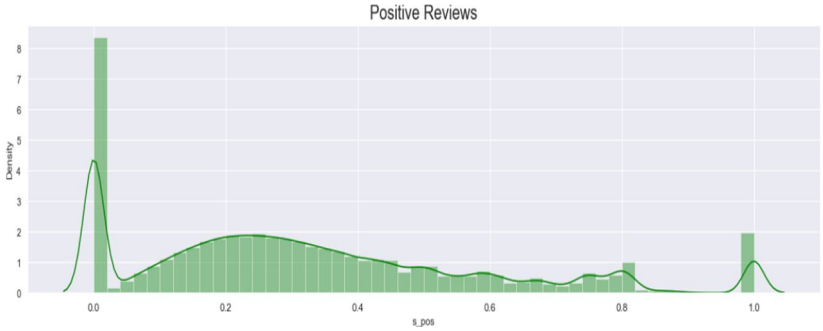


Fig 2 Positive reviews

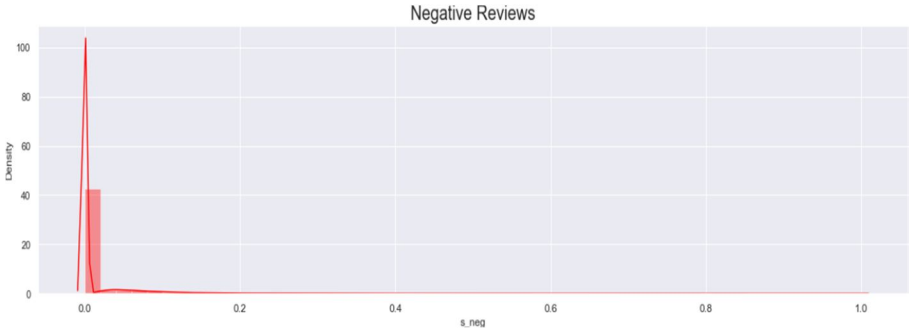


Fig 3 Negative reviews

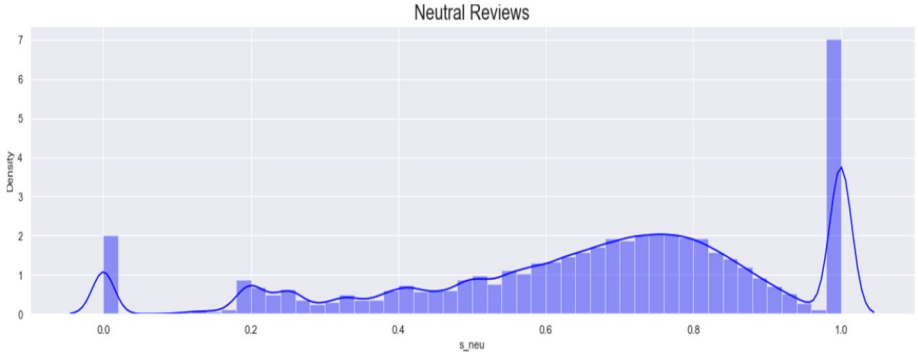


Fig 4 Neutral reviews

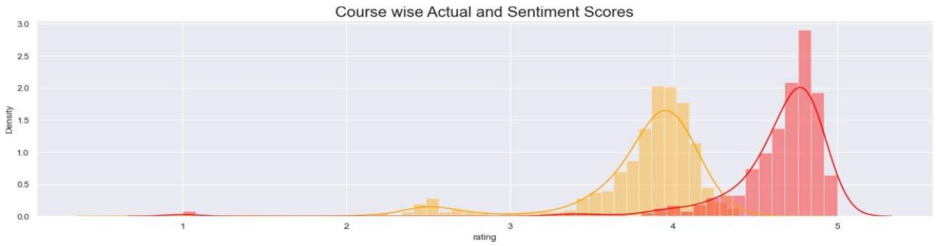


Fig 5 Course wise actual and sentiment scores

The above results show that with the help of text mining and techniques in natural language processing we can find the difference with actual reviews that are numeric in nature and sentimental scores that reveal the real essence of the review whether positive, negative or neutral.

IV. CONCLUSIONS

In conclusion, our project aims to support learners in their course selection process through large-scale analysis of open MOOC reviews using text mining and sentiment analysis techniques. We have developed a system that utilizes advanced technologies like BERT, VADER, and Python programming language to process and analyze massive amounts of data from MOOC reviews. By extracting valuable insights and providing course recommendations based on sentiment analysis, our system has the potential to enhance the decision-making process for learners, helping them make informed choices about which courses to enroll in.

REFERENCES

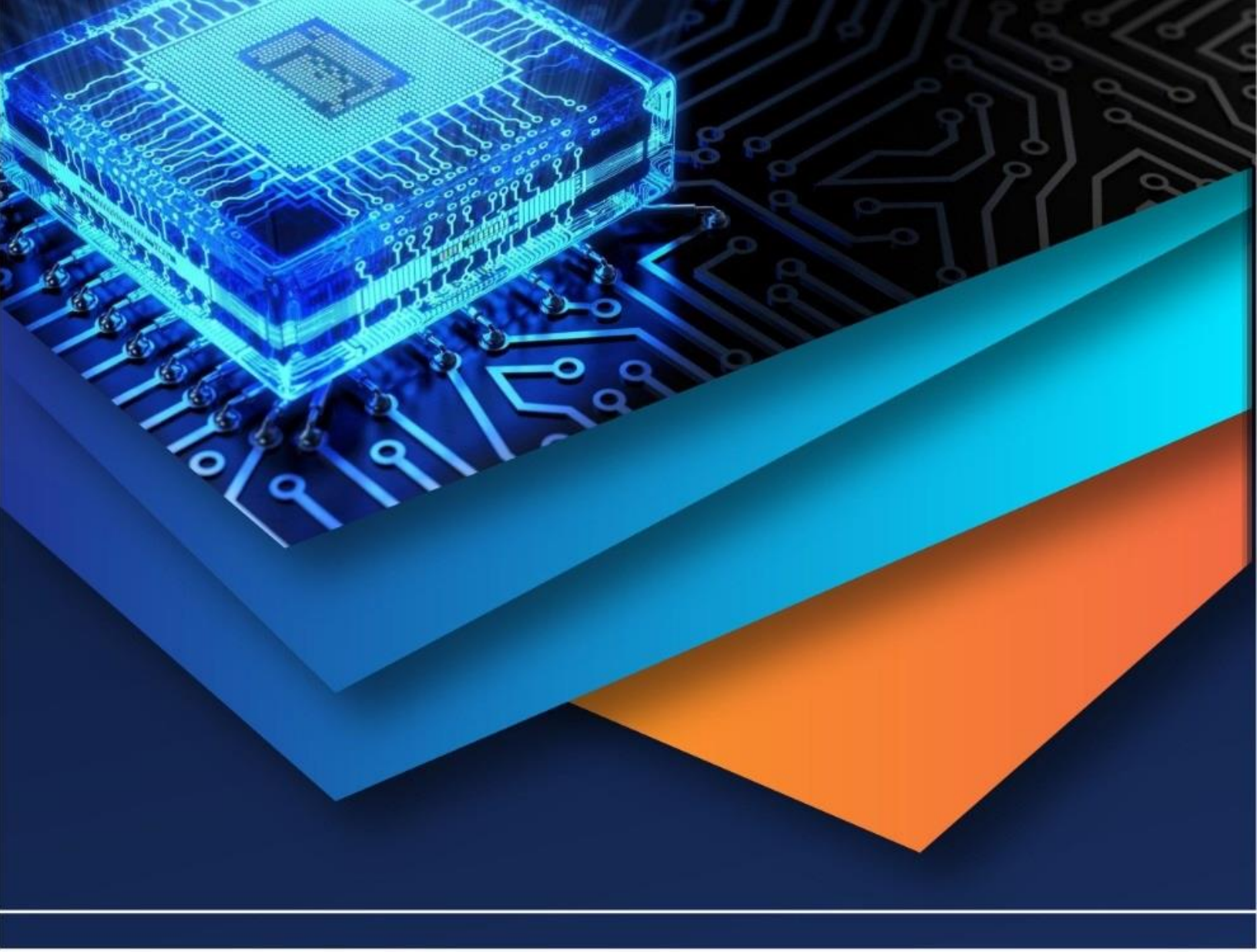
[1] M.Munezero, C.S. Montero, M. Mozgovoy, and E. Sutinen, “Exploiting sentiment analysis to track emotions in students' learning diaries,” Association for Computing Machinery, New York, NY, USA, 145–152. In Proceedings of the 13th Koli Calling International Conference on Computing Education Research (Koli Calling '13),2013.

[2] Q.Zhao, H.Wang,P.Lv, and C. Zhang, “A Bootstrapping Based Refinement Framework for Mining Opinion Words and Targets,” Acm International Conference. ACM,2014.

[3] J. Devlin, M. Chang, W. M and K. Lee, “Bert: Pre-training of deep bidirectional transformers for language understanding,” arXiv preprint, vol. 10, no. 04805, 2018.

[4] K Liu L Xu J Zhao Coextracting opinion targets and opinion words from Online Reviews Based on the Word Alignment Model,” Knowledge and Data Engineering, IEEE Transactions on, 27(3):636- 650,2015.

[5] https://www.javatpoint.com/text-data-mining



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)