Assessing Teacher Competencies in Higher Education: A Sentiment Analysis of Student Feedback

Charalampos Dervenis[1,*], Panos Fitsilis[1], Omiros Iatrellis[2], and Athanasios Koustelios[1]

[1]Department of Business Administration, University of Thessaly, Larisa, Greece
[2]Department of Digital Systems, University of Thessaly, Larisa, Greece
Email: dervenisch@gmail.com (C.D.); fitsilis@uth.gr (P.F.); iatrellis@uth.gr (O.I.); akoustel@uth.gr (A.K.)
[*]Corresponding author
Manuscript received November 6, 2023; revised December 5, 2023; accepted December 20, 2023; published April 7, 2024

***Abstract*—One of the most important issues concerning education nowadays is that of mapping the quality of teaching, and teacher competencies. Simultaneously, the need to exploit the huge amount of data derived from student feedback, and in particular the comments on open-ended questions, constitutes a huge challenge for both universities and researchers. In this study, we use sentiment analysis methods to measure teachers' communication competencies. Utilizing the data of over 700 feedback comments from students of our university, we assessed specific competencies through the sentiment intensity that these comments contained. The model designed for the sentiment analysis, as well as the entire experimental phase, were implemented using an open source data mining and visualization platform. Our research revealed that certain competencies are highlighted by the nature of the course while others do not depend on it. In addition, findings indicate both a homogeneous and convergent view among students, thus strengthening the validity of the students' opinion and their valuable contribution to the mapping of teaching quality in general with the ultimate goal of enhancing education.**

***Keywords*—teacher competencies, student feedback, sentiment analysis, teacher effectiveness**

I. INTRODUCTION

Nowadays, with the rapid spread of data and its huge availability arising from various educational processes, the majority of universities frequently use data mining methods to examine data collected and to extract hidden knowledge from student feedback with the aim of improving teaching quality [1–5]. However, the data sets are too large or complex to deal with using traditional data application software. Therefore, institutions are confronted with the challenge of how to tap into this vast amount of data with the ultimate goal of enhancing student learning. These data generally consist of student perspectives regarding various aspects such as their attainment of a course's learning objectives, the effectiveness of instructional methods, the competence of the instructor, the content of the course, and even the students' overall views on their academic institutions. When taken into account, this data helps academic institutes in their curriculum design and in the management decisions of their organizations. It also gives the respective teacher the opportunity to understand their teaching approaches, strengths and weaknesses. In this way, teachers are better equipped to make appropriate changes thus contributing to the overall effectiveness of their teaching and student learning.

The traditional techniques for student feedback analyses are based on close-ended questionnaire data collection and analysis [6, 7]. These forms, based on Likert-scale items of various types, provide quantitative data. However, the majority of academic institutions, such as our university, additionally offer students the ability to express their feedback opinions through text field (open-ended questionnaires). According to previous research, this institutional educational text data is commonly collected but not fully exploited and remains untapped, thus depriving the acquisition of knowledge that would contribute to the improvement of teaching and learning [8]. This is because unlike Likert-scale based data, which is quickly and easily analyzed using simple statistical methods, open-ended comments require special treatment and analysis.

This paper aims to investigate and highlight how text educational data from student feedback could provide valuable insight to academic institutions. In this study we will attempt to exploit the data gathered from over 700 student comments retrieved from open-ended questions regarding their teacher's specific competencies as a continuation of our previous research. This data will be analyzed and studied as to how it could contribute to the assessment of teacher communication competencies through opinion mining methods and Sentiment Analysis. We will also investigate whether students perceive their professors' competencies differently depending on the nature of the course. Accordingly, the research questions that guided this study were:

- RQ1: How can written text educational data generated from student feedback be used to measure specific teacher competencies through the Sentiment Analysis approach?
- RQ2: Does a student's perception of a teacher's specific competencies vary depending on the nature of the course?

This paper is structured as follows. After the introduction made in Section I, a brief overview of related work follows in Section II. Next, we describe the methodology and present a case study with real data provided by our university (Section III). Section IV is devoted to a discussion on our results. Lastly, Section V provides concluding remarks.

II. RELATED WORK

A. The Value of Student Feedback in Teaching Progress

Academic institutions have always aimed to collect and take their students' views into account, in an effort to improve their teaching and curriculums [9, 10]. Flodén [11], with thorough research, focuses on student feedback and its positive acceptance from university teachers, emphasizing that both positive and negative student feedback are equally valuable to their teachers. Mandouit [12] concluded that when

doi: 10.18178/ijiet.2024.14.4.2074

provided with professional learning to support the process, student feedback has the potential to have a positive impact on teaching practices and contributes to opening up a dialogue about teaching and learning.

Another pioneering study dealing with an alternative approach to student feedback on courses is that conducted by Marley [13] which claims that the positive and negative experiences of online learners are extremely useful for educators involved in the design and development of online teaching. According to Nasim [14], in their feedback, students' views touch on multiple aspects of their teachers such as teaching style, lecture organization, knowledge, punctuality, communication competence and are valuable in the learning process.

Other research [15, 16] has found that, student feedback contributes to various aspects of academic institutions by exporting information and knowledge to facilitate decision making in their organization. Marsh [15] claims that student opinion mining, beyond serving as diagnostic feedback for academics about the effectiveness of their teaching, provides a measure of teaching effectiveness for decisions regarding their appointment and promotion and a valuable component for use in the quality assurance processes of academic institutions.

Student feedback is collected in different forms, such as close-ended questions usually in Likert scales, open-ended questions and free text comments. Different mediums are also used, such as, classroom feedback, social media, online platforms, etc., by students to express their perceptions of issues about their instructors and institutions [17]. This whole array of educational data has led to more demanding data management systems and complex analytical approaches resulting in the emergence of a specialized research field known as Educational Data Mining (EDM) [18].

B. Sentiment Analysis in Education

Sentiment analysis, also known as opinion mining, is a field of study, directly related to natural language processing that analyzes views, sentiments, evaluations, attitudes and positions of people on issues through the computational processing of subjectivity in text [19]. Sentiment analysis aims at taking the sentiment of a simple word, phrase, sentence or even an entire document into consideration and rating it appropriately. Essentially, using appropriate lexicons or machine learning approaches [20], opinion polarity is determined, as to whether it is positive, negative or neutral [21].

The initial challenges encountered in the practical application of sentiment analysis and opinion mining stemmed from the huge pace and volume of comment content observed on social media on the web. Since then, Sentiment Analysis (SA) has penetrated and found application in a wide range of fields, focusing on customer opinions about products and services through their written comments. As a result, with the growing familiarity of users with the Internet and online review writing, SA has established itself as a very powerful method of deciphering customer desires in various marketing companies, in the movie industry, in policy-making, in public transportation, in the travel industry, as well as in academic institutions. For the first time in history, we now have a huge volume of opinionated data recorded in digital form for analysis [19]. As a result, Sentiment analysis, considered as a big data task, is constantly arousing the interest of researchers.

Traditionally, in academic institutes, student feedback analysis systems, as a rule, were based on close-ended questionnaires that included a set of questions with predefined possible answers [16]. With Sentiment Analysis techniques, students have no limitation to grade only those features that are included in the questionnaire, but instead opinions are freely expressed. According to Nasim [14], textual feedback is the best way for students to express their views spontaneously, without being limited through predetermined answers. Despite its potentially huge size, this quality educational data coming from a collection of texts can now be easily analyzed with the help of computer systems using Sentiment Analysis techniques [22].

Kumar and Jain [16] proposed an automatic evaluation system based on sentiment analysis collecting student feedback in the form of text. Misuraca *et al.* [3] used a dataset of text, built by Welch and Mihalcea [23], with 1042 comments extracted from a Facebook group where students expressed their opinions on courses and instructors pertaining to the Computer Science department at the University of Michigan (US). They observed that, concerning the polarity score calculation, it is worth adjusting lexicons to the specific language commonly used by students, particularly because the polarization of some terms may be different from the polarization used in common language. Nitin *et al.* [24] observed that students provide feedback that is usually related to labs, projects, skills, etc. when it comes to components of a course which are non-theoretical (interactive, practical, collaborative) in nature. On the contrary, when it comes to theoretical components, student feedback is concerned directly with the profile of their teachers.

Although there is research that has shown that the neutral comments, in sentiment analysis, are not of particular research interest [25], even student feedback categorized as neutral can contribute positively in terms of understanding educational issues. Ortigosa *et al.* [26], especially in the education domain, point out the necessity, even of the students' neutral feedback, in order to comprehend their attitudes.

It is not our intention to review the entire body of literature concerning sentiment analysis from a technical point of view. Nevertheless, we will attempt to provide a brief, and in our opinion necessary, overview in the following section.

C. Technical Background

Most sentiment analysis approaches rely heavily on lists of lexical features (e.g., words, phrases), called sentiment lexicons, generally labeled according to their semantic orientation as either positive, negative or neutral [19]. These lexicons are classified into two main categories: (1) Polarity-based Lexicons and (2) Valence-based Lexicons.

In the first case, a Polarity-based Lexicon is used to determine the binary polarity of words i.e., either positive or negative. For example, "angry" has negative polarity while "family" positive. To calculate the total polarity of a text, the polarity of each word is searched in the lexicon and if found, is added to obtain an overall score. These lexicons, such as LIWC, Hu and Liu, etc., are widely known and validated.

However, the sentiment analysis in practice has become more demanding, applying more complex procedures in order to express, not only whether a linguistic term is positive or negative, but also its polarity's intensity. These valence-based lexicons, such as Valence Aware Dictionary for sEntiment Reasoning (VADER), Affective Norms for English Words (ANEW), SentiWordNet etc., use word lists with a valence score. For example, using the VADER lexicon measuring sentiment-intensity on a scale from −4 to +4, the term "okay" scores +0.9, "great" scores 3.1, "horrible" −2.5, ":(" or "☹" scores −2.2 and so on. Beyond the sentiment word intensity provided, there are sentiment lexicons, such as VADER, which identify properties and specific characteristics of the text, especially in short comments, which affect the perceived sentiment intensity of the text. For example, the exclamation mark, adds more orientation intensity. "Excellent!" is stronger than "Excellent". In addition, it understands many sentiment-laden emoticons such as ":)", "D" and sentiment-laden initialisms and acronyms such as "lol". It is worth noting that many lexicons of the 2 aforementioned categories, Polarity-based Lexicons and Valence-based Lexicons, were created based on or combining earlier lexicons.

In other cases, machine-learning approaches are incorporated to achieve automated sentiment-related tracing processes in the text, considering the knowledge base obtained. Kang *et al.* [27] proposed an improved Naïve Bayes algorithm to solve problems related to classification performance. Chen and Tseng [28] used the Support Vector Machine (SVM), a machine learning algorithm to classify reviews. Al Amrani *et al.* [29], proposed a hybrid approach by making a combination of machine learning techniques, using Random Forest and SVM algorithms, when applying sentiment analysis to comments from product reviews offered by Amazon. In another powerful study, Kabir *et al.* [30] used different machine learning algorithms and concluded that Boosting and Maximum Entropy outperform the other examined machine learning algorithms for detecting sentiments in online user reviews. However, it should be noted that, most machine-learning approaches have drawbacks that are mainly related to the requirement of training data set. This data set must be extensive and representative. In addition, significant training and classification [21] time is required and therefore computing power (CPU processing) and high memory.

In each case, sentiment analysis can be conducted using different techniques with some alterations. In this paper, aiming at presenting a well-defined approach, we describe it step-by-step and next apply it on student feedback comments of our university, with no intention to challenge any other similar or alternative approach. Rather, we aim to highlight the diversity of approaches and to encourage the exchange of views for the purpose of exploring the specific subject.

III. Methodology and Sentiment Analysis Approach

A. *Methodology*

In this section, we provide the method that we applied for the sentiment analysis procedure based on the lexicon approach. The initial stage (Stage 1) concerns the data set imported to the input channel. This data set contains the student responses in text. The data collection process is referred to and thoroughly described in the "Implementation" subsection of our approach. The data can be imported in various file formats such as Excel (.xlsx), comma-separated (.csv) and native tab-delimited (.tab) files and others, depending on the system used. Then, in Stage 2, the text preprocessing process takes place, during which some natural language issues such as *Transformation*, *Tokenization, Stopwords filtering, Normalization (stemming and lemmatization)* etc. should be considered. The process of text preprocessing consists of a sequence of steps and its objective is to prepare the raw text data for the next phase, thus acquiring a clean and consistent form that can then be fed into a model for further analysis. In general, text preprocessing is an integral part of Natural Language Processing (NLP) applications. Fig. 1 illustrates the order of distinct steps followed in our approach during the text preprocessing phase. Additionally, Table 1 summarizes thoroughly all of the text data challenges handled in these steps.

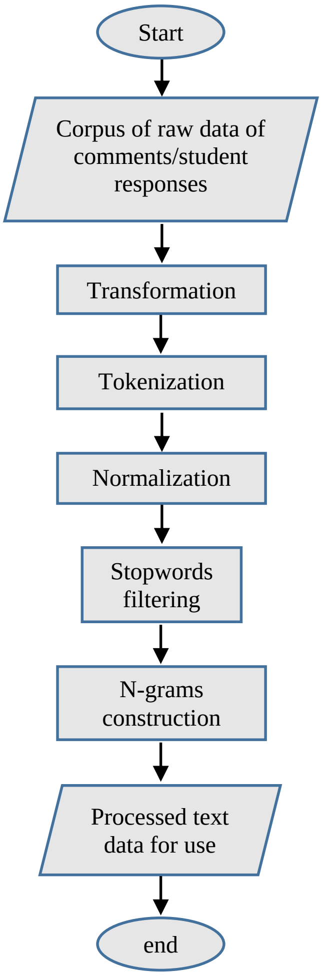


Fig. 1. Text preprocessing phase (Stage 2).

In Stage 3, the sentiment intensity for each segment of the data processed in the previous stage is detected. To do this, a prepared sentiment lexicon was used containing words and corresponding sentiment scores for each word. This assignment of polarity is quite important, as the purpose is to calculate the overall score of a student comment on a particular topic, although in the individual words, aspects, etc. different grades may be assigned. For instance, a comment consisting of 7 tokens may have a positive average grade, even though 5 of the tokens have negative scores and only the other two have a positive one.

Table 1. Text data challenges

Process	Description
Transformation	Removes urls from text. Detects and retains only the text between html tags. For example, from valuable text retains the "valuable text". If necessary, applies lowercase transformation.
Tokenization	Tokenization is the method of splitting text into smaller components, called tokens or units. These units may be words, bigrams, sentences and even paragraphs. In addition to breaking up a sequence of strings into pieces, a characteristic task in the process of tokenization is the discarding of certain characters such as punctuation.
Normalization (stemming and lemmatization)	This method is used to find out the stem or lemma of a word. Words are stemmed using different stemming algorithms such as, Porter stemmer [31], Snowball stemmer [32] etc. For instance, the word "flying" is stemmed as "fly" by removing the "ing" suffix. It is very important to remove the endings of the words, to return the root word so that it is given the intensity of polarity based on the dictionary used.
Stopwords filtering	At this stage, stopwords from texts are removed, i.e., pronouns, articles, that carry very little useful information such as "he", "it", "the", "and", etc., without changing the semantics of a text. These words occur commonly across all the texts in the corpus and need to be dissociated in order to retain only the meaningful words and improve the performance of our model. Furthermore, at this stage we can remove words that make no contribution when analyzed, such as verbs or prepositions with high frequency of occurrence and of negligible interest (be, with, etc.).
N-grams construction	The N-gram consists of successive tokens, with determined window (N tokens), that occur in texts. N-grams of texts are extensively useful in text mining and natural language processing tasks [33]. For instance, "Difficult lesson" is a 2-gram and "Difficult lesson in lab" is a 4-gram with different meaning. It is quite interesting to determine one or the other case and to find out the frequency of their occurrence and polarity in a plethora of student feedback. Fig. 2 illustrates an example of a 1-gram and a 2-gram collection of tokens in the same corpus.

In the final stage (Stage 4), all the sentiment analysis insights are converted into relevant reports. Reports enable interactivity through different forms such as charts, graphs, etc. At this stage of the sentiment analysis steps, visualization is a very important segment for the overall monitoring of the results and the receiving of actionable insights. To enable the decision of the right course of action, it is crucial to know which aspects got high a score, which ones got low scores and generally, to have an idea as to which areas need our attention more than others.

B. Implementation

At the end of each semester at our university, students are asked to answer a questionnaire about each course and its instructor, anonymously online. For this purpose, the Quality Assurance Unit of the University of Thessaly, which is the central body for the coordination and support of the quality assurance procedures, sends a relevant link to the students' email. The questionnaire is harmonized with the requirements of the National Higher Education Assessment Authority of Greece and contains close and open-ended questions. The data from student responses are gathered online and are available to authorized administrators. However, as a continuation of our previous research [34] and independently of the aforementioned process, we aimed to measure the teacher competencies referring to the dimension of "Communication" (Table 2).

Table 2. Teacher competencies that refer to the "Communication" dimension with the corresponding open-ended questions asked to students

Teacher competence to be measured	Open-ended question asked to students.	Question coding
C22: Listening	Note general comments about your professor regarding their competence in actively listening (i.e., making a conscious effort to hear not only the words said but, more importantly, the complete message being communicated).	Q1
C23: Persuasion	Note general comments about your professor regarding their powers of persuasion used to inspire students and enhance / boost their interest in the subject.	Q2
C24: Empathy	Note general comments about your professor: regarding the empathy they feel for their students. (Empathy as a basic dimension in emotional intelligence, is the ability to put oneself in another person's shoes and understand what they are feeling).	Q3
C25: Presentation	Note general comments about your professor regarding their presentation abilities (i.e., do they deliver their presentation in a clear and concise manner?).	Q4
C26: Collaboration	Note general comments about your professor regarding their collaboration with students (i.e., are they cooperative, do they detect an unwillingness to cooperate and prevent it, do they take into consideration the smooth functioning of the course?).	Q5

For this reason, we additionally developed a specific questionnaire, oriented to our purpose which was provided online to students of our university through a Google Forms link sent to their emails. Students were asked to provide text responses anonymously to open-ended questions about *Listening, Persuasion, Empathy, Presentation and Collaboration* demonstrated by their teacher. Our purpose was to gather all text responses in one excel format file which would be the input file to our model. For the needs of our research, we initially focused on a course taught by two different professors in order to determine any differences in the nature of the teachers and not of the course. Subsequently, we repeated the same procedure with the same teacher in two different courses, to examine this time whether possible differences were due to the nature of the course and not of the teacher.

It is worth noting that, from a technical point of view, the platform allows us to choose the preferred language for data entry and processing. For our purposes, we judged that we should rely on the VADER lexicon which is fully open-sourced under the MIT License, in the English language, fully validated and widely accepted across multiple distinct domains including the field of education [35, 36]. Furthermore, the VADER sentiment lexicon has been compared to many other well-established sentiment analysis lexicons such as Linguistic Inquiry Word Count (LIWC), General Inquirer (GI), Affective Norms for English Words (ANEW), SentiWordNet (SWN), SenticNet (SCN), WordNet and the Hu-Liu, and the results were indeed quite remarkable.

The comparison results revealed that the VADER lexicon performs at a similar level to that of individual human raters. The classification accuracy metric "Precision" varied between 0.69 and 0.99 across various domain contexts [21].

At the same time, it was essential that the students express themselves in their mother tongue, Greek, in order for their responses to convey meaning and sentiment as accurately as possible [37]. Therefore, to achieve maximum accuracy and efficiency, all the comments were initially collected in the Greek language and subsequently translated into English by accredited translators and interpreters of Greek and English.

In this way, we ensured the highest degree of accuracy both from the students' feedback in their mother tongue, as well as by utilizing the VADER English dictionary which provided us with the validity needed for the purposes of our research. We then entered the final data into our system.

The entire experimental phase was implemented using the Orange data mining and visualization platform [38]. Orange is a powerful data mining visual programming tool for explorative qualitative data analysis and interactive data visualization. Our model takes the sequences of a sentence as input and gives the intensity of the sentiment polarity of the sentence as output. In Fig. 2, the workflow depicts the model designed according to our approach. After importing the raw data from the student feedback into the system, attributes definition of data takes place. That is, we assign, which data will be the target for analysis and which will not. In our case, the questions and their answers content are considered for analysis. Then, we processed the "raw" data in order to clean it up and to eliminate the noise, i.e., words that carry very little useful information such as "a", "and", "is", "the", etc.

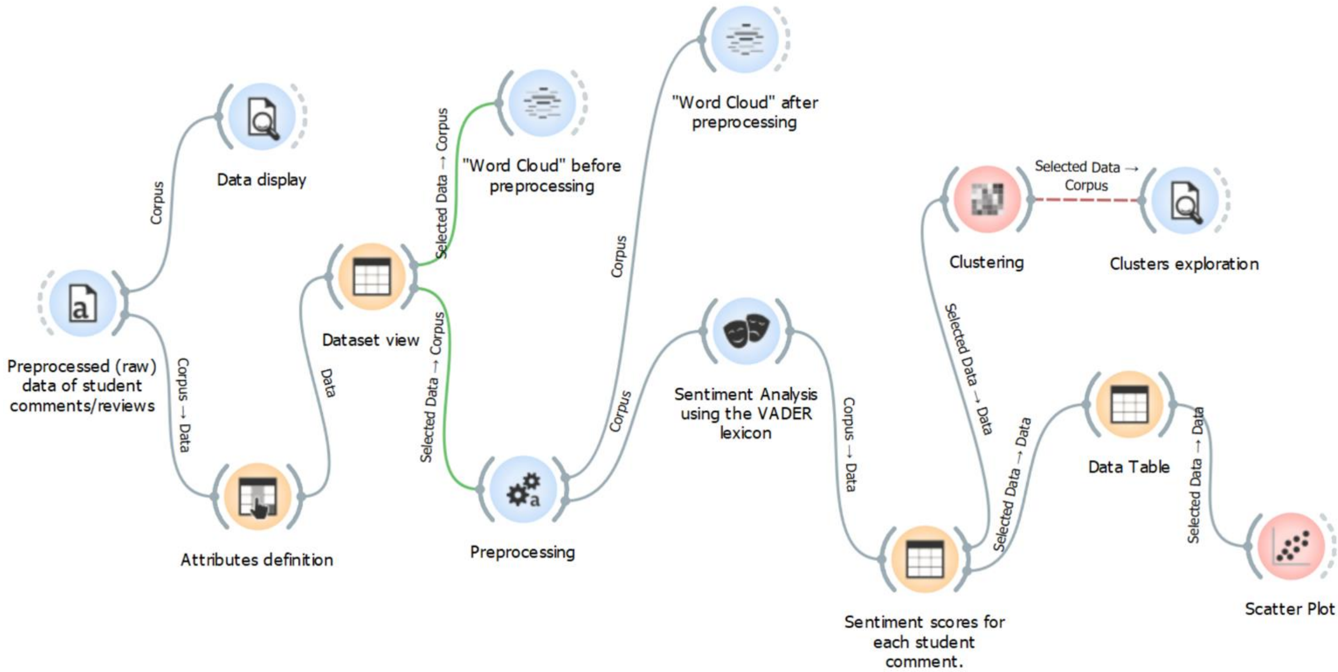


Fig. 2. Workflow of model designed for sentiment analysis.

Then, we utilized the VADER lexicon, which covers multiple domains and special cases for sentiment analysis that include the proper handling of sentences with typical negations (e.g., "not good"), the use of contractions as negations (e.g., "wasn't very good"), conventional use of punctuation to signal increased sentiment intensity (e.g., "Good!!!"), conventional use of capital letters to signal emphasis (e.g. "NICE") and other special cases such as those mentioned in the previous section (Section II). Subsequently, we computed and reported the sentiment score as positive, negative or neutral to reach the compound score of each student comment received. The compound score of a specific sentence is determined by summing up the sentiment scores of each individual word within the sentence [39]. This calculation is achieved by applying the following formula:

$$compound\ score = \frac{x}{\sqrt{x^2 + a}}$$

where, compound score: the sentiment score received of a specific sentence

x: the sum of individual sentiment scores for each word in the specific sentence (as described in detail in Section II)

a: the normalization factor that typically has the value of 15

It is worth noting that in instances where VADER encounters a word, it identifies the sentiment polarity of a word by comparing it with the words contained in its lexicon and have predefined positive, negative or neutral sentiment intensity. Simultaneously, as previously highlighted, VADER considers a range of linguistic and grammatical cues to enhance the precision of sentiment analysis, particularly when handling words that are not explicitly covered in the lexicon. The sentiment score resulting, indicates the overall intensity of sentiment in the sentence and falls within the range of −1 to +1. Fig. 3 shows a representative sample of student responses. The first column refers to the question asked to students, the second column refers to the feedback we received, and then the next three columns refer to the positive, negative, and neutral scores each response received. Finally, the last column "compound" expresses the corresponding score (sentiment intensity) of each student response based on the aforementioned formula.

Question	Students Comment	pos	neg	neu	compound
Q4	He is to the point and ...	0.176	0.059	0.765	0.5964
Q1	I believe he presents th...	0.175	0.066	0.758	0.3182
Q1	I think that students pa...	0.108	0.072	0.821	0.1779
Q1	Your active listening ca...	0.387	0.075	0.538	0.6997
Q2	His persuasion skills ar...	0.148	0.076	0.776	0.2115
Q2	he has told us very poli...	0	0.076	0.924	-0.1027
Q2	The professor uses ever...	0.176	0.088	0.736	0.3678
Q2	His calm way of teachi...	0.333	0.09	0.577	0.5859
Q2	He tried to help us wit...	0.272	0.091	0.638	0.6597

Fig. 3. Sample comments from students with their corresponding scores.

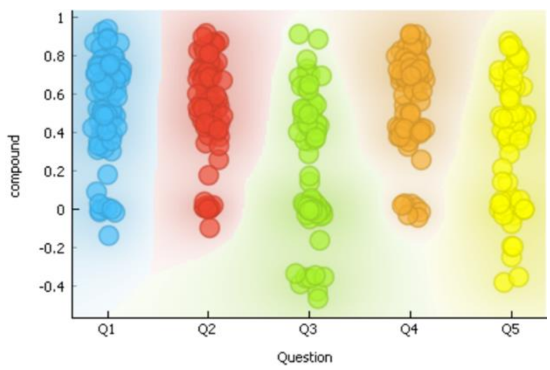
IV. Results and Discussion

Our research was developed in 2 dimensions (cases). In the first, we aimed to investigate and compare the competencies that students perceive in two different professors. In order to ensure a common base, we focused on one course, taught by both professors. Our approach detected varying degrees of competencies of the two teachers, as expected. Fig. 4 illustrates the results extracted, referring to the grade of sentiment received from student comments regarding prof_1 in course_1 and prof_2 in course_1 and course_2. The data is displayed as a collection of points, with the x-axis depicting the question number (in various colors) and the y-axis depicting comment intensity (ranging from −1 to +1) as defined by our system. Note that the compound sentiment intensity of a specific comment corresponds to its combined positive, negative and neutral score according to our model as described in detail in the previous section.

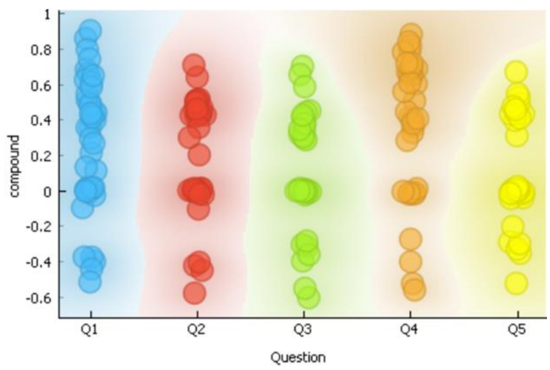
As we delve into the analysis of these results, two distinct dimensions emerge, each providing valuable insights into teacher competencies and student perceptions.

A. Comparative Analysis of Different Professors of the Same Course

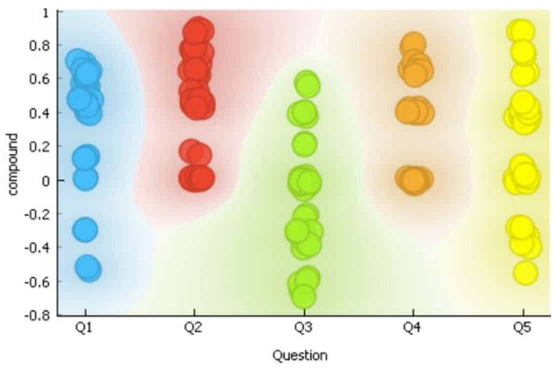
In the first case, our objective was to discern and compare the competencies perceived by students in two different professors who taught the same course. The sentiment analysis results, as depicted in Fig. 4, reveal variations in sentiment intensity across specific competencies. More specifically, Q1, Q2 and Q4, for prof_1, received the most positive responses with the greatest intensity, thus indicating that the corresponding competencies of "*Listening*", "*Persuasion*" and "*Presentation*" were evident in prof_1. On the contrary, the feedback comments referring to "*Empathy*" and "*Collaboration*", had the lowest sentiment score. In addition, for each question, it should be noted that only a few student comments received a negative compound intensity.



Prof_1 in course_1



Prof_2 in course_1



Prof_2 in course_2

Fig. 4. Sentiment intensity of student comments for prof_1 in course_1, prof_2 in course_1 and prof_2 in course_2.

Also, in Fig. 4, it is observed that the scores received by the student comments concerning the competencies of prof_2 in the same course (course_1) are distributed over a wider range of values of intensity, without clearly indicating the existence of the specific competencies concerning prof_2.

Summarizing and comparing the results of the first case, depicted in Fig. 4, concerning the competencies of 2 different professors in the same course (course_1), the students' point of view is clearly established as shown in Fig. 5. Despite the obvious differences between the two professors that the students discerned, as for example in the "Presentation" skill, it is found that in both professors the students "see" the lack of teacher empathy as evident. This finding is notable and in line with research findings from Meyers *et al*. [40] who highlight the need to measure teachers' empathy.

Our findings also extend previous research [3, 41] by directly applying sentiment analysis to evaluate specific teacher competencies in higher education. Our approach, also, differs from Syauqi *et al.* [42], which used data collection from students' feedback to provide an overview of their perceptions through survey methods based on Likert scale questionnaires.

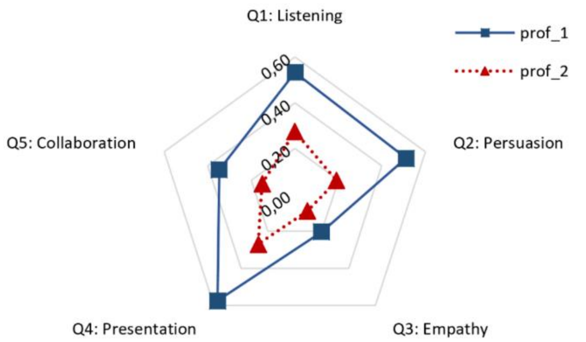


Fig. 5. Mean sentiment score of student comments received, regarding the competencies of 2 different instructors of the same course (course_1).

Next, we presented our data in a cluster matrix using the k-means algorithm. We aimed to group student responses into clusters based on their similarity. The choice of k-means algorithm was made due to its simplicity and interpretability. In addition, the k-means algorithm provides high reliability and efficiency in clustering responses in various domains, such as education [43–45]. Given our data, and guided by the elbow method [46] we implemented the k-means algorithm over a range of *k* values to ensure the optimal choice for the number of clusters. The following heat map diagram (Fig. 6), concerns the comment clustering of the case of prof_1, course_1.

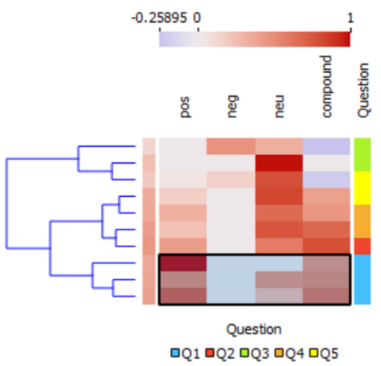


Fig. 6. Data clustering using K-means algorithm.

The far-right column indicates the prevailing question in the specific comment cluster. An interesting finding emerging through the cluster exploration was the nature of the content of each cluster. We found that comments referring to the same question, and in extension to a specific teacher competence, were clustered together, based solely on sentiment and not on question type. For example, it seems clear that the comments related to Q1 were "grouped" together because of the common characteristics detected in the specific comments and not because they were stated to be related to Q1. In other words, the student opinions referring to specific teacher competencies showed a homogeneity in terms of the sentiment they expressed and what they believed about their instructors without an extensive dispersion of opinions regarding the same competence, thus strengthening the value of our findings.

B. Comparative Analysis of the Same Professor Across Different Courses

Case 2, which concerns the same professor in different courses (course_1 and course_2), is also quite interesting. The aim of this case was to investigate whether students perceive one professor's competencies differently depending on the nature of the course. For this reason, we attempted to measure the specific competencies of prof_2 in course_2, in addition to course_1. Fig. 4 depicts the results referring to prof_2 in course_2. As observed, there are questions (Q2 and Q4) that correspond to the "Persuasion" and "Presentation" competencies which received comments with negligible negative sentiment scores, in contrast to the case of prof_2 in course_1. However, a moderate amount of feedback comments referring to "Listening", "*Empathy*" and "*Collaboration*" received negative compound scores, as in case 1 for prof_2 in course_1.

Taking into account the results regarding course_1 and course_2 for prof_2 we found that certain competencies were consistently perceived by students, regardless of the course's nature. More specifically, the competence, or lack thereof, when it comes to empathy expressed, listening and collaboration, is equally evident in students regardless of the nature of the course. Simultaneously, teacher competencies such as powers of persuasion used to inspire students and enhance/boost their interest, as well as presentation competencies, seem to depend on the nature of the course. In order to reinforce this observation, we posed an additional question to the students of courses 1 and 2. We asked them to express their opinion about which of the 2 courses they considered to be more difficult or complex. The results showed an overwhelming difference between the 2 courses, since more than 90% of the students considered course_1 to be more difficult than course_2. We concluded that there are indeed some competencies such as persuasion, presentation, etc. which depend on the nature of the course. For example, according to students, a course that is easy to understand contributes to the emergence of the presentation skill possessed by the professor who teaches it.

This research represents the first known endeavor to employ sentiment analysis on open-ended student feedback, shedding light on the assessment of specific teacher competencies. While recent review papers have thoroughly examined sentiment analysis in education, none have ventured into the domain of teacher competences [47–49]. Notably, sentiment analysis has demonstrated its versatility by finding application in diverse fields related to competence management, including assessing operator skills in manufacturing systems [50], modeling doctor communication skills in the medical domain [51] and addressing ethical considerations, such as evaluating the honesty of lawyers, within the legal domain [52]. This serves as motivation for utilizing sentiment analysis in competency assessment within the education domain, a fact confirmed by our experiments. Utilizing sentiment analysis, our approach efficiently processes streaming data, even with standard computer specifications, without encountering computational limitations. In contrast, most machine-learning approaches present drawbacks concerning big training data sets, significant training and classification time, CPU processing, high memory, etc. In summary, our contribution extends prior research by directly applying sentiment analysis to evaluate teacher competencies in higher education, opening up new avenues for further research and discussion, offering opportunities to refine and improve the assessment of teacher competencies in higher education.

V. Conclusion and Future Work

In this paper, we highlighted the necessity of extracting valuable knowledge from student feedback and exploiting this data with the aim of improving teaching and learning in universities. We also emphasized that student feedback comments regarding open-ended questions are unique but unfortunately remain untapped due to the fact that current analyses are primarily limited to traditional techniques focused on closed-ended questions. As a result, not only do teachers miss out on the opportunity to gain insight into teaching procedures and the unique elements that characterize them in the educational process, but the educational system is also deprived of the potential benefits for all involved.

In this study, we proposed a sentiment analysis approach for mining and quantifying student opinion in order to bridge the aforementioned gap. Following our previous research and using our approach with a dataset from the University of Thessaly, teacher communication competencies were assessed measuring the intensity of sentiment of each student comment in two cases; for two teachers in the same course and for one teacher in two different courses. The analysis of the results, indicated that students expressed a homogeneous and convergent view regarding competencies referring to the same instructor, thus both strengthening their opinion and also providing added weight to our approach. From our research, it was also revealed that teacher competencies such as "Presentation" and "Persuasion" are more evident in courses with a lower degree of difficulty and vice versa in the opposite case. More specifically, according to our research and based on student opinion, a course that is easy for students to understand, contributes to the emergence of the presentation and persuasion competence possessed by the professor who teaches it. Simultaneously, according to the data collected from the students in our study, competencies such as "Listening", "Empathy" and "Collaboration" possessed by a teacher were revealed to be independent of the nature of the course.

While conducting this study, we encountered certain limitations that are important to acknowledge. More specifically, during the data collection process, our research was limited to gathering data of feedback from students of our university on two courses and two professors. Although an expanded collection of student feedback including multiple courses with different characteristics such as semester, cognitive area, degree of course difficulty, etc., would require additional time and effort, we believe it could contribute to revealing important findings and conclusions. In addition, the language in which the students expressed themselves was their mother tongue, Greek, in order for their responses to convey meaning and sentiment as accurately as possible. Although from a technical point of view in our model, we had the ability to rely on a multilingual sentiment lexicon, we preferred the validity of the VADER lexicon. Because this choice limited us to the use of the English language, in order to achieve maximum accuracy and efficiency, all the student comments initially collected in the Greek language, were subsequently translated into English by accredited translators and interpreters of Greek and English.

Looking ahead, we plan to conduct extensive research on various methods of measuring and leveraging student feedback using machine learning algorithms. We strongly believe that through our research we can contribute to how institutions face the challenge of harnessing this vast amount of data, with the ultimate goal of enhancing the teaching and learning processes.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, C.D. and P.F.; methodology, C.D., O.I. and P.F.; software, C.D. and O.I.; formal analysis, C.D. and A.K.; investigation, C.D. and P.F.; resources C.D., P.F. and A.K; data curation, O.I.; writing—original draft preparation, C.D. and P.F.; writing—review and editing, C.D., O.I. and P.F.; visualization, P.F.; supervision, A.K.; project administration, P.F.; All authors have read and agreed to the published version of the manuscript.

References

[1] L. Zeng, Z. Tan, L. Xia, Y. A. Xiang, and Y. Ke, "Behavior analytics, sentiment analysis, and topic detection of Danmaku from online electronics courses on Bilibili," *International Journal of Information and Education Technology*, vol. 13, no. 2, 2023.

[2] R. M. Mayordomo, A. Espasa, T. Guasch, and M. Martínez-Melo, "Perception of online feedback and its impact on cognitive and emotional engagement with feedback," *Education and Information Technologies*, vol. 27, no. 6, pp. 7947–7971, 2022.

[3] M. Misuraca, G. Scepi, and M. Spano, "Using opinion mining as an educational analytic: An integrated strategy for the analysis of students' feedback," *Studies in Educational Evaluation*, vol. 68, 100979, 2021.

[4] L. A. T. Nguyen and A. Habók, "Tools for assessing teacher digital literacy: a review," *Journal of Computers in Education*, pp. 1–42, 2023.

[5] K. Rybinski, "Exploring the influence of student emotions and professor behaviour on course ratings: A quantitative analysis," *Quality Assurance in Education*, 2023.

[6] J. T. Richardson, "Instruments for obtaining student feedback: A review of the literature," *Assessment & Evaluation in Higher Education*, vol. 30, no. 4, pp. 387–415, 2005.

[7] B. T. M. Wong and K. C. Li, "A review of learning analytics intervention in higher education (2011–2018)," *Journal of Computers in Education*, vol. 7, no. 1, pp. 7–28, 2020.

[8] J. McDonald, A. C. M. Moskal, A. Goodchild, S. Stein, and S. Terry, "Advancing text-analysis to tap into the student voice: A proof-of-concept study," *Assessment & Evaluation in Higher Education*, 2019.

[9] F. Su, D. Zou, L. Wang, and L. Kohnke, "Student engagement and teaching presence in blended learning and emergency remote teaching," *Journal of Computers in Education*, pp. 1–26, 2023.

[10] M. Lane and D. Meth, "Exploring impacts on students as givers of teaching feedback," *Quality Assurance in Education*, vol. 29, no. 2/3, pp. 225–237, 2021.

[11] J. Flodén, "The impact of student feedback on teaching in higher education," *Assessment & Evaluation in Higher Education*, vol. 42,no. 7, pp. 105420171068, 2017.

[12] L. Mandouit, "Using student feedback to improve teaching," *Educational Action Research*, vol. 26, no. 5, pp. 755–769, 2018.

[13] C. Marley, A. D. Faye, E. Hurst, J. Moeller, and A. Pinkerton, "Moving beyond 'You Said, We Did': Extending an ethic of hospitality to the student feedback process," *Online Postgraduate Education in a Postdigital World: Beyond Technology*, pp. 1–19, 2021.

[14] Z. Nasim, Q. Rajput, and S. Haider, "Sentiment analysis of student feedback using machine learning and lexicon based approaches," in *Proc. 2017 international conference on research and innovation in information systems (ICRIIS)*, IEEE, July 2017, pp. 1–6.

[15] H. W. Marsh, "Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness,"

The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective, pp. 319–383, 2007.
[16] A. Kumar and R. Jain, "Sentiment analysis and feedback evaluation," in *Proc. 2015 IEEE 3rd International Conference on MOOCs, Innovation and Technology in Education (MITE)*, IEEE, 2015, pp. 433–436.
[17] K. Sangeetha and D. Prabha, "Sentiment analysis of student feedback using multi-head attention fusion model of word and context embedding for LSTM," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 4117–4126, 2021.
[18] S. Slater, S. Joksimović, V. Kovanovic, R. S. Baker, and D. Gasevic, "Tools for educational data mining: A review," *Journal of Educational and Behavioral Statistics*, vol. 42, no. 1, pp. 85–106, 2017.
[19] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1–167, 2012.
[20] C. Salazar, E. Montoya-Múnera, and J. Aguilar, "Sentiment analysis in learning resources," *Journal of Computers in Education*, pp. 1–26, 2022.
[21] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. the International AAAI Conference on Web and Social Media*, May 2014, vol. 8, no. 1, pp. 216–225.
[22] T. Shaik, X. Tao, C. Dann, H. Xie, Y. Li, and L. Galligan, "Sentiment analysis and opinion mining on educational data: A survey," *Natural Language Processing Journal*, 100003, 2022.
[23] C. Welch and R. Mihalcea, "Targeted sentiment to understand student comments," in *Proc. COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, December 2016, pp. 2471–2481.
[24] G. I. Nitin, G. Swapna, and V. Shankararaman, "Analyzing educational comments for topics and sentiments: A text analytics approach," in *Proc. 2015 IEEE Frontiers in Education Conference (FIE)*, IEEE, October 2015, pp. 1–9.
[25] W. Wang and J. Wu, "Notice of Retraction: Emotion recognition based on CSO&SVM in e-learning," in *Proc. 2011 Seventh International Conference on Natural Computation*, IEEE, July 2011, vol. 1, pp. 566–570.
[26] A. Ortigosa, J. M. Martín, and R. M. Carro, "Sentiment analysis in Facebook and its application to e-learning," *Computers in human Behavior*, vol. 31, pp. 527–541, 2014.
[27] H. Kang, S. J. Yoo, and D. Han, "Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews," *Expert Systems with Applications*, vol. 39, no. 5, pp. 6000–6010, 2012.
[28] C. C. Chen and Y. D. Tseng, "Quality evaluation of product reviews using an information quality framework," *Decision Support Systems*, vol. 50, no. 4, pp. 755–768, 2011.
[29] Y. Al Amrani, M. Lazaar, and K. E. El Kadiri, "Random forest and support vector machine based hybrid approach to sentiment analysis," *Procedia Computer Science*, vol. 127, pp. 511–520, 2018.
[30] M. Kabir, M. M. J. Kabir, S. Xu, and B. Badhon, "An empirical research on sentiment analysis using machine learning approaches," *International Journal of Computers and Applications*, vol. 43, no. 10, pp. 1011–1019, 2021.
[31] V. Balakrishnan and E. Lloyd-Yemoh, "Stemming and lemmatization: A comparison of retrieval performances," *Lecture Notes on Software Engineering*, vol. 2, no. 3, 2014.
[32] D. Khyani, B. S. Siddhartha, N. M. Niveditha, and B. M. Divya, "An interpretation of lemmatization and stemming in natural language processing," *Journal of University of Shanghai for Science and Technology*, vol. 22, no. 10, pp. 350–357, 2021.
[33] L. S. Xun, S. Gottipati, and V. Shankararaman, "Text-mining approach for verifying alignment of information systems curriculum with industry skills," in *Proc.* 2015 *International Conference on Information Technology Based Higher Education and Training (ITHET)*, June 2015, IEEE, pp. 1–6.
[34] C. Dervenis, P. Fitsilis, and O. Iatrellis, "A review of research on teacher competencies in higher education," *Quality Assurance in Education*, vol. 30, no. 2, pp. 199–220.
[35] A. R. F. Shafana and S. M. Safnas, "Does technology assist to continue learning during pandemic? A sentiment analysis and topic modeling on online learning in south Asian region," *Social Network Analysis and Mining*, vol. 12, no. 1, 65, 2022.
[36] M. Wook, S. Vasanthan, S. Ramli, N. A. M. Razali, N. A. Hasbullah, and N. M. Zainudin, "Exploring students' feedback in online assessment system using opinion mining technique," *International Journal of Information and Education Technology*, vol. 10, no. 9, pp. 664–668, 2020.
[37] R. Sharma, S. Nigam, and R. Jain, "Opinion mining in Hindi language: a survey," arXiv preprint, arXiv:1404.4935, 2014.
[38] J. Demšar, T. Curk, A. Erjavec, Č. Gorup, T. Hočevar, M. Milutinovič, and B. Zupan, "Orange: data mining toolbox in Python," *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 2349–2353, 2013.
[39] A. Amin, I. Hossain, A. Akther, and K. M. Alam, "Bengali VADER: A sentiment analysis approach using modified vader," in *Proc. 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, IEEE, February 2019, pp. 1–6.
[40] S. Meyers, K. Rowell, M. Wells, and B. C. Smith, "Teacher empathy: A model of empathy for teaching for student success," *College Teaching*, vol. 67, no. 3, pp. 160–168, 2019.
[41] L. F. Gómez and M. G. Valdés, "The Evaluation of Teacher Performance in Higher Education," *Journal of Educational Psychology-Propositos and Representaciones*, vol. 7, no. 2, pp. 499–515, 2019.
[42] K. Syauqi, S. Munadi, and M. B. Triyono, "Students' perceptions toward vocational education on online learning during the COVID-19 pandemic," *International Journal of Evaluation and Research in Education*, vol. 9, no. 4, pp. 881–886, 2020.
[43] R. Nainggolan, F. A. T. Tobing, and E. J. Harianja, "Sentiment; Clustering; K-Means Analysis Sentiment in Bukalapak Comments with K-Means Clustering Method," *JNMT (International Journal of New Media Technology)*, vol. 9, no. 2, pp. 87–92, 2022.
[44] D. Saputra, H. Haryani, A. Junaidi, T. Baidawi, and A. Surniandari, "Application of K-mean clustering algorithm in grouping data prospective new students," in *Proc. AIP Conference*, AIP Publishing, May 2023, vol. 2714, no. 1.
[45] H. Xie, L. Zhang, J. Sun, and X. Yu, "Application of K-means clustering algorithms in news comments," in *Proc. 2010 International Conference on E-Business and E-Government*, IEEE, May 2010, pp. 3759–3762.
[46] K. U. Sarker, M. Saqib, R. Hasan, S. Mahmood, S. Hussain, A. Abbas, and A. Deraman, "A ranking learning model by k-means clustering technique for web scraped movie data," *Computers*, vol. 11, no. 11, 158, 2022.
[47] K. Mite-Baidal, C. Delgado-Vera, E. Solís-Avilés, A. H. Espinoza, J. Ortiz-Zambrano, and E.Varela-Tapia, "Sentiment analysis in education domain: A systematic literature review," in *Proc. International conference on technologies and innovation*, Cham: Springer International Publishing, October 2018, pp. 285–297.
[48] J. Zhou and J. M. Ye, "Sentiment analysis in education research: A review of journal publications," *Interactive Learning Environments*, vol. 31, no. 3, pp. 1252–1264, 2023.
[49] Pooja and R. Bhalla, "A review paper on the role of sentiment analysis in quality education," *SN Computer Science*, vol. 3, no. 6, 469, 2022.
[50] D. Mourtzis, J. Angelopoulos, V. Siatras, and N. Panopoulos, "A methodology for the assessment of Operator 4.0 skills based on sentiment analysis and augmented reality," *Procedia CIRP*, vol. 104, pp. 1668–1673, 2021.
[51] T. Sen, M. R. Ali, M. E. Hoque, R. Epstein, and P. Duberstein, "Modeling doctor-patient communication with affective text analysis," in *Proc. 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, IEEE, October 2017, pp. 170–177.
[52] N. Mehrabi, P. Zhou, F. Morstatter, J. Pujara, X. Ren, and A. Galstyan, "Lawyers are dishonest? quantifying representational harms in commonsense knowledge resources," arXiv preprint, arXiv:2103.11320, 2021.