Expert Systems With Applications 210 (2022) 118400

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Large scale analysis of open MOOC reviews to support learners' course selection☆

Manuel J. Gomez [a,*], Mario Calderón [b], Victor Sánchez [b], Félix J. García Clemente [a], José A. Ruipérez-Valiente [a]

[a] University of Murcia, Calle Campus Universitario, 30100, Murcia, Spain
[b] SkillMapper, 6 rue de Steinkerque, 75018, Paris, France

ARTICLE INFO

Keywords:
Massive Open Online Courses
Natural language processing
Sentiment analysis
Recommendation systems
Online education

ABSTRACT

The recent pandemic has changed the way we see education. During recent years, Massive Open Online Course (MOOC) providers, such as Coursera or edX, are reporting millions of new users signing up on their platforms. Though online review systems are standard among many verticals, no standardized or fully decentralized review systems exist in the MOOC ecosystem. In this vein, we believe that there is an opportunity to leverage available open MOOC reviews in order to build simpler and more transparent reviewing systems, allowing users to really identify the best courses out there. Specifically, in our research we analyze 2.4 million reviews (which is the largest MOOC reviews dataset used until now) from five different platforms in order to determine the following: (1) if the numeric ratings provide discriminant information to learners, (2) if NLP-driven sentiment analysis on textual reviews could provide valuable information to learners, (3) if we can leverage NLP-driven topic finding techniques to infer themes that could be important for learners, and (4) if we can use these models to effectively characterize MOOCs based on the open reviews. Results show that numeric ratings are clearly biased (63% of them are 5-star ratings), and the topic modeling reveals some interesting topics related with course advertisements, the real applicability, or the difficulty of the different courses.

1. Introduction

Massive Open Online Courses (MOOCs) are a relatively recent educational phenomenon that has received a great deal of attention during the last decade (Pina & Steffens, 2015). This attention is due to their potential to disrupt the traditional educational pathways through ease of access and free or low-cost contents (Castillo et al., 2015). Furthermore, the recent COVID-19 pandemic has completely shifted the behavior of students around the world (Dhawan, 2020). Strict lockdowns and a growing fear of returning to classrooms were the perfect storm for the unparalleled acceleration in the adoption of MOOCs. Thus, it is not surprising that the largest providers of MOOCs such as Coursera or Udemy reported a +50% growth in revenue in 2020 (Lohr, 2020), the most significant year-over-year increase since the beginning of the MOOC era. Taking advantage of the momentum, ed-tech providers attempted to capitalize by widening their offer of courses (Shah, 2021).

Though a sudden increase in the educational offer seems like an idyllic scenario as it provides a broader range of educational options, it is yet to be confirmed whether the aforementioned is net positive for the system as a whole. As a result, we might be reaching what some scholars call "choice overload", as a higher number of options might decrease the motivation to choose or the satisfaction with the final preferred option (Scheibehenne et al., 2010). Arguably, this lack of satisfaction with the preferred option contributes to low completion rates in the MOOC space, which are reported to be around 10% (Khalil & Ebner, 2014). Moreover, Hew and Cheung (2014) reported that the quality of MOOC education and the MOOC business model are still unresolved issues. In addition, more than 55% of the students who were part of a research conducted in LATAM reported spending days and weeks choosing a course (SkillMapper, 2022); thus, shedding light on how time-consuming the process of finding and purchasing courses could be. A process that ultimately fails to deliver on the promise of democratizing education.

In addition to the trend above, students do face other challenges when choosing courses. Though online review systems are standard among many verticals, no standardized or fully decentralized review systems exist in the MOOC ecosystem. Whereas some MOOC providers

☆ Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
* Corresponding author.
E-mail addresses: manueljesus.gomezm@um.es (M.J. Gomez), mario@skillmapper.com (M. Calderón), victor@skillmapper.com (V. Sánchez), fgarcia@um.es (F.J.G. Clemente), jruiperez@um.es (J.A. Ruipérez-Valiente).

https://doi.org/10.1016/j.eswa.2022.118400
Received 25 January 2022; Received in revised form 29 July 2022; Accepted 3 August 2022
Available online 6 August 2022

rely on the standard 5-star rating system and free form feedback, such as Class Central (Central, 2021), other popular ones do not even offer the alternative to leave a single review. Moreover, many of these MOOCs only allow the user to leave a review once the course has been completed (Gamage et al., 2016). Thus, the opinions of around the 90% of students who never finished the MOOC are missing, hence usually creating a very positive and skewed perspective. In this vein, NCA (research agency) found that around 18% of students that do not finish a MOOC experienced teachers with low pedagogical qualities (SkillMapper, 2022). Naturally, the restrictiveness of many of these review systems limits the voice of many students, thus suppressing the distribution of these kinds of relevant insights for MOOC purchase decision making. This transparency in the reviewing systems could allow new approaches to facilitate students' choices when comparing different courses.

The inconsistent review systems, high search costs, and meager completion rates contribute to many MOOC providers' poor experience. Current textual reviews are considered as additional information to the rating, to know how the course is, but it is not considered in the rating of the review itself and in personalized recommendations (Kapoor et al., 2020). We believe that there is an opportunity to leverage available open MOOC reviews, in order to build simpler and more transparent reviewing systems, allowing users to really identify the best courses out there. In fact, Natural Language Processing (NLP) provides us a set of techniques and approaches to analyze textual information (e.g., sentiment analysis, topic modeling, text classification) that could be used to add value to the reviews and provide useful insights for learners. Ultimately, there is an opportunity to provide a framework that could benefit the overall MOOC industry. Thus, enable more transparency, better discoverability, and increased satisfaction with choice, also helping to greatly improve the professional development of our society to current industry needs. As a by-product, these enhancements could also lead to improved overall satisfaction with MOOC providers. In our work, we aimed to collect the largest dataset of MOOC reviews from five different platforms via web scraping, in order to analyze around 2.4 million reviews and to transform these big data into useful insights and actionable information for learners and other stakeholders, such as industry and educators. Specifically, we state four research questions (RQs):

- **RQ1.** Do the crowdsourced numeric ratings of the courses provide discriminant information to prospective learners?
- **RQ2.** Does NLP-driven sentiment analysis on the textual reviews provide valuable information to prospective learners?
- **RQ3.** Can we leverage NLP-driven topic analysis techniques to find themes that can be important for prospective learners?
  - **RQ3.1.** Using words describing the course in a qualitative way.
  - **RQ3.2.** Using words describing the courses' content.
- **RQ4.** Can we use these models to effectively characterize MOOCs based on the open reviews?

The rest of the paper is organized as follows. Section 2 reviews background literature on MOOCs, previous course selection approaches, and previous analyses using MOOCs reviews. Section 3 describes the complete methodology, and Section 4 shows the results of our analysis, answering each one of the RQs. Then, we finalize the paper with discussion in Section 5 and conclusions and future work in Section 6.

2. Related work

Technological advances, particularly in the Web 2.0, have brought a major transformation in the delivery of education, and one of the most recent innovations in e-learning is the development, roll out and uptake of MOOCs (Terras & Ramsay, 2015). MOOCs represent the next stage in the evolution of open educational resources (Al-Rahmi et al., 2018). In addition, they have the advantage of being available for all and open to an unlimited number of students (Al-Rahmi et al., 2019). The increase in numbers of MOOCs has been dramatic in recent years, and their scope, available in every sector of education, makes them very strong and being considered to be a new form of virtual technology enhanced learning environments (Pina & Steffens, 2015). In 2013, Billington and Fronmueller (2013) stated that some experts considered MOOCs as a bubble that would burst soon. However, by the end of 2020, more than 180 million students have registered in one or more of the 16,300 courses offered by over 950 universities worldwide (Shah, 2021), indicating that MOOCs continued to capture the attention of many institutions and the public worldwide since the "year of the MOOC" in 2012 (Lan & Hew, 2020; Pappano, 2012). Moreover, they have overcome all the challenges of the COVID-19 pandemic and have been instrumental in providing courses to the learners as they do not have physical boundaries (Purkayastha & Sinha, 2021).

Although MOOCs have a great potential for educating large numbers of students in a wide variety of settings, previous research has also identified some problematic issues with this area. One of the major problems is the high dropout rate for learners: a small percentage (generally around 10%) of the large numbers of participants enrolling in MOOCs manage to complete the course, which is a poor result compared to traditional courses (Liyanagunawardena et al., 2014). Bezerra and Silva (2017) conducted a review on the different reasons causing these high dropout rates, identifying 24 different reasons, such as the lack of prior knowledge, the difference between the expected and the real level of the course, or the quality of materials. We see another example in Jobe et al. (2014), where authors reported challenges associated with recognition, validation, and accreditation of learning. Of course, another area of major concern is academic integrity, including verifying that a student who is registered for a course is the same person who completed the tasks, and the potential for widespread plagiarism or homework assignments. Finally, another common problem that usually is not discussed in the literature is the course selection. Given the large numbers of courses (most of them with very high ratings) and information out there, learners are easily disoriented by the information overload (Zhang et al., 2018), and they usually find difficulties choosing which is the appropriate course for them (taking several days to decide); this course selection difficulty is also related to the high dropout rate as they may not be choosing the best course to fit their preferences. Some of the reasons of the high dropout rate reported previously could be related with the overload and misunderstandings that learners suffer, since they do not usually know the real characteristics of the course that they are choosing: in a survey conducted by Gütl et al. (2014), 8.96% of the students reported that the course was too difficult, meanwhile 7.46% emphasized that the course was not challenging. Moreover, 14.93% of students also highlighted that the learning environment was not personalized, and another 6.72% indicated that the courses were poorly taught. The selection of courses is a crucial decision for learners, and the use of existing reviews from other learners to reveal the real courses' characteristics could be a interesting solution to make this choice easier and fruitful.

Learners are exposed to various challenges with this excess of learning resources and information: which provider they have to choose to search for a specific MOOC? Who is the best provider? Are the available MOOC reviews reliable? (Ouertani & Alawadh, 2017). Usually, in literature we find different approaches trying to perform an efficient course selection: collaborative filtering (according to users in same interests on same courses), content-based filtering (using courses features from courses that the user likes to recommend similar items), knowledge-based approaches, or hybrid recommender approaches, among others (Al-Badarenah & Alsakran, 2016). Ouertani and Alawadh (2017) presented the idea of a system that could satisfy learners' needs when searching for suitable courses among different providers, using previous experiences of its users. Moreover, Al-Badarenah and Alsakran (2016)

developed a recommendation system that employed an association rules algorithm to recommend university elective courses to a target student based on what other similar learners have taken. We found many other studies that used these type of approaches to improve course selection (Aher, 2014; Aher & Lobo, 2012; Huang et al., 2019; Vanitha et al., 2019). In fact, in their review, Khalid et al. (2020) reported that most of the work on the implementation of the recommender system for courses used collaborative and content-based filtering. In our research, we decided to use a novel approach that aims to leverage existing open MOOC reviews using NLP to gather information that could be useful for learners' course selection.

Moreover, as we do in our case study, previous researchers have explored the potential of using open MOOC reviews and NLP. For example, authors in Deng and Benckendorff (2021) reported that it is critical that student reviews are systematically analyzed and interpreted in conjunction with student ratings data. In their work, they presented a study using *Leximancer*, a data-mining tool which extracted the key concepts from the data collection, based on the frequency of occurrence of concepts and co-occurrence use. This study highlighted some interesting key terms inferred, such as "videos", "engaging", "understanding", "easy", or "problems". Furthermore, Wen et al. (2014) applied sentiment analysis toward the curriculum and key course tools from forum posts, finding that the emotion of students was correlated with the number of students dropping out of the course each day. Moreover, Chen et al. (2020) employed an innovative structural topic modeling technique to analyze 1,920 reviews of 339 courses regarding computer science to understand what primary concerns the learners had. They found that 64.2% of the reviews in their data collection were 5-star reviews, while 16.7% were 4-star reviews, revealing clearly biased rating values. Regarding structural topic modeling, they found topics such as "Course levels", "Teaching style", or "Course content". Furthermore, Chang et al. (2016) presented a novel approach that used the courses' videos to locate "hot" video segments (i.e., the parts that students reviewed more) to generate their subtitles and extract the main keywords, being useful for the teacher to better understand which parts of the contents of each topic are most difficult for the learners. Our main contributions to existing literature are:

- We use the largest MOOC review dataset to date in literature.
- We analyze numeric ratings' distribution to identify possible bias.
- We build two different NLP models (topic finding and sentiment analysis), oriented to provide help and support learners' course selection easily and quickly, and not only extracting terms or revealing important topics within a word collection.
- The topic modeling approach considers two different aspects of the reviews: on the one side, it considers words in reviews describing the course in a qualitative way; on the other side, it considers words in reviews describing the courses' content. This aims to gather insights from two different points of view, also reducing the bias produced by other types of words.

3. Methodology

We conducted our analysis in six different stages: (1) MOOC review providers, (2) Data scraping, (3) Data cleansing and wrangling, (4) Data collection, (5) Data pre-processing, and (6) Data modeling. We can see the entire methodology process represented in Fig. 1. Next, we explain each stage in detail:

3.1. Selection of MOOC review providers

Initially, we considered five different MOOC review providers for our analysis: Udemy, Coursera, Domestika, Platzi, and Crehana. Next, we make a brief introduction to each platform:

- **Udemy** is a for-profit MOOC provider that offers more than 183,000 courses taught by 65,000 different instructors and with 75 different languages available (Udemy, 2021).
- **Coursera** is an American MOOC provider. In 2021, 150 universities offered more than 4,000 courses through Coursera, reaching more than 190 different countries (Coursera, 2021a, 2021b).
- **Domestika** is an online learning platform for creatives, with paid courses on software, crafts, illustration, and more. Its major feature is the exceptionally high course quality (Rowan, 2021).
- **Platzi** is a MOOC provider founded in Colombia, offering more than 700 open courses within different areas, such as data science, videogames, or blockchain (Platzi, 2021).
- **Crehana** is an online education provider for today's student which offers subjects such as illustration, design, or business, bringing education closer to all types of students (Dañino, 2021).

3.2. Data scraping

Depending on how websites were structured, we used two different approaches: to work with websites using static pages, we used *Scrapy*, a web framework designed to extract massive amounts of data in a robust and efficient manner (Kouzis-Loukas, 2016); to work with more dynamic websites (with click buttons or infinite scrolls), we used *Selenium*, a portable open source software (Bruns et al., 2009). Using these tools, we started scraping the existing reviews in each one of the platforms. We worked with the metadata divided into two different levels:

- **Review level**. For each review, we collected the following information: URL, review, rating of the review (from one to five), the platform, the username, the date of the review, and the identifier of the related course.
- **Course level**. For each course, we collected the following information: the course identifier, URL, title, platform related with the course, the content category of the course, and the name of the teacher.

First, we downloaded all the data crawled and stored it directly in Firestore, a non-relational database from Google Cloud. For storing the data, we used a dictionary structure adapted to each platform.

3.3. Data cleansing and wrangling

Next we proceeded to apply a step wise approach to clean and process the data. In this part of the process (ran in Python), we use a series of rules that mostly leverage a set of regular expressions and variable conversions. These functions allow us to remove elements that do not add value, such as commas or special characters, and normalize types of variables when necessary.

Since we needed to identify the language of each review, and this information is not available in the original metadata, we identified it using the raw text of each review. Taking into account that we cannot rely on the audio feature to ensure that all the reviews are in the same language as the course, we finally decided to work with *Fasttext* library (Bhattacharjee, 2018) to infer the language of each review.

The last step was to store the data in a PostgreSQL database, a relational database that will help us to understand the data collected running some analytics easily. The objects that we built are inserted in each of the two data tables that we consider for this study: **final_courses_info**, which contains all the metadata at course level, and **final_reviews_info**, which contains all the metadata at review level.

3.4. Data collection

We initially considered all the data scraped from the five different platforms to collect the metadata. However, since for this study we decided to focus only on English reviews, the data collection was distributed as follows:

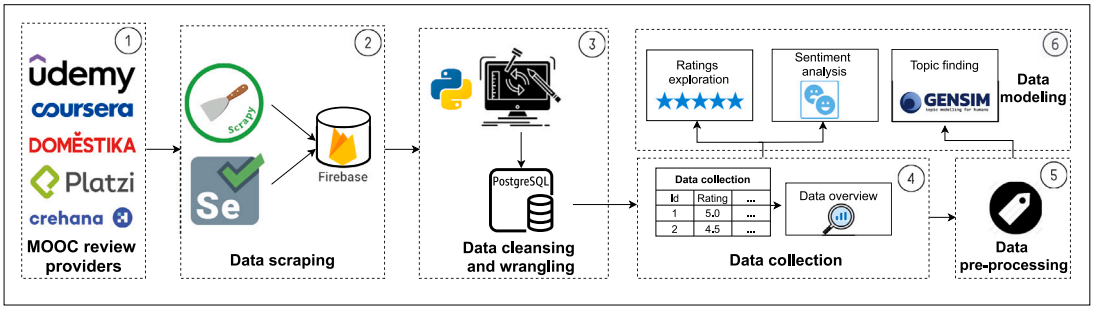
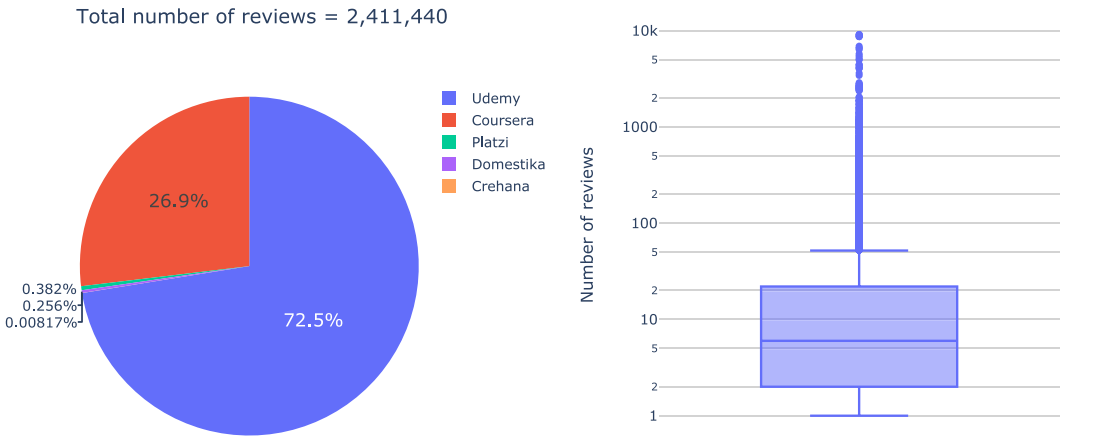


Fig. 1. Methodology diagram.



(a) Distribution of reviews by course provider.

(b) Distribution of number of reviews per course.

Fig. 2. Data collection overview.

- *Udemy* platform: It has 95,143 courses and 71,518 of them have reviews, meaning that 24.8% of them do not have any review.
- *Coursera* platform: It considers 4,154 courses and 2,932 of them have reviews, so 29.4% of courses do not have any review.
- *Domestika* platform: It has 558 courses in English and 541 of them have reviews, meaning that 3% of the courses do not have any review.
- *Platzi* platform: It has 451 courses in English and 450 of them have reviews, meaning that 0.002% of the courses do not have any review.
- *Crehana* platform: It has 1 course with reviews (100%).

We established a threshold of one minimum review for each course to be considered for the analysis, meaning that 25% of the courses were removed from our work. In our data collection, some of the reviews are concentrated on some courses. This can be related to several factors, such as important differences in popularity of each one of the courses, marketing efforts from the platforms, or hot topics. For example, a "Roman history" MOOC will attract a very different population of learners in quantity and background than a "Python" course. The distribution of reviews by the different providers can be seen in Fig. 2(a). Furthermore, in Fig. 2(b) we can see the distribution of the number of reviews per course. The maximum number of reviews in a single one is 9,149, the minimum is one review, and the median is six reviews per course.

3.5. Data pre-processing

When trying to leverage NLP-driven topic finding in such a large set of reviews, we found that a large proportion of the words do not provide useful information to learners, such as "lot", "give", or "thing", among many others. Thus we collected the most frequent words in the collection (words that appear more than 500 times in the entire collection), and we manually classified every word within two different categories:

- `QualitativeDescription`: if the word is related to a qualitative description of the course (e.g., easy, clear, practical). This category includes 357 different words.
- `Content`: if the word is related to the content of the course itself (e.g., machine, yoga, cooking). This category includes 759 different words.

This categorization will help us to create the two different topic finding models in our research. Every word not matching any of those two categories will be excluded from our later analysis. The next step in the pre-processing is to clean each textual review, keeping only the review's main words and removing, for example, unnecessary URLs, numbers, or additional space characters. Moreover, to apply NLP techniques afterward, we need to define a set of "stop words" (i.e., words that will not be considered in the text analysis). Some examples of stop words are "would", "be", or "however", which are common words that appear in every document but do not provide helpful information to the analysis. We can see this process represented in Fig. 3.

From now, we can start treating each review as a "document", since most of the cleaning process has ended. Once the full review collection is cleaned, we lemmatized every document using *pywsd* library (lemmatization is the process of converting a word to its base form).

3.6. Data modeling

This subsection is divided into four separate parts, each one explaining the analysis to address each RQ:

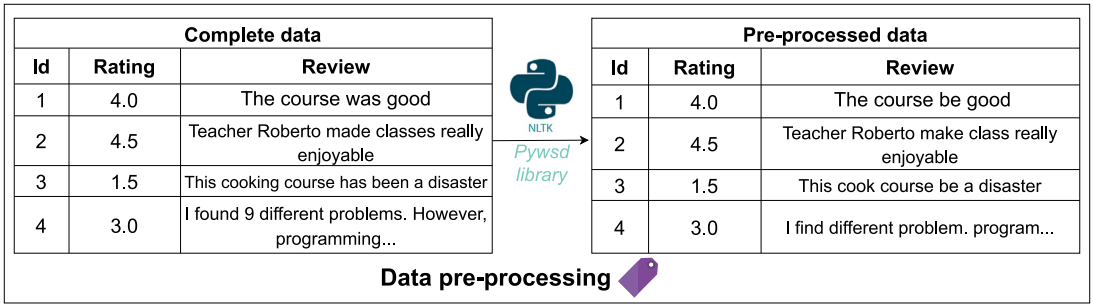


Fig. 3. Data pre-processing using Python NLTK and Pywsd library.

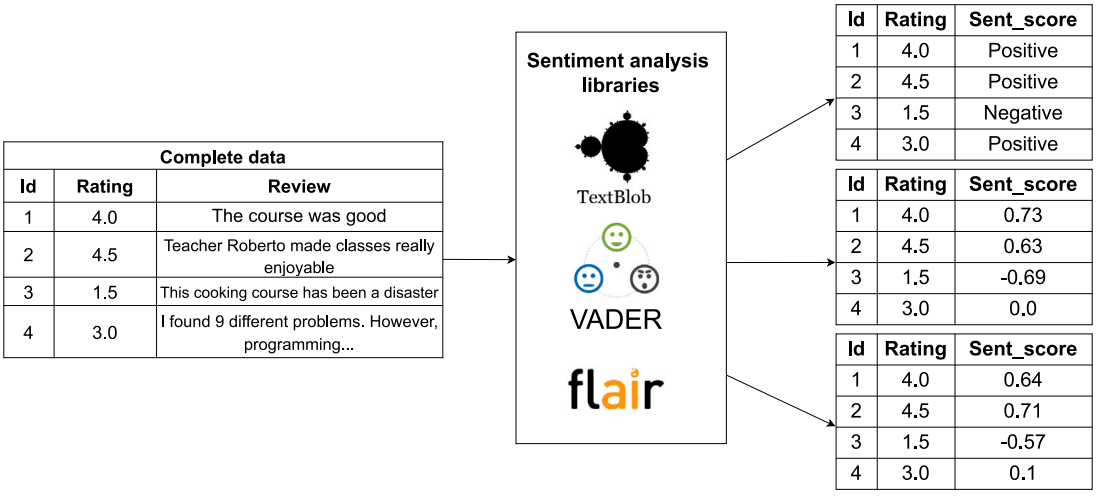


Fig. 4. Sentiment analysis methodology process.

3.6.1. Ratings exploration

To discover how was the distribution of ratings within the data collection, we checked two different approaches: in the first approach, we simply count every single numeric rating to see which are the most common ones; in the second approach, we calculated the mean rating of each one of the courses in order to see the distribution of ratings within the different courses separately.

3.6.2. Sentiment analysis

The main aim of sentiment analysis (or opinion mining) is to discover emotion, opinion, subjectivity and attitude from a natural text. There are many different techniques that are frequently applied to natural text to determine the sentiment such as feature extraction, emoticon study, tokenization etc. In our work, we explored the use of three different libraries in *Python*, namely *TextBlob*, *VADER*, and *Flair*:

- *TextBlob* is a Python library for processing textual data (Loria, 2018). It returns a score within the range [-1.0, 1.0], depending on how positive or how negative that sentence is.
- *VADER*: Valence Aware Dictionary for Sentiment Reasoning (VADER) is specifically attuned to the sentiments expressed in social media (Elbagir & Yang, 2019). Given its implementation, we used the "compound" value, which is a number within the range [-1.0, 1.0].
- *Flair* is a framework which provides a pre-trained model, which returns a tag ("Positive" or "Negative") depending on the sentiment of the text.

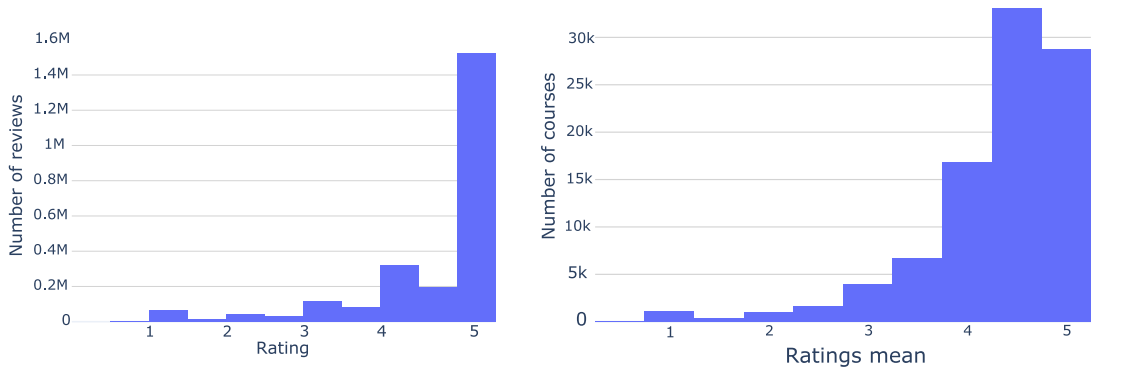
This sentiment analysis process is represented in Fig. 4.

3.6.3. Topic finding

To discover which are the main topics within the review collection, we applied Latent Dirichlet Allocation (LDA) topic modeling to the data provided. Specifically, we use *gensim* library (Saxton, 2018) and its *ldaMallet* model (Graham et al., 2012). Generally in LDA, each document can be described by a distribution of topics, and each topic can be described by a distribution of words; *ldaMallet* uses an optimized Gibbs sampling algorithm for LDA (Yao et al., 2009).

There are multiple metrics for evaluating the optimal number of topics. Recent studies have shown that the classic predictive likelihood metric (or equivalently, perplexity) and human judgment are often not correlated, and even sometimes slightly anti-correlated (O'callaghan et al., 2015). This has led to many studies that have focused upon the development of topic coherence measures. To determine the optimal number of topics, we used two of these coherence measures: C_v (combining the indirect cosine measure with the normalized pointwise mutual information and the boolean sliding window) and C_{umass} (based on document co-occurrence counts, a one-preceding segmentation and a logarithmic conditional probability as confirmation measure) (Kapadia, 2019; Röder et al., 2015).

Based on these two coherence measures, we chose the number of topics of the two final models calculated. The first of our models aimed to identify topics that described the course in a qualitative way (e.g., easy, clear, detailed, hard), thus we selected, for each review, only words that were classified as `QualitativeDescription`, removing the rest of the review to avoid bias produced by other type of words. Then, the second of our models aimed to identify topics that described the content of the courses, thus we selected only words that were previously classified as `Content`. Moreover, to perform the analysis,



(a) Numeric ratings distribution. (b) Distribution of courses based on their ratings mean.

Fig. 5. Raw ratings and course rating mean distribution.

we assumed that all the reviews have the same weight (i.e., all reviews are equally important).

Then, based on the topics discovered by the topic finding algorithm, we calculated the proportion of each topic across the entire corpus. To do that, we evaluated each review to get its topics associated (note that, in LDA, each document can be assigned to several topics with a certain weight). We calculated the proportion of each topic as follows:

$$Proportion\_topic_j = \frac{\sum_{i=1}^{N} weight\_topic_{ij}}{N} * 100 \quad (1)$$

Therefore, the proportion of topic j would be the summation of each weight assigned to the topic j in each document from i to N, divided by the number of documents in the corpus (N).

To provide a more complete analysis, we wanted to see the relationship between the sentiment analysis and the qualitative topic model (e.g. see if a course with a bad mean rating also is related to more negative topics). That way, we calculated the distribution of topics in each course separately (the distribution in each course will be the sum of proportions of each review separately), and compared them to the positive or negative values given by the sentiment classifier. A given course will be labeled as `Positive` if the majority of its reviews are labeled as `Positive`, and `Negative` if the majority of its reviews are labeled as `Negative`.

4. Results

4.1. RQ1. Do the crowdsourced numeric ratings of the courses provide discriminant information to prospective learners?

In Fig. 5(a) we can see the distribution of numeric ratings available in our collection. About 1.52 million reviews (63% of the total) are 5-star ratings, and also about 519,000 reviews (21.5% of the total) are 4 and 4.5 star ratings, thus we have a clearly biased data, where only 17.5% of the ratings are 3.5-star ratings or worse. Moreover, in Fig. 5(b) we see that, out of 93,678 total courses, 48,640 (52%) have a mean rating between 4.5 and 5 stars, and 34,770 (37%) courses have a mean higher of equal than 3.5 and below 4.5, highlighted again the systematic biases existing in open MOOC reviews. Therefore, this motivates alternative models to facilitate course selection to learners'.

4.2. RQ2. Does NLP-driven sentiment analysis on the textual reviews provide valuable information to prospective learners?

As we stated in Section 3.6.2, we performed sentiment analysis on textual reviews based on three different libraries: *TextBlob*, *VADER*, and *Flair*. Since *TextBlob* and *VADER* provide numeric values, we can compare both libraries directly. After applying the pre-trained sentiment models to every review in our corpus, we obtained a mean compound value of 0.54 (SD = 0.37) using *VADER*, and a mean compound value of 0.37 (SD = 0.30) using *TextBlob*. Then, we aggregated the sentiment values by course. On the one hand, we find that 67,019 courses (71.5%) have a mean compound value above 0.4 (high positive mean) using *VADER*. On the other hand, we find that 29,777 courses (31.8%) have a mean compound value above 0.4 using *TextBlob*.

We also calculated the Pearson's correlation coefficient and the Spearman's rank correlation coefficient between the compound sentiment value and the numeric ratings for both libraries. VADER library got a Pearson's coefficient value of 0.49, and a Spearman's coefficient value of 0.35. In the case of *TextBlob*, the Pearson's coefficient was 0.38, while the Spearman's coefficient was 0.31. Although all these coefficients show a positive correlation between the two variables, VADER library obtained a higher value, indicating a moderate level of correlation since it is below 0.5.

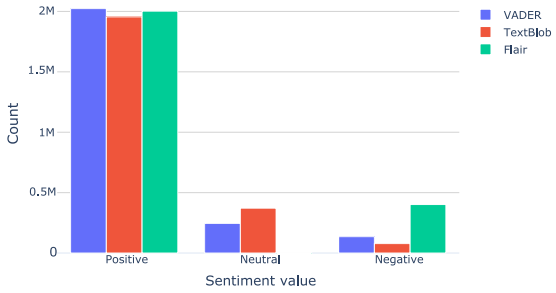
In the case of *Flair*, it provides a label indicating if the sentence is `Positive` or `Negative`. To compare the three libraries, we established a threshold to classify *TextBlob* and *VADER* compound values into one of the three (positive, negative and neutral) labels. That way, we classify a sentence into `Positive` when the compound value > 0.1, into `Neutral` when $-0.1 \leq compound\_value \leq 0.1$, and into `Negative` when the compound value < 0.1. In Fig. 6(a) we can see a comparison of the number of positive, neutral, and negative reviews classified by each one of the libraries. As we can observe, all three libraries classified the majority of the reviews as `Positive` (2M, 2.02M and 1.95M), indicating also a skewed data. However, we see a more significant difference if we take a look at the `Neutral` and `Negative` reviews. On the one hand, regarding *VADER*, we note that 245,538 reviews are `Neutral`, and 135,571 are `Negative`. On the other hand, *TextBlob* classifies 371,348 reviews as `Neutral`, and 78,198 as `Negative`. Finally, *Flair* library results show that 401,575 are `Negative` reviews, and only 516 reviews are *Neutral*. Furthermore, in Fig. 6(b) we can see a box plot comparing the sentiment values between the two libraries providing continuous values: *VADER* and *TextBlob*. As we observe, *VADER* library produces higher sentiment values (median= 0.62, $Q1 = 0.45$, $Q3 = 0.77$) than TextBlob (median= 0.35, $Q1 = 0.14$, $Q3 = 0.6$).

4.3. RQ3. Can we leverage NLP-driven topic analysis techniques to find themes that can be important for prospective learners?

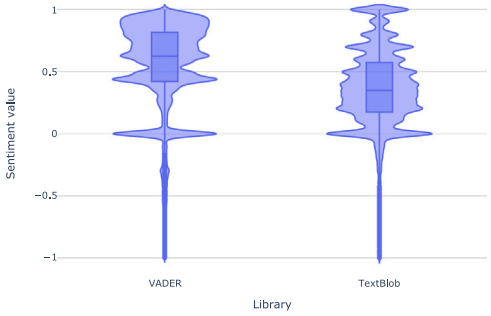
As described previously, we created two different topic finding models: the first one using words that describe the courses qualitatively; and the second one using words that described the courses' content.

4.3.1. Qualitative description model

The first model that we built was the qualitative description one. After applying the LDA algorithm as described in Section 3.6.3, we determined the optimal number of topics for this model taking into



(a) Categorized sentiment analysis comparison between libraries.



(b) Comparison of sentiment values between VADER and TextBlob library.

Fig. 6. Sentiment results.

Table 1
Summary of each detected topic regarding qualitative descriptions.

Topic	Main terms
1	Informative, easy, fun, enjoyable, helpful
2	Personal, authentic, productive, write, child, book
3	Basic, beginner, introduction, overview, simple
4	Simple, easy, effective, clear, short
5	Error, week, wrong, issue, outdated
6	Voice, fast, slow, hard, monotone
7	Money, worth, ad, free, quality
8	Specialization, team, rigorous, introduction
9	Step, explanation, detailed, easy, clear
10	Exam, test, practice, certification, real
11	Beginner, basic, advanced, intermediate
12	Real, worth, life, apply, practical
13	Theory, practical, theoretical, lab, hand
14	Slide, powerpoint, visual, lack, repetitive

account a range of topics between 2 and 25. We determined 14 as the optimal number of topics, with a C_v score of 0.20 and a C_{umass} score of −3.88. A summary of each topic, including the five most important words related, can be found in Table 1.

We can see a wide variety of topics, such as the first one, described by the words "informative", "easy", or "fun;" or the second one, described by the words "productive", "personal", or "authentic". Then, Fig. 7 shows the distribution of such topics across the entire collection of reviews. In the Figure, each topic is labeled by its three most important keywords (e.g., "simple_easy_effective"). As we can observe, the most frequent topics are "informative_easy_fun" (12.2%), "basic_beginner_introduction" (8.8%) and "personal_authentic_productive" (8.3%), and the less frequent topics are "real_worth_life" (5.3%) and "slide_powerpoint_visual" (5.0%).

To see the relationship between the qualitative topic model and the sentiment analysis, we calculated the distribution of topics in labeled courses based on their reviews' sentiment value. In this case, we used *Flair* library. Its sentiment classifier is based on a character-level LSTM neural network, and could also predict a sentiment for OOV (Out of Vocabulary) words including typos (Terry-Jack, 2019). In Fig. 8 we can see the distribution in both `Positive` and `Negative` labeled courses. As we can see, in the most frequent topics for positive courses, there are more positive topics such as "informative_easy_fun" (11.6%) and "simple_easy_effective" (8.6%). Then, in negative courses, the most frequent topics are the ones describing the course in a negative way, such as "voice_fast_slow"(11.7%) or "error_week_wrong" (10.4%). However, since most of the topics are positive, courses with negative sentiment values also show a high frequency of positive topics.

Then, we conducted a MANOVA test including the 15 different topics, confirming that there is a significant difference in topics proportion when comparing `Positive` and `Negative` courses ($F = 59.05$, $p <$ 0.0001). Thus, we can see that there is a relation between the sentiment analysis and the topic modeling approach that we have performed in this research.

4.3.2. Content model

Moreover, the second model that we built was the one describing courses' content. We determined 14 as the optimal number of topics, with a C_v score of 0.27 and a C_{umass} score of −5.08. In this case, we have assigned a label to each topic based on the existing content categories in the corpus and the keywords of each topic. A summary of each topic, including its name, description, and the five most important words related, can be found in Table 2.

Again, we see a large variety of topics referring to different content areas that courses aim to teach, such as "Programming", "Music", or "Language learning". Then, Fig. 9 shows the distribution of such topics across the entire collection of reviews. We see that there is a very uniform distribution of topics, since all of them have a similar value. Although there is a wide variety of knowledge areas among the different courses, there is only a 1.5% of difference between the most and less frequent topics.

4.4. RQ4. Can we use these models to effectively characterize MOOCs based on the open reviews?

To evaluate if we can effectively characterize MOOCs using the models developed in this research, we chose a set of four MOOCs belonging to different topics in order to analyze them:

- **Finance for Non-Finance Professionals**. The original category of this course is "Business" and has 342 available reviews.
- **Using Python to Access Web Data**. The original category of this course is "Computer science" and has 5,748 available reviews.
- **Mind Control: Managing Your Mental Health During COVID-19**. The original category of this course is "Health" and has 1,183 available reviews.
- **Teach English Now! Foundational Principles**. The original category of this course is "Social sciences" and has 2,447 available reviews.

As we can see in Fig. 10, all four courses have a rating mean above 4.7 stars, which confirms the biased data that we previously highlighted. Furthermore, we also see a positive correlation between the rating mean and the sentiment analysis results of the courses (calculated using *Flair* library), since the courses showing a higher rating mean also show a higher percentage of `Positive` labeled reviews. For example, the business related course has a rating mean of 4.83/5, and the 96.2% of its reviews are labeled as `Positive`, meanwhile the Python related course has a rating mean of 4.73/5, and the 88% of its reviews are labeled as `Positive`.

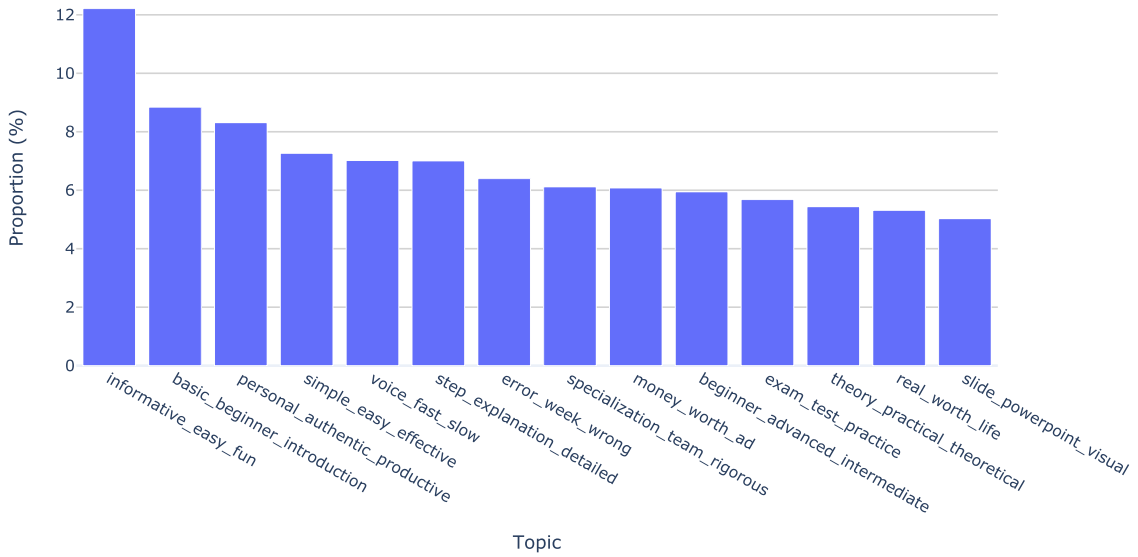


Fig. 7. Distribution of topics for the `QualitativeDescription` model.

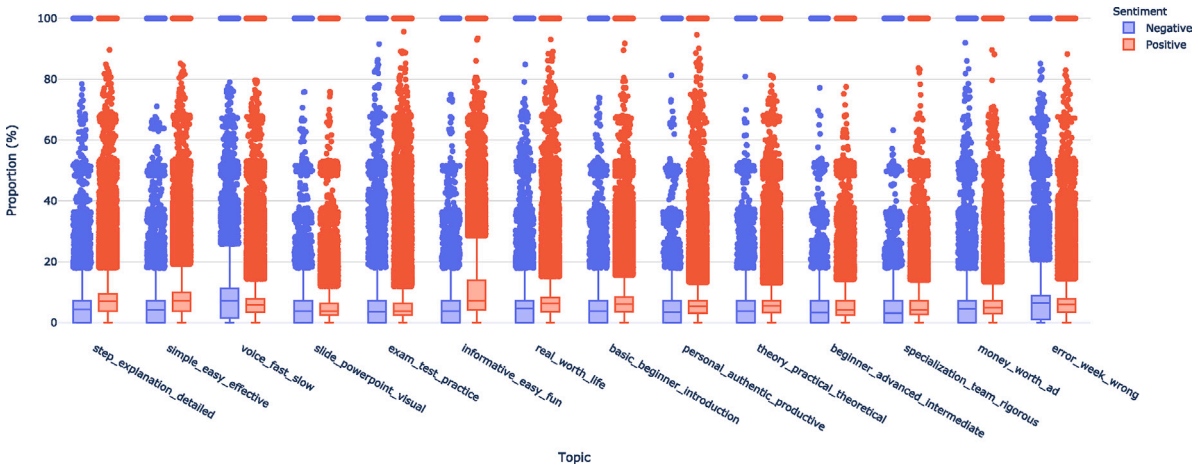


Fig. 8. Distribution of topics per course by sentiment values.

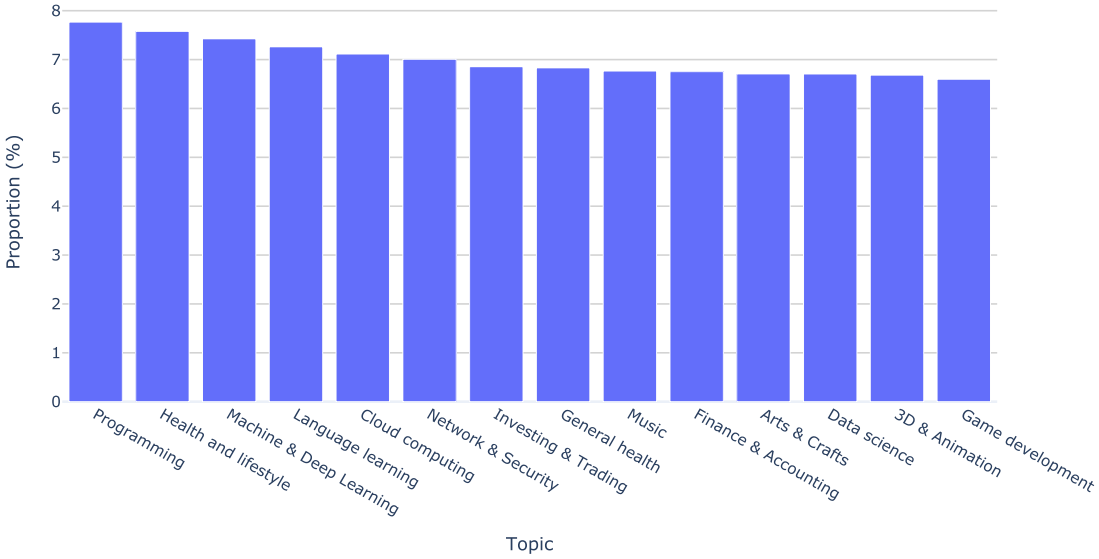


Fig. 9. Distribution of topics for the `Content` model.

Table 2
Summary of each detected topic regarding courses' content.

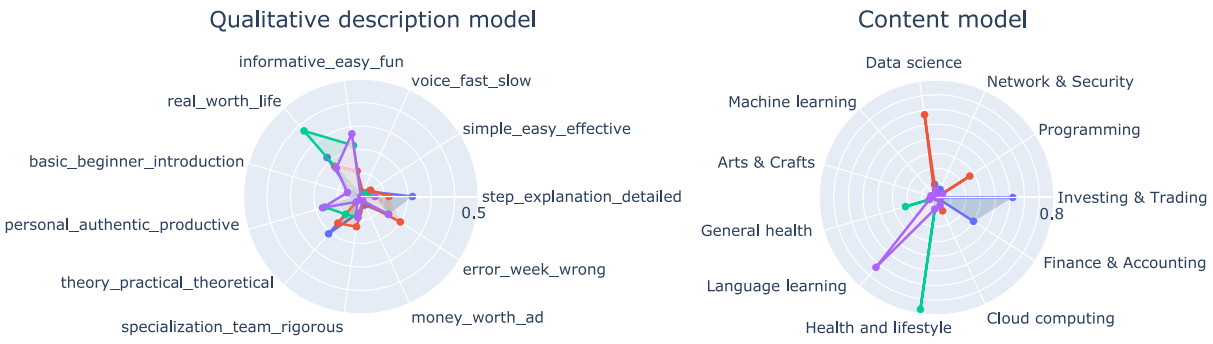
Topic label	Description	Main terms
Health and lifestyle	Courses that contribute to physical, mental, and social well-being (balanced diet, getting more rest, doing physical exercise...).	Life, exercise, food, calm, meditation
Programming	Courses that aim to teach programming knowledge.	Exercise, programming, code, language, programmer
Machine and deep learning	Includes courses that teach machine and deep learning techniques. Both aim the Artificial Intelligence (AI) to learn from data and then apply what they have learned to make informed decisions.	Machine, deep, learning, neural, network
Cloud computing	Courses teaching how to use cloud computing services (e.g., Microsoft Azure, Amazon Web Services).	Feature, topic, angular, azure, api
Investing & Trading	Includes courses that aim to teach investing and trading knowledge and techniques (i.e., methods of attempting to profit in the financial markets).	Business, trading, market, trade, financial
Music	Courses related with the music field, such as music production, how to sing, or how to play any instrument.	Music, play, audio, song, piano
Network & Security	Courses that aim to teach security and networks related content, such as how to prevent a hacking attack or how to provide more security to your devices.	Security, software, hack, project, management
Language learning	Courses dedicated to learn any language (e.g., Spanish, French).	Language, speak, accent, pronunciation, chinese
Finance & Accounting	Content related with accounting, which is an essential tool for providing information for decision-making, as well as for the evaluation of decisions previously made, and finance, that must seek resources at a reasonable cost and use them efficiently (Melé et al., 2017).	Management, business, leadership, law, economic
Arts & Crafts	Includes courses teaching knowledge about decorative design and handicraft.	Draw, art, paint, artist, oil
General health	Includes different general health sub-topics such as economic aspects, privacy, philosophy, or cultural aspects.	Life, topic, business, psychology, philosophy
Data science	Emerged as a new and important discipline, it can be viewed as an amalgamation of classical disciplines like statistics, data mining, databases, and distributed systems (Van Der Aalst, 2016).	Datum, science, statistical, excel, visualization
3D & Animation	These courses include 3D modeling and animation approaches, modeling objects or characters and utilizing motion in order to bring those characters, objects and more to life.	Design, software, graphic, animation, software
Game development	Includes different approaches that are part of developing a video game.	unity, game, unreal, software, engine

Then, we evaluated the two topic models obtained through NLP on the open reviews. For the qualitative description model on the left, the COVID-19 and English courses have a high proportion of the topics "real_worth_life", "informative_easy_fun", and "personal_authentic_productive", meanwhile the Python related course has a higher proportion of the topic "error_week_wrong", which indicates that reviews are revealing more problems in this last course than in the others. Furthermore, this finding is also confirmed by the fact that this course is the one with the lowest rating mean and sentiment analysis result of the selected courses. In addition, the topics that have a high proportion in the COVID-19 and English courses reveal additional positive aspects of these courses, which is the personalized and real application that they show. Regarding the content model, we see that our approach obtains an accurate result when modeling the content of each course automatically. For example, we see that the health related course, which was originally classified within the "Health" category, has a very high proportion of the topics "General health" and "Health and lifestyle". Furthermore, the business related course, which was originally labeled as a "Business" course, has a very high proportion of the topics "Investing & Trading" and "Finance & Accounting". These results validate that our models can automatically add additional and reliable information for learners to select the most appropriate course to their preferences.

5. Discussion

5.1. Interpretation of results

To achieve the proposed objectives in this research, authors used around 2.4 million reviews from learners who participated in 93,678 different courses across four MOOC platforms. Major findings revealed in this study are reported in this section and discussed in conjunction with the existing MOOC literature.



MOOC title	Rating mean	Sentiment analysis
Finance for Non-Finance Professionals (N = 342)	4.83	Positive: 96.2%, Negative: 3.8%
Using Python to Access Web Data (N = 5748)	4.73	Positive: 88%, Negative: 12%
Managing Your Mental Health During COVID-19 (N =1183)	4.81	Positive: 96.2%, Negative: 3.8%
Teach English Now! Foundational Principles (N = 2447)	4.91	Positive: 96.7%, Negative: 3.3%

Fig. 10. Rating mean, sentiment and topic analysis of four different MOOCs.

In Section 2 we identified some previous studies that also tried to analyze existing reviews in the MOOCs field. For example, Chen et al. (2020) analyzed 1920 reviews from 339 computer science courses, finding that 64.2% of the reviews in their data collection were 5-star reviews, while 16.7% were 4-star reviews. Similarly, Gamage et al. (2016) found that 18.24% of the courses had a 5-star rating, while 43.8% of the courses had a 4-star rating. The results of our research also confirm that the majority of reviews are 5-star and 4-star ratings. This bias is evident in the fact that most courses receive a majority of 5-star ratings. Further, using the 5-star rating system may provide limited insights in separating good from great products or supplies, and star ratings may be less meaningful to differentiate what learners like and dislike about a course design (Li et al., 2021). Instead, this research also explores sentiment analysis and topic modeling to gather insights into different course elements that were positive or negative to the learner experience.

The majority of MOOC reviews are less than 50 words long, and such short text is likely to create poor context-dependency, insufficient co-occurrence information, and sparse feature matrix, and result in an awful performance of LDA (Qi & Liu, 2021). In addition to having the biggest dataset analyzed in the area, our research found that the average length of the extracted reviews was 134.6 characters, which is larger than the typical average length of a review (Knöös & Rääf, 2021). In addition, since the coherence score is only an indicator of the model performance, we made iterations and manual reviews of the topics to ensure the relevance and coherence of the topics.

Our sentiment analysis results confirm that we have a skewed data, revealing also a positive correlation with the numeric ratings. Previous studies (Knöös & Rääf, 2021) used *VADER* sentiment analysis to predict the sentiment of reviews, finding that the library could predict the sentiment of positive reviews with a 93% accuracy. However, the accuracy in the prediction of negative and neutral reviews was less than 60%. Thus, since there might be some negative reviews classified as positive, we can only estimate the factors and topics that learners had a negative sentiment towards.

In their study, Deng and Benckendorff (2021) showed that learners placed high importance on making connections between MOOC content and their everyday lives. In our research, we found that reviews also made emphasis on the importance of MOOCs application in real life, with frequent topics and important keywords, such as topic 9 (“practice”, “real”) and topic 12 (“real”, “worth”, “life”, “practical”). Seeing this trend, we recommend that MOOCs content should be focused on a more practical approach, being applied in real life contexts and environments, and instructional conditions in MOOCs should be engineered to facilitate the acquisition of knowledge that can be transferred to real-world practices. Moreover, Chen et al. (2020) explored the relationship between ratings and topics identified. Following a similar approach, we decided to explore the relationship between sentiment analysis and our topic modeling approach, finding that courses with positive sentiment value tend to be related to positive topics, while courses with negative sentiment value tend to be more related with negative topics.

Furthermore, we can take a look at both of our topic models and discuss the coherence and usability of the topics revealed. Regarding the qualitative description model, we see that some topics could be really useful for students to identify the courses’ key features, such as topic 14 (“slide”, “powerpoint”, “lack”, “repetitive”) identifying courses that might be not as well designed as other, or topic 2 (“personal”, “authentic”, productive”) highlighting the personalization that many students missed in other courses (Gütl et al., 2014). However, we also identify other topics that might not add useful information, such as topic 11 (“beginner”, “intermediate”, “advanced”) or topic 13, mixing the keywords “theory” and “practical”. Regarding the content model, we believe that the topics revealed by the model are appropriate in comparison with the original categories defined by the courses themselves, and also consistent between them. However, a topic model with a larger number of topics may discover new ones that are not included in our current analysis.

5.2. Implications and limitations

This research’s findings validate that the existing open MOOC review systems contain courses with extremely positive ratings, making the decision-making process challenging. A potential MOOC student will face thousands of highly-rated MOOCs for popular topics such as Python or Excel, wherein the majority of courses ratings are around 4.7 stars or above, exacerbating what could be described as a “choice paradox” (Scheibehenne et al., 2010). No MOOC provider has improved significantly since MOOCs became popular in 2012, and it is unknown whether providers mostly display courses with the highest rating by curation or if a structural decision pushes the overall distribution of the system to a standard that does not reflect reality. Moreover, there is a clear economic incentive for MOOC providers to only display courses in the high four-stars range. A recent study from McKinsey found that even the slightest improvements in the star rating score (+0.2) in the retail sector could trigger a 37% increase in sales throughout a product life cycle (Briedis et al., 2020). The bias mentioned above, combined with an online education system with completion rates as low as 7%, creates the perfect scenario for MOOC students to be skeptical about the quality promoted by these providers.

This study also has some limitations. Non-English reviews were excluded from this study. An analysis of non-English reviews may provide additional insight into the learning experience idiosyncratic to these learners (e.g., individuals who use non-English MOOC platforms, individuals whose primary language is not English) (Deng & Benckendorff, 2021). Future research could overcome this limitation by analyzing non-English reviews to provide a more comprehensive and complete understanding of the learning experience in MOOCs. Moreover, the reviews considered for the analysis were not further filtered by length and contextual richness. Arguably, further research could utilize a subset of reviews that would provide more context, and ergo the results of a new topic labeling system could provide more nuance for the student.

There are reasons to believe artificial intelligence and community-driven tools can help improve the overall system. In this vein, the approach in our work, using topic modeling, and the standardization and decentralized review systems can provide more transparency to the decision-making process. Moreover, we can foresee increasing interest in platforms leveraging the aforementioned techniques and want to help students in their educational journey. Thus, creating a better system that increases the Return of Education (ROE).

6. Conclusions

In this study, we aimed to answer four research questions in order to support learners' course selection in MOOCs, using the largest dataset so far in the literature. First, we analyzed the numeric ratings in our data collection, finding a biased dataset with a majority of 5-star and 4-star ratings. Then, we performed sentiment analysis on textual reviews using three pre-trained models from different libraries. Results suggests that there is a positive correlation between the compound sentiment values and the numeric ratings. All three libraries show that the majority of textual reviews (with results varying between 80.8% and 83.8%) are classified as positive reviews. Then, we built two models based on a topic modeling approach. The first model used words that described the courses in a qualitative way, and revealed some interesting topics and keywords such as "informative_easy_fun", "money_worth_ad" or "slide_powerpoint_visual". This approach can effectively characterize courses, thus helping learners to decide which course to choose, highlighting features that have been important for learners over time. The second model used words that described the courses' content, and revealed frequent topics such as "Health and lifestyle", "Programming", or "Cloud computing". Since each course can cover different areas, this approach can help learners to discover some hidden topics in existing courses, based on other learners' reviews. We also calculated the distribution of each topic, which allowed us to see which were the most frequent and the less frequent topics across the entire collection. Moreover, we explored the relationship between the sentiment analysis results and the qualitative topic modeling approach, finding that courses with positive sentiment value tend to be related to positive topics such as "informative_easy_fun" or "real_worth_life", while courses with negative sentiment value tend to be more related with negative topics like "error_week_wrong". Finally, we tried to characterize four different MOOCs using our models, which confirmed the positive correlation between the sentiment analysis and the numeric rating, but also the difference in topic characterization when comparing courses with higher and lower sentiment analysis results.

As part of our future work, we would like to validate the `Content` topic model against the original content category in every course to see the accuracy of the model. Moreover, since the typical limited size in reviews could be a problem for the LDA model, future studies could consider only large reviews in order to explore more relevant or meaningful topics. In addition, an interesting approach would be to assign a different weight to each review depending on the reputation of each reviewer. This could be done by certifying which users have completed the course, assigning different weights accordingly. Regarding sentiment analysis, some of the pre-trained sentiment models tend to fail in some classifications (specially referring to neutral and negative reviews). Thus, in future studies, the design of a lexicon that better takes into account negative words could improve the analysis and results obtained. Finally, another step would be the development of a real recommendation system to validate the appropriateness of these discovered topics for its real purpose: allow students to identify the best courses among the almost infinite variety of them. To wrap up, we believe that there is an opportunity to leverage current approaches and create more powerful, yet simpler reviewing systems, using techniques similar to the ones applied in this research. Thus, being able to really identify the best courses out there, and turn a blind eye to the current highly biased MOOC review systems.

CRediT authorship contribution statement

Manuel J. Gomez: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Visualization. **Mario Calderón:** Investigation, Resources, Data curation, Writing – original draft. **Victor Sánchez:** Investigation, Resources, Data curation, Writing – original draft. **Félix J. García Clemente:** Conceptualization, Writing – review & editing, Supervision, Project administration. **José A. Ruipérez-Valiente:** Conceptualization, Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

References

Aher, S. B. (2014). Em&aa: An algorithm for predicting the course selection by student in e-learning using data mining techniques. *Journal of the Institution of Engineers (India): Series B*, *95*(1), 43–54.

Aher, S. B., & Lobo, L. (2012). Prediction of course selection by student using combination of data mining algorithms in E-learning. *International Journal of Computer Applications*, *40*(15), 1–7.

Al-Badarenah, A., & Alsakran, J. (2016). An automated recommender system for course selection. *International Journal of Advanced Computer Science and Applications*, *7*(3), 166–175.

Al-Rahmi, W., Aldraiweesh, A., Yahaya, N., Kamin, Y. B., & Zeki, A. M. (2019). Massive open online courses (MOOCs): Data on higher education. *Data in Brief*, *22*, 118–125.

Al-Rahmi, W. M., Alias, N., Othman, M. S., Marin, V. I., & Tur, G. (2018). A model of factors affecting learning performance through the use of social media in Malaysian higher education. *Computers & Education*, *121*, 59–72.

Bezerra, L. N., & Silva, M. T. (2017). A review of literature on the reasons that cause the high dropout rates in the MOOCS. *Revista Espacios*, *38*(05).

Bhattacharjee, J. (2018). *Fasttext quick start guide: get started with facebook's library for text representation and classification*. Packt Publishing Ltd.

Billington, P. J., & Fronmueller, M. P. (2013). MOOCs and the future of higher education. *Journal of Higher Education Theory and Practice*, *13*(3/4), 36–43.

Briedis, H., Kronschnabl, A., Rodriguez, A., & Ungerman, K. (2020). Adapting to the next normal in retail: The customer experience imperative. URL: https://www.mckinsey.com/industries/retail/our-insights/adapting-to-the-next-normal-in-retail-the-customer-experience-imperative. (Web; last Accessed 15 November 2021).

Bruns, A., Kornstadt, A., & Wichmann, D. (2009). Web application tests with selenium. *IEEE Software*, *26*(5), 88–91.

Castillo, N. M., Lee, J., Zahra, F. T., & Wagner, D. A. (2015). MOOCS for development: Trends, challenges, and opportunities. *International Technologies & International Development*, *11*(2), 35.

Central, C. (2021). Class central. URL: https://www.classcentral.com, (Web; last Accessed on 15 November 2021).

Chang, H.-M., Kuo, T. M.-L., Chen, S.-C., Li, C.-A., Huang, Y.-W., Cheng, Y.-C., Hsu, H.-H., Huang, N.-F., & Tzeng, J.-W. (2016). Developing a data-driven learning interest recommendation system to promoting self-paced learning on MOOCs. In *2016 IEEE 16th international conference on advanced learning technologies* (pp. 23–25). IEEE.

Chen, X., Zou, D., Xie, H., & Cheng, G. (2020). What are MOOCs learners' concerns? Text analysis of reviews for computer science courses. In *International conference on database systems for advanced applications* (pp. 73–79). Springer.

Coursera (2021). Impact report 2021: Serving the world through learning. URL: https://about.coursera.org/press/wp-content/uploads/2021/11/2021-Coursera-Impact-Report.pd, (Web; last Accessed on 15 November 2021).

Coursera (2021). Introducing new tools and features as demand for online learning grows. URL: https://blog.coursera.org/introducing-new-platform-innovations-as-demand-for-online-learning-grows/. (Web; last Accessed on 15 November 2021).

Dañino, R. (2021). Crehana. URL: https://www.startupranking.com/crehana. (Web; last Accessed on 16 November 2021).

Deng, R., & Benckendorff, P. (2021). What are the key themes associated with the positive learning experience in MOOCs? An empirical investigation of learners' ratings and reviews. *International Journal of Educational Technology in Higher Education*, *18*(1), 1–28.

Dhawan, S. (2020). Online learning: A panacea in the time of COVID-19 crisis. *Journal of Educational Technology Systems*, *49*(1), 5–22.

Elbagir, S., & Yang, J. (2019). Twitter sentiment analysis using natural language toolkit and VADER sentiment. In *Proceedings of the international multiconference of engineers and computer scientists, vol. 122, no. 16*.

Gamage, D., Perera, I., & Fernando, S. (2016). What do star rates for MOOCs tell you? An analysis of pedagogy and review rates to identify effective pedagogical model. In *24th international conference in education. India: Asia-pacific society for computers in education*.

Graham, S., Weingart, S., & Milligan, I. (2012). *Getting started with topic modeling and MALLET*: *Technical report*, The Editorial Board of the Programming Historian.

Gütl, C., Rizzardini, R. H., Chang, V., & Morales, M. (2014). Attrition in MOOC: Lessons learned from drop-out students. In *International workshop on learning technology for education in cloud* (pp. 37–48). Springer.

Hew, K. F., & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educational Research Review*, *12*, 45–58.

Huang, L., Wang, C.-D., Chao, H.-Y., Lai, J.-H., & Philip, S. Y. (2019). A score prediction approach for optional course recommendation via cross-user-domain collaborative filtering. *IEEE Access*, *7*, 19550–19563.

Jobe, W., Östlund, C., & Svensson, L. (2014). MOOCs for professional teacher development. In *Society for information technology & teacher education international conference* (pp. 1580–1586). Association for the Advancement of Computing in Education (AACE).

Kapadia, S. (2019). Evaluate topic models: Latent Dirichlet allocation (LDA). URL: https://towardsdatascience.com/evaluate-topic-model-in-python-latent-dirichlet-allocation-lda-7d57484bb5d0, (Web; last Accessed on 20 October 2021).

Kapoor, N., Vishal, S., & Krishnaveni, K. (2020). Movie recommendation system using nlp tools. In *2020 5th international conference on communication and electronics systems* (pp. 883–888). IEEE.

Khalid, A., Lundqvist, K., & Yates, A. (2020). Recommender systems for MOOCs: A systematic literature survey. *International Review of Research in Open and Distributed Learning*, *21*(4), 255–291.

Khalil, H., & Ebner, M. (2014). MOOCs completion rates and possible methods to improve retention-A literature review. In *EdMedia+ innovate learning* (pp. 1305–1313). Association for the Advancement of Computing in Education (AACE).

Knöös, J., & Rääf, S. A. (2021). Sentiment analysis of MOOC learner reviews: What motivates learners to complete a course?.

Kouzis-Loukas, D. (2016). *Learning scrapy*. Packt Publishing Ltd.

Lan, M., & Hew, K. F. (2020). Examining learning engagement in MOOCs: A self-determination theoretical perspective using mixed method. *International Journal of Educational Technology in Higher Education*, *17*(1), 1–24.

Li, L., Johnson, J., Aarhus, W., & Shah, D. (2021). Key factors in MOOC pedagogy based on NLP sentiment analysis of learner reviews: What makes a hit. *Computers & Education*, Article 104354.

Liyanagunawardena, T. R., Parslow, P., & Williams, S. (2014). Dropout: MOOC participants' perspective.

Lohr, S. (2020). Remember the MOOCs? After near-death, they're booming. *The New York Times*, *26*.

Loria, S. (2018). Textblob documentation. *Release 0.15*, *2*, 269.

Melé, D., Rosanas, J. M., & Fontrodona, J. (2017). Ethics in finance and accounting: Editorial introduction. *Journal of Business Ethics*, *140*(4), 609–613.

O'callaghan, D., Greene, D., Carthy, J., & Cunningham, P. (2015). An analysis of the coherence of descriptors in topic modeling. *Expert Systems with Applications*, *42*(13), 5645–5657.

Ouertani, H. C., & Alawadh, M. M. (2017). MOOCs recommender system: A recommender system for the massive open online courses. In *Innovations in smart learning* (pp. 139–143). Springer.

Pappano, L. (2012). The year of the MOOC. *The New York Times*, *2*(12), 2012.

Pina, A. R. B., & Steffens, K. (2015). Are MOOCs promising learning environments? *Comunicar. Media Education Research Journal*, *23*(1).

Platzi (2021). Platzi. URL: https://platzi.com. (Web; last Accessed on 16 November 2021).

Purkayastha, N., & Sinha, M. K. (2021). Unstoppable study with MOOCs during COVID 19 pandemic: A study. *Library Philosophy and Practice*, 1–12.

Qi, C., & Liu, S. (2021). Evaluating on-line courses via reviews mining. *IEEE Access*, *9*, 35439–35451.

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on web search and data mining* (pp. 399–408).

Rowan, N. (2021). What is domestika?. URL: https://tinyworkshops.com/domestika/. (Web; last Accessed on 16 November 2021).

Saxton, M. D. (2018). A gentle introduction to topic modeling using Python. *Theological Librarianship*, *11*(1), 18–27.

Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2010). Can there ever be too many options? A meta-analytic review of choice overload. *Journal of Consumer Research*, *37*(3), 409–425.

Shah, D. (2021). By the numbers: MOOCs in 2020. URL: https://www.classcentral.com/report/mooc-stats-2020/. (Web; last Accessed on 15 November 2021).

SkillMapper (2022). Did you pay for an online course and never finish it? You aren't alone and lack of time isn't always to blame. URL: https://blog.skillmapper.com/?p=18. (Web; last access on 10 January 2022).

Terras, M. M., & Ramsay, J. (2015). Massive open online courses (MOOCs): Insights and challenges from a psychological perspective. *British Journal of Educational Technology*, *46*(3), 472–487.

Terry-Jack, M. (2019). NLP: Pre-trained sentiment analysis. URL: https://medium.com/@b.terryjack/nlp-pre-trained-sentiment-analysis-1eb52a9d742c. (Web; last Accessed on 20 October 2021).

Udemy (2021). About us. URL: https://about.udemy.com. (Web; last Accessed on 15 November 2021).

Van Der Aalst, W. (2016). Data science in action. In *Process mining* (pp. 3–23). Springer.

Vanitha, V., Krishnan, P., & Elakkiya, R. (2019). Collaborative optimization algorithm for learning path construction in E-learning. *Computers and Electrical Engineering*, *77*, 325–338.

Wen, M., Yang, D., & Rose, C. (2014). Sentiment analysis in MOOC discussion forums: What does it tell us? In *Educational data mining 2014*. Citeseer.

Yao, L., Mimno, D., & McCallum, A. (2009). Efficient methods for topic model inference on streaming document collections. In *Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 937–946).

Zhang, H., Huang, T., Lv, Z., Liu, S., & Zhou, Z. (2018). MCRS: A course recommendation system for MOOCs. *Multimedia Tools and Applications*, *77*(6), 7051–7069.