Sentiment Analysis of Social Media for Evaluating Universities

Anas Abdelrazeq Daniela Janssen Christian Tummel Sabina Jeschke Anja Richert
IMA/ZLW & IfU at RWTH Aachen University
Dennewartstrasse 27, 52068 Aachen, Germany
Anas.Abdelrazeq@ima-zlw-ifu.rwth-aachen.de, Daniela.Janssen@ima-zlw-ifu.rwth-aachen.de,
Christian.Tummel@ima-zlw-ifu.rwth-aachen.de, Sabina.Jeschke@ima-zlw-ifu.rwth-aachen.de,
Anja.Richert@ima-zlw-ifu.rwth-aachen.de

ABSTRACT

In the age of digitalization, a huge amount of sentiments are expressed daily on university related topics using social media platforms. Particularly, posted statements from students and teachers can provide a potential source for evaluating universities. Twitter as one of the most popular microblogging platforms is a rich data resource for opinion mining. Stimulated by this fact, ways to analyze Twitter for information in the context of universities are sought. This paper looks at the analysis of social media sentiment as a complementary source for evaluating universities. The extracted results can support university rankings that experience criticism in terms of measuring vital indicators. This paper relays on sentiment analysis methods to analyze opinions published on Twitter. For this purpose, at first, tweets that are related to selected universities in Germany were collected. Second, the tweets were classified based on their sentiment into “Positive” and “Not Positive” tweets. At last, the results were analyzed providing information about the communicative topics at the universities. This paper gives an outlook to further research in context of an automated analysis of social media content in order to support the evaluation of universities.

KEYWORDS

Social Media, Universities, Higher Education, Twitter, Sentiment Analysis, Data Analysis, Natural Languages Processing, Learning Environments

1 INTRODUCTION

Nowadays, social media platforms form a substantial data source for opinions on any topic. This include statements that are related to universities’ topics and events. Particularly, microblogs appeared as one of the most common social media tools, in which, countless users can participate and interact at any time. Using microblogging platforms, content from every user can be published, read, commented, linked or forwarded as desired. As a result, an open communication medium is established. This open exchange with the participation of different users helps to bridge and interlock formal and informal learning contexts. Consequently, microblogging is a medium that offers real-time communication of its users thoughts and ideas.

Twitter is one of the most opinion-rich resources, where huge amounts of opinions on different topics are expressed. Such opinion-rich data can be used for extracting and analyzing the opinions in terms of specific questions. In this context, data mining techniques have the potential to detect opinions from large amounts of data. With the help of existing tools from machine learning and natural languages processing fields, concepts of sentiment analysis can be applied on the collected data from social media. This can provide vital measurements and insights over the communicative topics in the university environment.

Analyzing students and teachers opinions as well as comparing and evaluating universities is important on different scales. On the one hand, universities would like to have a performance measurement mechanism for improvement and adjusting plans. On the other hand, students relay on these comparisons to support their decision for joining a specific university to proceed with their studies.

The standardized process of comparing and evaluating universities is based on university rankings on a national and international level. It operates on the basis of certain criteria (i.e.

indicators). Recently, university rankings experienced a lot of criticism in terms of: measurement accuracy, measuring the university as a whole institution, and the way data is being collected for measuring specific indicators at universities.

Besides the official standardized university rankings, alternative and complementary approaches must be investigated. Standard rankings try to search for statements from teachers and students that are related to their experiences at the universities. A promising addition lies in extending the search to include new platforms such as social media.

The main contribution of this papers is to bring sentiment analysis tools in analyzing social media content - especially microblogs- to act as a supportive indicator for evaluating universities.

In section 2, related work is presented. The paper concept of using social media as supplementary tool for evaluating university is discussed in section 3. Afterwards, the paper method used for data processing is presented in section 4. Later on, the results and discussion on using sentiment analysis for Twitter data in university context are shown in section 5. Finally, a conclusion and outlook are presented in section 6.

2 RELATED WORK

This paper aims at investigating new indicators for universities evaluation and comparison. Such indicators can be provided by applying sentiment analysis concept and methods on the available data from social media. In the upcoming sections, the related work for using social media platform at universities, evaluating universities by university rankings and sentiment analysis concepts are presented.

2.1 Social Media at Universities

Recently, social media platforms have become a main medium for people to express their daily activities, reactions and emotions. Blogs and microblogs are the most common form of social media [1]. Blogs (the abbreviated form of Weblog) are informal sites on the world-wide web where users are used to post (i.e. publish) ideas, discussions, thoughts, etc. [2]. While, microblogs are smaller blogs with short posts up to a limited number of signs [3]. One thing they have in common is that they consist of entries that are listed in a chronologically descending order (i.e. the latest news is on top). They are tools that enable discussions and comments on information shared with other users. They are characterized by its dynamic and up-to-datedness.

Students use social media platforms and internet on a daily basis, often times more frequently than other mass media such as newspaper or television. Therefore, they are experienced in how to use social media [4]. Based on this fact and the increasingly usage of social media, it seems reasonable to use social media in university environments for the purpose of engaging the social media tools in the teaching process, as well as for the purpose of analyzing opinions regarding university topics like teaching and learning. On university level, microblogs are tools for simultaneous communication, knowledge management and publication service. In formal and informal teaching and learning contexts, microblogs can support individual and cooperative communication, knowledge management, reflection and feedback processes.

Teachers use microblogs to encourage motivation and participation of students. This is justified by two features [5]: First, it raises interactivity of students and creates the opportunity to implement social and team building aspects even in large classes. Second, it makes use of the media usage of smartphones, tablets and notebooks of today's students in learning contexts [1]. The fields of application of microblogs in teaching scenarios can be divided into three parts: collaboration, feedback-giving/discussion and public scientific communication. By adding the ability to share information and knowledge via social media, mentioning Twitter specifically, collaborative work and learning is enabled with an open feedback-giving space.

Furthermore, general and personal questions as well as lecture or seminar contents can be reflected. Twitter can be utilized to collect questions and feedback of students and discuss them [6]. One of the interesting use cases is the "Twitter wall", where the concept is realized by projecting Twitter's posts on a big screen during the class time [7]. In addition to discussing the most relevant questions at the end of the lecture, other questions and feedback can also be answered after the lecture via Twitter [6].

Besides the well-known applications of microblogs in different teaching and learning scenarios, semantic technologies are often put to use in order to analyze and evaluate automatically opinions made in social media, e.g. Twitter or Facebook. Brauer and Bernroider [8] conduct an international study analyzing the usage of Facebook within higher education institutes in Germany, Austria and Switzerland. In this case, the social media strategy that is related to Facebook of selected universities is analyzed. Another field of social media analytics in universities is the analysis of students and teachers opinions which are made in e.g. Twitter or Facebook [1].

2.2 Evaluating Universities via University Rankings

The comparable analysis of universities worldwide has been established over the past few decades. National and international university rankings try to compare and evaluate universities through various criteria and try to make differences in quality measurable.

According to a user study from the Center for Higher Education Development (CHE), the most common reason to make use of university rankings is to acquire information about universities current or potential status [9]. Therefore, the main goal of any university ranking is to process and present data in order to make comparing universities with each other clearer. It also helps in the decision making process towards the future of teaching and learning environment based on the data collected and information provided [10].

Via comparing the available study programs and conditions of universities, the transparency of performance is improved [11]. University rankings have to meet certain methodical standards to be helpful for the decision making process of students, as well as an orientation tool for university's improvements. At first, university rankings have to be disciplinary because universities are versatile and have different foci [11]. Another important requirement for university rankings is being multidimensional. Thus, other indicators, such as timetables, external funding or the state of the university's library as examples, have to be put in comparison.

On an international level, the most popular rankings in Germany are the "Ranking of World Universities" by the magazine "Times Higher Education (THE)", as well as the "QS World University Ranking" [9]. The THE reviews research-led universities across all their core missions in teaching, research, knowledge transfer and international outlook using 13 indicators. The "QS World University Ranking" compares universities worldwide on a basis of eight indicators. In addition to the state of research, publications and Nobel prizes are also emphasized in the QS. The QS ranking does not only compare universities on a global level, but also individually for each country or region.

Even though rankings have become an increasingly popular way to compare higher education performance and productivity, university rankings have been criticized over the recent years in three major aspects [12][13].

First, most rankings measure universities as whole institution. Thereby, only the average quality of a university is measured. Individual subjects are not taken into account. Based on weighing single indicators a total value (i.e. "composed indicator") is calculated [12].

The second aspect relates to the way different aspects and dimensions have been measured in some indicators, and the way the related data is collected. In total, some indicators (e.g. university reputation) have a higher influence value than some others. [12].

Third, the measurement of educational quality such as the quality of teaching and learning or the quality of student experience is underrepresented in the existing ranking [13].

2.3 Sentiment Analysis

Sentiment analysis is a method that analyzes how opinions, reactions, impressions, emotions and perspectives are expressed in a language. Its algorithms can extract evaluative information from large text databases and summarize it [14].

In order to analyze the opinion of people and customers, sentiment analysis appears as the main tool in different contexts. As an example, sentiment analysis has been used to measure customers satisfaction via statements they comment on a specific product they bought or a service they were delivered [15]. It also appeared in the extend of detecting different opinions regarding political events such as elections [14].

Sentiment analysis methods are well developed in the domain of blogs and product reviews [16] [17]. Researchers have been working on detecting sentiment in text via presenting different algorithms for detecting semantic orientation [18].

In favor of producing meaningful information from tweets, sentiment analysis has been also used [14] [15] [19]. Different features selection techniques have been investigated, establishing a comparison between different ones such as n-grams, part of speech, lexicons, etc. [20]. Besides, different classifiers with their learning performance have been tested in different contexts [21].

This paper applies the existing developed approaches in sentiment analysis to microblogging platforms data such as Twitter in order to explore complimentary resources for university evaluation and comparison.

3 TWITTER SENTIMENT ANALYSIS FOR EVALUATING UNIVERSITIES IN GERMANY

This paper suggests that social media content is a vital source for collecting feedback and reactions on the daily events and activities that is related to universities. To prove this hypothesis, a case study is established which evaluates the reactions and feedback from the social media data that is related to nine universities in Germany that are part of the TU9[1] German Institutes of Technology. The TU9 universities are: RWTH Aachen University, TU[2] Berlin, TU Braunschweig. TU Darmstadt, TU Dresden, Leibniz Universität Hannover, Karlsruher Institut für Technologie, TU München and Universität Stuttgart.

3.1 Data Collection

Out of many social media platforms, Twitter is the most popular microblogging platform [22]. Therefore, it was chosen to be the source of the social media data for this paper.
Students in universities message about anything and everything in their day-to-day lives. Twitter users post about their reactions and feedback in a form of microblogs, each is a text up to 140 letters that is called a tweet. The tweets that are related to the TU9 form the data set.
The Twitter API[3] enables different ways to search and filter the tweets. In the literature, different ways for extracting the tweets are adapted. For example, Pak and Paroubek [23] and Bifet and Frank [24] have extracted the tweets based on specific mentioned emoticons. Others, such as Davidov et al. [25], have extracted the tweets based on a list of mentioned hashtags in the tweets.
To collect tweets which are related to the TU9, all tweets which have a matching word from a list of keywords were extracted via Twitter

[1]TU9 is an incorporated society of the nine most prestigious, oldest, and largest universities focusing on engineering and technology in Germany.

[2]TU: Technical University

[3]API: Application Programming Interface

API. The keyword list contains combinations of the universities names and titles.
The data extraction script began running on October 1st, 2014 till March 31st, 2015. In the German academic calendar, this time span is the entire winter semester of 2014/2015. The script catches every tweet that matches any of the keywords and saves it in a database. As a result, the script collected 16488 tweets.
The collected tweets were posted in both English and German. Also, it included original tweets and the re-posted (i.e. retweeted) ones. Table 1 shows how tweets are divided over the TU9 [4].

Table 1. Tweets count for each of the TU9.[4]

University	Count
RWTH Aachen	8073
TU Dresden	4075
Universität Stuttgart	1014
TU Darmstadt	951
Karlsruher Institut für Technologie	767
Leibniz Universität Hannover	450
TU Braunschweig	445
TU Berlin	401
TU München	300

Table 2. Tweets count for each language.

Language	Count
German	12906
English	3582

Considering the fact that German universities - especially the TU9 - rely on multiple languages for communication and enroll a large number of international students, it makes sense to keep the tweets in both English and German. Table 2 shows the number of collected tweets in each language. Retweets begin with ”RT” and are mostly copies of other original tweets, with some possible text at the beginning [26]. Still, retweets are useful to be considered as they emphasize the statement and facts included in the original tweet. Table 3 shows the number of original and retweeted tweets in the data set.

[4]Few tweets belong to more than one university, that leads to have the tweets summation larger than the total count of collected tweets.

Table 3. Number of original and retweeted Tweets.

Tweet Type	Count
Original Tweets	10189
Retweets	6287

3.2 Defining Tweets Sentiment

This paper's approach classifies the sentiment of each tweet to be either “Positive” or “Not Positive”. This is known as a two way sentiment classification [15].
A “Positive” tweet refers to text that indicates a positive statement regarding an event such as a lecture, class or activity that is related to one of the TU9. A “Not Positive” tweet can be either a negative statement regarding an event, or a neutral one, such as an announcement or advertisement regarding an event in the university.
Adapting two way classification can be considered as a limitation. Nevertheless, it is easier to process in the classifier learning step. The same approach was adapted by Go et al. [20] who consider that the “Not Positive” tweets are actually “Negative” ones, ignoring the neutral nature of some tweets. Pak and Paroubek [23] proved that adapting a three way classification leads to bad performance which can be avoided by the two way classification [21].

3.3 Tweets Sentiment Analysis Challenges

Dealing with social media as a source of information - especially microblogging platforms such as Twitter - adds extra difficulties to the sentiment analysis process [19]. Tweets are plain text written in an informal manner and its processing face challenges such as:

- Length: Tweets have a limited text length, which is 140 characters. This forces users to start using some common and uncommon abbreviations and phrases. As an example, abbreviations such as OMG[5], WTH[6], DKDC[7], TY[8], etc. appears often in Twitter.

[5]OMG: Oh My God!
[6]WTH: What the hell!
[7]DKDC: Don't know, don't care
[8]TY: Thank you

- Informality: Twitter is mostly used as a non-formal communication medium. This leads to many informal statements which probably contain errors such as misspellings, unstructured sentences and slang. Informality may also infer sarcasm, which adds an extra layer of difficulty in guessing the right sentiment of each tweet.

- Credibility: This paper's approach of gathering the tweets is based on a list of keywords. This does not guarantee the credibility of who and what tweets are generated on Twitter. This leaves the possibility that one anonymous user has generated all the content about a specific university with different usernames, rather than the students.

- Data availability: collecting the right data is always a challenge, but having enough data is another critical issue. The target data for this study is very specific, which can be problematic for collecting data over a six-month period of time. It is significant to note that more data leads to more trusted results.

4 DATA PROCESSING

Figure 1 illustrates the main steps of the data processing. The starting point is a set of tweets which was extracted via Twitter API. Based on sentiment analysis approach, a sentiment classifier will be build by learning from previously annotated subset of tweets in order to classify the rest of tweets. The classifier to be built will be able to learn the defined sentiment: "Positive" and "Not Positive". The processing steps are divided into three main steps: Tweets' text filtering, feature extraction, and sentiment classification.

4.1 Tweets Text Filtering

As previously mentioned, tweets are informal sentences that have to pass through a filtering stage before it can be processed for the up coming steps. Filtering is the process of cleaning

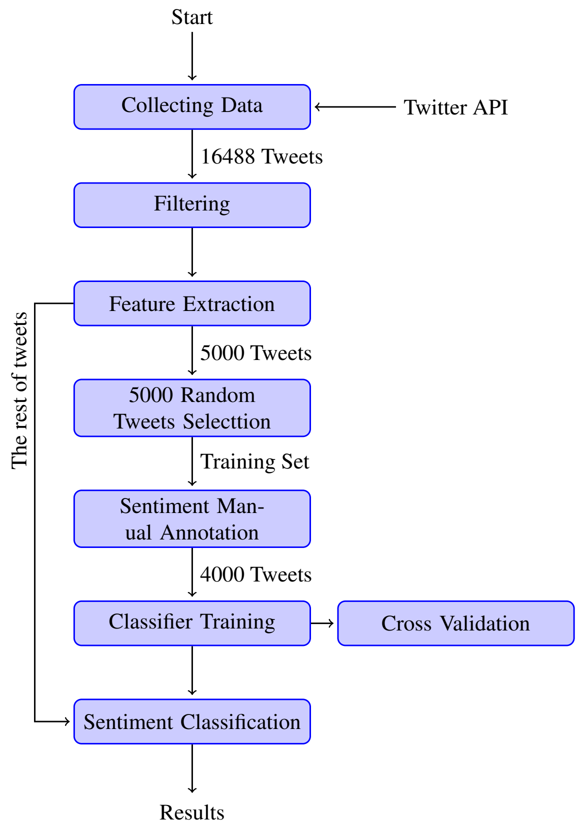

Figure 1. Data processing flow

the tweets text removing all irrelevant text for the sentiment classifier learning step. The following are tweets filtering steps mentioned in the order they were performed:

1. All text is switched to lowercase including those words which are completely capitalized. Despite the fact that some users tend to emphasize specific words with capitalization, this was not the general case with the collected tweets. Many names and sentences are found completely capitalized indicating no emphasizing on the meaning. Taking the capitalization into account in such cases might lead to false results. Therefore, all text is changed to lowercase.

2. All hyperlinks are removed. Tweets mostly contain hyperlinks to other sites and photos which does not contribute to the sentiment of the tweet.

3. All mentioned usernames (identified by words that start with @) are removed and

ISBN: 978-1-941968-26-0 ©2015 SDIWC

all hashed words with the # symbol are replaced with the word itself. These specific symbols and markups mentioning usernames or include hashed words that tag a place, name, etc. are so general to contribute to a specific tweet sentiment.

4. The "RT" text which indicates a retweet is removed.

5. Repeated lettered are filtered. Often, users emphasize words by repeating letters such as: "I am Happyyyy". Go et al. [20] suggest to remove out repeated letters leaving only two of them. This also guarantees that words such as "cool" with original double letters are left unaffected.

6. Common emoticons are replaced with their semantic. Emoticons are often used in social media language to indicate the users' emotions [23]. The found emoticons are classified as:

   - Happy emoticons: ":)", ":-)", ":D", ";)", ";']", etc. which are replaced by "HAPPY_FACE".
   - Sad emoticons: ":-(", ":(", "=(", ";(", ":[", etc. which are replaced by "SAD_FACE"

7. Negations are detected in the tweet. Depending on the language, negation appears in different forms. Accordingly, the sentiment of the words appear before and after the negation are changed. For example, "I don't like exams" is changed to "I NOT_ do NOT_ Like exams".

8. All words which do not start with a letter are removed. This eliminates all phone numbers and dates included in the tweet.

9. Extra spaces and punctuation marks are removed.

10. All stop words and keywords (including the universities' names) are removed based on the language of the tweet.

After these steps, every tweet text is left only with words that can play a role in indicating the sentiment of the tweet. Here is an example of a tweet after applying all filtering steps:

"I would tell you. That I loved you. If I thought ... ♫ Boys Don't Cry by Grant-Lee Phillips (at RWTH Bibliothek 2) https://t.co/qGoJm1bmV8"

"tell loved thought boys NOT_ do NOT_ cry grantlee phillips bibliothek"

4.2 Features Selection and Extraction

An important part of the sentiment analysis process is features selection. Features are the sentence properties that are analyzed in an attempt to correlate it to the tweet sentiment (i.e. "Positive" or "Not Positive"). A feature can be the fact that the tweet contains a word, emoticon, a combination of words, etc. The selection of the features is very important as they act as the input for the classifier in the next step.

Different features selection approaches appear in the literature in the context of tweets sentiment analysis. Some algorithms use unigrams (i.e. single words), some others use bigrams or trigrams (i.e. two or three consecutive words, respectively) [20]. Also, many algorithms use part-of-speech tags and lexicons [19]. Hashtags and emoticons also appeared as features in some algorithms [25].

Unigram features give a wide coverage for the tweet's text. On the other hand, n-grams with $n \geq 2$ show a higher ability to capture the sentiment expression patterns. As larger as n gets, the sentiment is more specific. N-grams with $n \geq 2$ are used to find out the domain specific language. This is useful for this research, as the aim is to build a set of vocabulary that is related to the universities context.

Depending on the context, the chosen size of n-grams affects differently. For example, unigrams are a better choice than bigrams when performing the sentiment classification of movie reviews [16]. On the contrary, bigrams and trigrams worked better for the product review polarity classification [27]. In gen-

eral, using only bigrams as features is not useful because the feature space becomes very sparse [20].
The best settings of the tweets features has to be determined. This paper adapts a combination of uni- and bigrams features to benefit from the unigrams' coverage of the data and the bigrams' ability to capture the sentiment expression patterns. The same approach was adapted by Pak and Paroubek [23].
Using part-of-speech features leads to a drop in the sentiment classification performance. These results were proven by Go et al.[20] and Kouloumpis et al. [19]. Therefore, the part-of-speech features were neglected.
The sentiment of the emoticons can be part of the n-grams features. On the one hand, stripping out the emoticons from the tweets leads the classifier to learn from other words that forms unigrams and bigrams [20]. On the other hand, including the emoticons data showed a performance improvement referring to the results of Kouloumpis et al. [19].
Based on the filtering step, emoticons are replaced by their sentiment (i.e. "HAPPY_ FACE" and "SAD_ FACE"). In the features extraction step, they take part in forming the unigrams and bigrams features of the tweets. This leads to a better performance for the classifier [20].

4.3 Sentiment Classification

Different classifiers have been presented in the literature for Twitter sentiment analysis. Supervised classifiers are the focus. They require a training set to be prepared forehead.
The training set has to be annotated. Pak and Paroubek [23] did this automatically based on the fact that all tweets in their dataset contains emoticons. They labeled each tweet based on the emoticon sentiment to be either "Positive" or "Negative".
For the training set, it got annotated manually by different people based on their own feeling whether the tweet indicate a "Positive" or "Not Positive" sentiment. Automatic labeling was not possible at this stage as the tweets lack a common feature for sentiment labeling.
Several classifiers can be used when it comes to the tweets sentiment analysis. Mainly, three common classifiers in the field of machine learning have been used in the literature: Naive Bayes classifier, Support Vector Machines (SVM), and Maximum Entropy.
Naive Bayes and SVM have been compared by Pak and Paroubek [23] and Go et al. [20] [21]; Naive Bayes has performed better.
Theoretically, Maximum Entropy performs better than Naive Bayes as it handles feature overlap better. However, in practice, Naive Bayes showed better performance on a variety of problems [20].
Naive Bayes classifier is adapted by this paper's approach. It is a common method for text categorization. It appeared often for solving the problem of determining the category or class of documents that belongs to using word frequencies as the features.
In machine learning, Naive Bayes classifier belongs to the family of probabilistic classifiers based on applying Bayes' theorem with the assumption that features are conditionally independence from each other given a specific class.

$$P(s|f) = \frac{P(s).P(f|s)}{P(f)} \tag{1}$$

Equation 1 shows the basic formula of Naive theorem where s is the sentiment class (i.e. "Positive" or "Not Positive") and f is a specific feature. This equation computes the probability of having a tweet with the sentiment s when it contains the feature f. It is calculated based on the probability of having a specific sentiment, probability of the feature existence in all tweets, and the probability of finding the feature in the tweets that belongs to that specific sentiment.

5 RESULTS AND DISCUSSION

The data set holds 16488 tweets. Each tweet contains a statement regarding a university or more from the TU9 in Germany. For the training set, 5000 tweets were chosen randomly and got annotated manually by one sentiment either a "Positive" tweet or "Not Positive" tweets.

From the 5000 tweets, 4000 tweets where chosen randomly divided equally between 2000 "Positive" and 2000 "Not Positive" tweets. They are the input for the training step of the Naive Bayes classifier (see Figure 1).
The results section evaluates three main aspects of the presented method:

1. Measuring the classifier efficiency based on the suggested filtering and features extraction steps.
2. Establishing a comparison between the TU9 based on each university's tweets trying to prove the hypothesis that social media content may act as an indicator for university comparisons.
3. Investigating the tweets sentiment on daily basis for each university to obtain feedback on different events and activities.

Each is presented in the following sections.

5.1 Classifier Efficiency

Each tweet has been processed through the filtering step. Then, its unigram and bigram features were extracted. For the purpose of cross checking validation, 25% (i.e. 1000) of the training set tweets were chosen randomly leaving out 3000 tweets for training. The 1000 tweets are used for testing the classifier performance, which were excluded from the training. A script was written in Python using the Natural Languages Tool Kit (NLTK) python library. The NLTK classifier was used for performing this paper's approach. Part of the results of the trained classifier are shown in Table 4. The table shows the most informative features which were learned by the classifier. In other words, each listed feature was more found several times for the corresponding feature class. The results show that some of these features have a dominant effect on the classification of the tweet. As an example, whenever the word "Pegida[9]" is mentioned, tweets are probably "Not Positive" 27 more times than being "Positive". On the other hand, whenever the word "Glückwunsch[10]" is mentioned, tweets are probably "Positive" 16.3 more times than being "Not Positive".
It can be also noticed that both unigrams and bigrams played a role as the most informative features.
The classier performance was verified with the rest of the annotated tweets (i.e. 1000 tweets). 73.6% of the tweets sentiment were guessed correctly by the classifier. This result is considered to be good considering the data size and compared to the other developed classifier by Pak and Paroubek [23] and Go et al. [20], which achieved around 80% success rate.

Table 4. The most informative features learned by the Naive Bayes classifier.

Feature	Sentiment Class	× Times
Pegida Demonstrant	Not Positive	27.0
Glückwunsch	Positive	16.3
Mittelschicht	Not Positive	14.6
Ojeu	Not Positive	13.0
Pegida Studie	Not Positive	13.0
Berufstaetig	Not Positive	12.3
Studie Pegida	Not Positive	12.1
Maennlich Jahre	Not Positive	11.7
Ausgebildet	Not Positive	11.0
Ausgebildet berufsttig	Not Positive	11.0
Schreiben	Positive	11.0
NOT_ gut ausgebildet	Not Positive	11.0
Kritik Pegida	Not Positive	11.0
Klausuren	Positive	10.3
Willkommen	Positive	10.3
Mittelschicht gut	Not Positive	9.7
Herzlichen	Positive	9.7
Demonstrant mittelschicht	Not Positive	9.7
Geheimdienst	Not Positive	9.0
Wünschen	Positive	8.3

5.2 Tweets Sentiment for Universities Comparison

The rest of the tweets have been classified by the learned classifier. Each tweet belongs to one sentiment class. Tweets were divided over the TU9, showing how many "Positive" and

[9] Patriotic Europeans Against the Islamisation of the West

[10] Congratulations

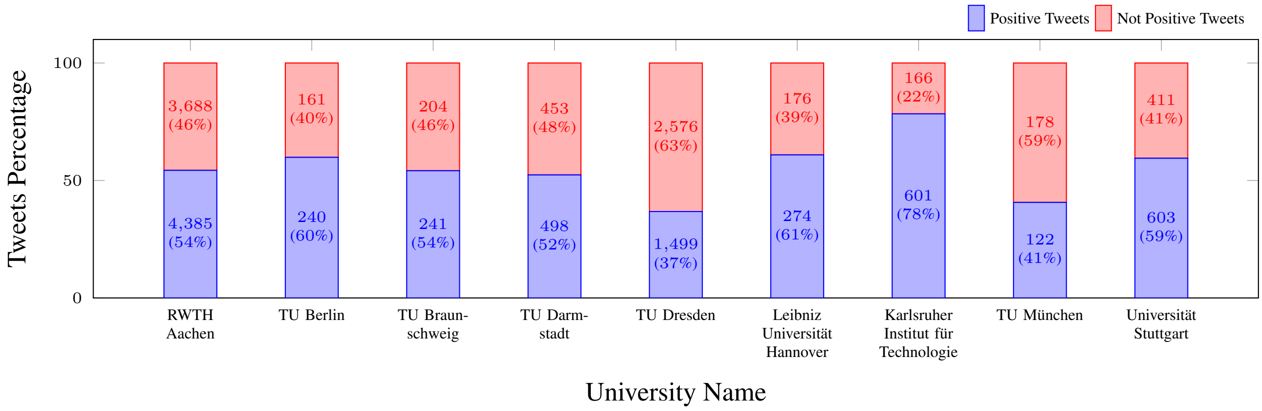

Figure 2. Percentage of “Positive” and “Not Positive” tweets per university in the time span from October 1st, 2014 till March 31st, 2015

“Not Positive” tweets each university had in the specified time span (see Table 5).

Table 5. Classified tweets counts per university of the TU9.

University	Positive	Not Positive
RWTH Aachen	4385	3688
TU Berlin	240	161
TU Braunschweig	241	204
TU Darmstadt	498	453
TU Dresden	1499	2576
Leibniz Universität Hannover	274	176
Karlsruher Institut für Technologie	601	166
TU München	122	178
Universität Stuttgart	603	411

The number of the total tweets per university varied depending on how active Twitter users are at that university. This makes it difficult to establish a comparison between the universities. Nevertheless, to get an overview on how positive the given feedback was by Twitter users on each university, the percentage of the “Positive” tweets is considered for carrying out the comparison. Results are shown in Figure 2. Karlsruher Institute für Technologie had the highest percentage of “Positive” tweets forming 78% out of its total 767 tweets. RWTH Aachen University got the highest number of positive tweets with 4385 tweets. Meanwhile, TU Dresden had the lowest percentage of the positive tweets (37%).
The results might act as an indicator on how the higher education environment at each university is perceived by Twitter users. Such indicator which is supported by the social media content can play a role in enhancing the universities rankings.

5.3 Daily Scale Tweets Sentiment

The next goal is to build a more detailed evaluation by having a closer look into the tweet sentiments regarding daily events and activities at each university. The classified tweets are divided over time on daily basis, where the “Positive” and “Not Positive” tweets count and percentages are considered.
In order to present the case, the two universities with the highest number of tweets were chosen, namely: RWTH Aachen University and TU Dresden.
Figure 3 shows how RWTH Aachen University tweets are spread over the six month period (i.e. winter semester 2014/2015). The upper two figures show the number of “Positive” and “Not Positive” tweets respectively. The third figure shows the ”Positive” tweets percentage on each day.
The “Positive” tweets in RWTH Aachen University data varied over the semester in different frequencies. The graph in Figure 3 can track when and what events or activities are interesting for the students.
Few local maxima points are investigated. The tweets at each point were extracted and word frequencies are analyzed. For RWTH Aachen University, the highest “Positive” tweets count

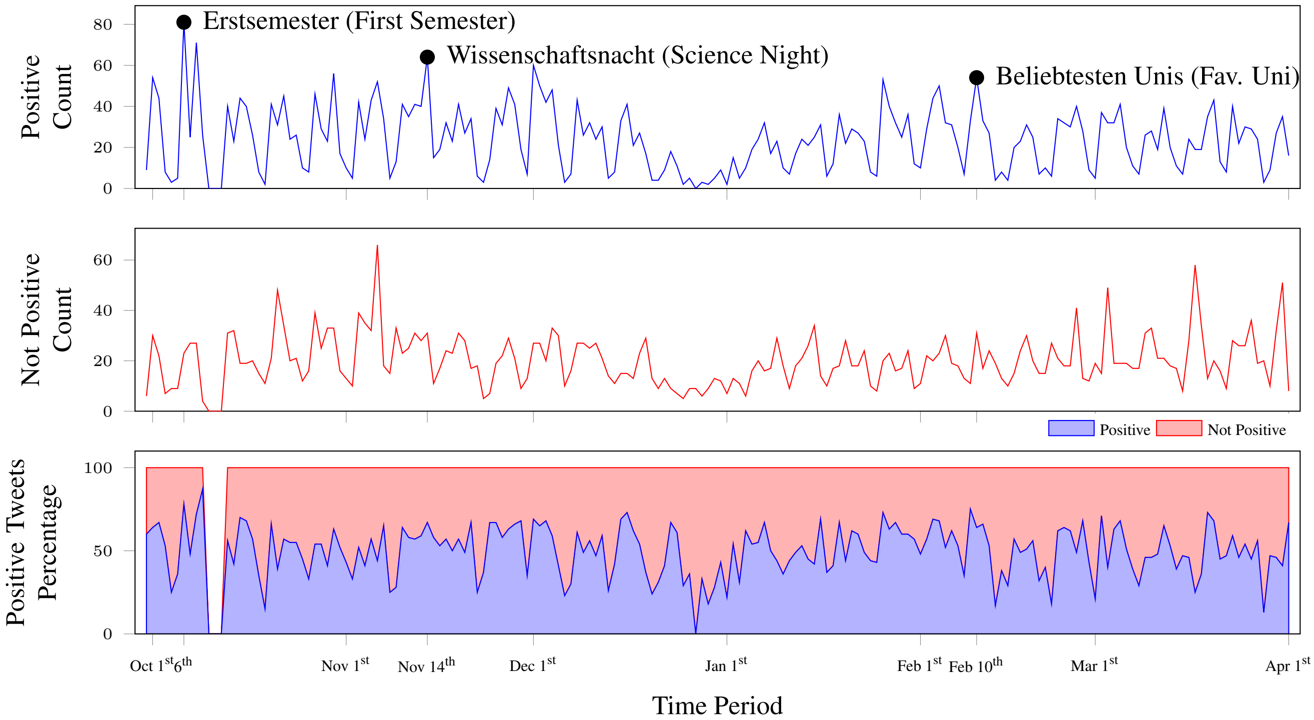

Figure 3. RWTH Aachen University tweets sentiment on daily bases over the period from October 1st, 2014 till March 31st, 2015

occurred on October 6th, 2014, when the word frequency analysis shows that many users that tweeted were excited about the beginning of the new semester. Table 6 shows the most frequent words appeared in that day's tweets. On October 6th the orientation sessions and welcoming event took place, in which many new and old students were tweeting "positively" about it.

Looking at other points similarly, on November 14th, 2014 another peak was experienced where the event of the "Wissenschaftsnacht" (Science Night) took place. It is an annual event where students from different institutes present their work and achievements. Many users tweeted about the event, and the activities which took place at it. One more example is on February 10th, 2015, when the RWTH Aachen University was chosen as one of the top three "most favorite" universities in Germany.

The same analysis steps were applied to TU Dresden tweets. In total, TU Dresden tweets experienced the most of the "Not Positive" sentiments. Figure 4 shows the tweet sentiments on a daily basis over the semester. The

Table 6. Most frequent words which were mentioned in the "Positive" tweets for RWTH Aachen University on October 6th, 2014.

Word	Translation
Erstsemester	First Semester
Heute	Today
Ersti	Newbie
:)	:)
Alle	All
Viel	A lot
Filmstudio	Filmstudio
Wünschen	Wish
neuen	New
Erstis	Newbies
Semesterstart	Semester Start
Bietet	Offers
Herzlich [Willkomen]	Warmly Welcomed

maximum number of the "Not Positive" tweets appeared on January 14th, 2015. As shown in Table 7, the word frequency analysis indicates that many tweets were related to "Pegida demonstrations" that took place on January 2015 in the city of Dresden. This was a reason for many users to tweet "Not Positively" about the activities that happened regarding these demonstration.

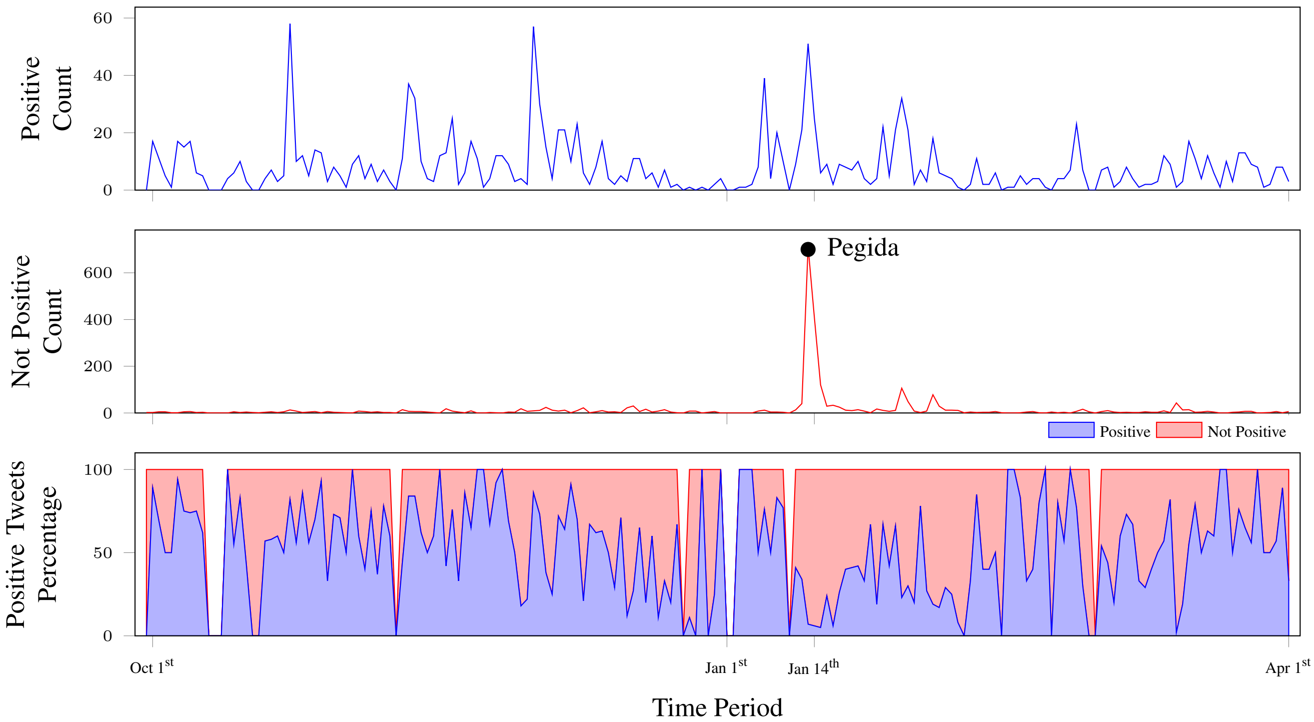

Figure 4. TU Dresden University tweets sentiment on daily bases over the period from October 1st, 2014 till March 31st, 2015

Table 7. Most frequent words which were mentioned in the "Positive" tweets for TU Dresden on January 14th, 2014.

Word	Translation	Count
Pegida	Pegida	1095
Studie	Study	895
Männlich	Male	243
Mittelschicht	Middle Class	196
Demonstrant	Demonstrator	187
Demonstranten	Demonstrators	181
Gut	Good	168
Typische	Typical	158
Typischen	Typical	155

Such analysis for the daily tweet sentiments in the context of higher education can give an insight over what activities and events which attract the attention of the Twitter users, and how they react based on it (i.e. positively or not). This analysis can be used by the university's administration to gather feedback. Based on it, actions can be delivered.

6 CONCLUSION AND OUTLOOK

Students nowadays use different social media platforms including Twitter to express their reactions, and tell about their daily activities. Such platforms can be a vital source for opinion mining related to the universities.

Standard university rankings are quite widespread in order to compare and evaluate universities. University rankings require complementary sources specially those which identify statements from students and teachers expressing their experiences.

This paper established a case study to prove the hypothesis that Twitter sentiment can potentially support the universities ranking system by analyzing posted statements and opinions of students and teachers in higher education institutions context.

The case study used sentiment analysis methods to analyze 16488 collected tweets about the TU9. 4000 annotated tweets between "Positive" and "Not Positive" were used to train a Naive Bayes classifier. The classifier performance achieved 73.6% success rate. The classifier was used to guess the sentiment for the rest of the collected tweets.

The tweets sentiment was analyzed on two different scales. First, the percentage of "Posi-

tive" tweets was calculated to establish a comparison between the TU9. Second, the tweets were analyzed on daily basis for the RWTH Aachen and TU Dresden in order to discover the communicative topics and events at the universities.
The analysis established a new approach of using social media in comparing universities.
As a future improving steps, there is a lot of room in improving the classifier performance regarding the filtering, features selection and chosen classifier type. The sentiment classification can be extended to include a third neutral class. Also, the tweets can be classified based on topics to give a better insight on what attracts Twitter users in the context of higher education institutes.

REFERENCES

[1] Ilona Buchem, Ralf Appelt, Sascha Kaiser, Sandra Schön, and Martin Ebner. Blogging und microblogging-anwendungsmöglichkeiten im bildungskontext. booklet, 2011.

[2] Danah Boyd. A blogger's blog: Exploring the definition of a medium. *Reconstruction, online booklet*, 6(4), 2006.

[3] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pages 56–65. ACM, 2007.

[4] Harald Gapski and Lars Gräßer. Medienkompetenz im web 2.0–lebensqualität als zielperspektive. *Praxis Web*, 2:11–34, 2007.

[5] Julia Hisserich and Jasmin Primsch. Wissensmanagement in 140 zeichen: Twitter in der hochschullehre. webSquare, online booklet, 2010.

[6] Michael Kerres and Annabell Preussler. Soziale netzwerkbildung unterstützen mit microblogs (twitter). Handbuch E-Learning, Köln: Deutscher Wirtschaftsdienst, online booklet, 2009.

[7] Jana Herwig, A. Kittenberger, M. Nentwich, and J. Schmirmund. Microblogging und die wissenschaft. das beispiel twitter. *Steckbrief IV im Rahmen de Projekts "'Interactive Science"' ITA-Reports*, 54-4, 2009.

[8] Claudia Brauer and Edward WN Bernroider. Social media analytics with facebook-the case of higher education institutions. In *HCI in Business*, pages 3–12. Springer, 2015.

[9] Isabel Roessler. Hochschulentwicklung, was war? was bleibt? was kommt? 15 jahre erfahrungen mit rankings und indikatoren im hochschulbereich, arbeitspapier nr 167. booklet, 2013.

[10] Carmen Lebherz, Cordula Mohr, Matthias Henning, and Peter Sedlmeier. Wie brauchbar sind hochschul-rankings? eine empirische analyse. In *Hochschullandschaft im Wandel.*, pages 188–208, 2005.

[11] Sonja Berghoff, Gero Federkeil, Petra Giebisch, Cort-Denis Hachmeister, Detlef Müller-Böling, and Daniel Rölle. Che-hochschulranking. vorgehensweise und indikatoren, arbeitspapier nr 119. booklet, 2006.

[12] Gero Federkeil. Internationale hochschulrankings–eine kritische bestandsaufnahme. *Beiträge zur Hochschulforschung*, 35(2):34–48, 2013.

[13] Ellen Hazelkorn. *World-Class Universities or World Class Systems?: Rankings and Higher Education Policy Choices*. Book of Dublin Institute of Technology, 2013.

[14] Diana Maynard and Adam Funk. Automatic detection of political opinions in tweets. In *The semantic web: ESWC 2011 workshops*, pages 88–99. Springer, 2012.

[15] Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. Sentiment analysis of twitter data. In *Proceedings of the Workshop on Languages in Social Media*, pages 30–38. Association for Computational Linguistics, 2011.

[16] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135, 2008.

[17] Bo Pang and Lillian Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*, pages 271–279. Association for Computational Linguistics, 2004.

[18] Peter D Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th*

annual meeting on association for computational linguistics, pages 417–424. Association for Computational Linguistics, 2002.

[19] Efthymios Kouloumpis, Theresa Wilson, and Johanna Moore. Twitter sentiment analysis: The good the bad and the omg! *in Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM-2011)*, pages 538–541, 2011.

[20] Alec Go, Richa Bhayani, and Lei Huang. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1:12–18, 2009.

[21] Alec Go, Lei Huang, and Richa Bhayani. Twitter sentiment analysis. Final Project Report, 2009.

[22] Farzindar Atefeh and Wael Khreich. A survery of techniques for event detection in twitter. *Computational Intelligence*, 31:132–164, 2013.

[23] Alexander Pak and Patrick Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In *In Proceedings of the Seventh Conference on International Language Resources and Evaluation*, pages 1320–1326, 2010.

[24] Albert Bifet and Eibe Frank. Sentiment knowledge discovery in twitter streaming data. In *Discovery Science*, pages 1–15. Springer, 2010.

[25] Dmitry Davidov, Oren Tsur, and Ari Rappoport. Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 241–249. Association for Computational Linguistics, 2010.

[26] Amy Mollett, Danielle Moran, and Patrick Dunleavy. Using twitter in university research, teaching and impact activities. in Impact of Social Sciences, online booklet, 2011.

[27] Kushal Dave, Steve Lawrence, and David M Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web*, pages 519–528. ACM, 2003.