Sentiment Analysis of Twitter Posts on COVID-19 Cultural Dimensions: Collectivist vs. Individualist

Daniel Dobler*, Leo Donisch†, Melanie Koeppel‡, Patricia Brockmann§
Computer Science Department Nuremberg Institute of Technology
Nuremberg, Germany
Email: *doblerda75546@th-nuernberg.de, †donischle75565@th-nuernberg.de,
‡koeppelme76459@th-nuernberg.de, §patricia.brockmann@th-nuernberg.de

***Abstract*—Social distancing requirements during the COVID-19 pandemic have led to an increase in the importance of social media to maintain communication channels. This paper describes an initial investigation to English Twitter posts about the COVID-19 epidemic. The goal is to determine whether differences in opinion between users from different cultural backgrounds can be discerned. As a first prototype, a classification of tweets according to collectivist and individualistic cultures is attempted. Training data is used to generate feature vectors to train a neural network. Sentiment analysis is employed to classify the posts as positive, negative or neutral. Potential consequences for education and possible adaptive measures for collectivist and individualistic cultures are suggested.**

***Index Terms*—social media; sentiment analysis; cultural; collectivist; individual; education.**

I. Introduction

Physical and social distancing requirements during the COVID-19 pandemic have made it more difficult for people to physically spend time with friends, family members and colleagues. The need to discuss experiences and exchange ideas with others remains a fundamental human need. To fill this void caused by restrictions of in-person meetings, the importance of social media channels has increased [1].

Attitudes toward contact restrictions imposed to combat the spread of COVID-19 have varied considerably among citizens of different countries. Levels of resilience in dealing with stress situations caused by lock-downs have also varied considerably around the world, especially among young people. Cultural dimensions, as described by Hofstede [2], may play a role in explaining some of these different responses. A high level of power distance may positively affect respect for positions of authority and thus increase acceptance of temporary restrictions. Conversely, a culture which highly values individualism may experience lower compliance with health regulations. In collectivist cultures, which value the group higher than the individual, people may willingly adhere to health measures, in order protect weaker members of the society. Depending on the cultural dimensions of a country, educational measures could be specially adapted to help students cope with pandemic measures.

The goal of this work is to build a proof-of-concept prototype to investigate whether it is possible to differentiate between tweets on Covid-19 from different cultures by applying sentiment analysis. Sentiment Analysis is defined as the computational analysis of opinion, analysis and subjectivity in text [3]. Using natural language processing techniques, text can be classified according positive or negative polarity. To perform this investigation, a large collection of Twitter posts were cleaned, pre-processed and their sentiments analyzed. An evaluation was made to determine whether different types of cultures express more positive, negative or neutral opinions on the COVID-19 virus. For this first prototype, a focus is placed on collectivist and individualist cultures.

The research questions examined in this study are:

- R1: Can sentiment analysis deliver meaningful insights into the opinions of the COVID-19 pandemic expressed in Twitter posts?
- R2: Do cultural dimensions associated with users from collectivist vs. individualist cultures affect their expressed opinions?

First, an overview of the related literature is surveyed in Section II. The methods employed in this work are described in Section III. In Section IV, initial results of the prototype model for sentiment analysis are presented. Finally, conclusions and plans for future work are discussed in Section V.

II. Related Work

A. Cultural Dimensions

Hofstede [2] was one of the first investigators to apply multivariate statistical methods to analyze data from a large, international survey of thousands of information technology professionals. The differences observed in cultural perspectives among respondents from different countries were scored according to six dimensions:

1) Power distance: How a society views inequalities between individuals
2) Collectivism vs. individualism: Preference for loosely or tightly-knit social frameworks
3) Masculine vs. feminine: Achievement and assertiveness vs. cooperation and caring
4) Uncertainty avoidance: Degree to which unknown or ambiguous situations are viewed as threatening
5) Long-term vs. short-term orientation: Thrift and planning for the future vs. challenges of the present
6) Indulgence: Immediate gratification vs. restraint.

In addition to these six cultural dimensions, Hall [4] differentiates between high and low context cultures. In low context cultures, explicitly written and spoken words are the primary

 ISBN: 978-1-68558-002-5

source of meaning. Thus, communication in low context cultures can often seem quite verbose. Western countries, such as Germany, tend to be classified as low context cultures and also place a high value on individualism. In high context cultures, personal relationships between people, such as their level of familiarity or differences in societal status, can play an intrinsic role in communication. Unspoken communication, such as facial expressions, gestures and pauses can sometimes convey more meaning than the actual written or spoken words. East Asian countries, such as Japan, are classified as high context cultures and also tend to value collectivism.

Kim et al. [5] found major differences in the usage of social media by university students in Korea and the U.S. Korean students were more motivated to use social media to obtain social support from existing relationships, while American students were more interested in seeking entertainment. Korean students also had a smaller number of contacts in their networks than American students. This appears to coincide with Hofstede's [2] findings, which categorize Korea as a collectivist culture and the U.S. as an individualistic culture.

B. Sentiment Analysis of Social Media

A number of researchers have applied the technique of sentiment analysis to social media. Chakraborty et al. [6] conducted a widespread literature review of over 200 papers on the subject of social networks and the use of sentiment analysis in social media. They review different techniques of sentiment analysis and point out important challenges which should be addressed, such as rumor detection and community shaming. Strathern et al. [7] explored the use of sentiment analysis to detect so-called "firestorms" on Twitter. These firestorms can be triggered by negative online dynamics, which result in uncontrollable escalation which result in real harm to people. Tsao et al. [8] performed a literature review of 81 studies on online social media and COVID-19. They identified five main public health themes: surveying public attitudes, identifying infodemics, assessing mental health, detecting or predicting COVID-19 cases, analyzing government responses to the pandemic and evaluating the quality of health information in prevention education videos. Their main criticism is the scarcity of studies documenting real-time surveillance with data from social media. Aggregated data from Facebook was found to show that COVID-19 is more likely to spread between regions with stronger social network connections [9].

Sentiment analysis conducted during a nationwide lockdown in one single country, India, showed that although a number of negative sentiments were expressed, such as fear, disgust, and sadness, the overwhelming sentiments were positive, especially trust [10]. A different study from India utilized sentiment analysis on tweets. They found that popularity has an effect on the accuracy of information disseminated over social media. The most popular retweets skewed highly negative and did not contain any significant information [11]. Another study compared topic modeling for English and Portuguese tweets related to COVID-19. They found that the top ten topics for both languages were mostly similar [12].

Kruspe et al. [13] analyzed Twitter messages collected during the first few months of the pandemic in Europe. They performed a sentiment analysis using multilingual sentence embeddings and separated the results by country of origin. They found that lockdown measures correlated with a deterioration of sentiment in almost all of the countries surveyed. A sentiment analysis Twitter messages from different countries was performed by Imran et al. [14]. They divided up countries into geographic regions. The U.S. and Canada were grouped together as North American countries. India and Pakistan were grouped together as South Asian countries. Sweden and Norway were grouped together as Nordic countries. They found a high level of correlation between countries in the North American group. The South Asian group also showed a high level of correlation within the group. Sweden and Norway, however, showed opposite trends in polarity. This result is quite surprising and inspires further inquiry.

One cause which may be explain highly different sentiments between geographically close countries may lie in different values for certain cultural dimensions. Sentiment analysis geared toward clusters of countries which share similar values on specific cultural dimensions has not yet been handled in the literature. This research gap is addressed in this paper.

III. Methods

The goal of this study was to investigate possible differences in the sentiment of groups with similar values of cultural dimensions. For this first proof-of-concept experiment, individualist vs. collectivist cultures were examined.

A. Data Source

This work was conducted on a publicly available data set which contains tweets about the COVID-19 pandemic. This data set is freely available on the open data platform on Kaggle and includes 44,955 tweets from March 12th through March 16th, 2020 [15]. Each data record includes information about the user, their screen name (encoded to preserve user privacy) and the date the tweet was posted. In addition to the text content of each tweet, information about the location where it was posted and the sentiment of the text was included.

The sentiment of each tweet (very positive, positive, neutral, negative, very negative) was determined manually by the author of the data set and serves as the label value which the algorithm used in this experiment attempts to predict. Manual labeling can be quite difficult, even for a human. Furthermore, individual, subjective opinions may also bias this labeling process. To get a more objective evaluation, three of the authors of this work manually labeled 100 tweets. In 41 % of these tweets, the three given labels were unambiguous. Through a majority decision, a label could be assigned to 96 % of the tweets. Four examples could not be labelled, since they received one vote each of positive, neutral and negative. For 47 % of the tweets, the appropriate value was determined. For 53 % of the tweets, the manual evaluation would have given a different label to the tweet than the author of our source.

 ISBN: 978-1-68558-002-5

B. Data Partitions

The data set first needs to be partitioned according to the country of origin. These countries will then be assigned to groups with similar values on cultural dimensions. Twitter users input their geographic locations as strings. As a consequence, a location can sometimes contain a country, federal state, city or any other character data. The only location data which can be easily mapped to a certain value for one cultural dimension is the country. Therefore, only the data samples which included a country as part of the location data were used in this first prototype experiment. To refine samples with an easy to convert country, all of the data samples with country values that consist of only one word were selected. In the next step, a function from Kaggle [16] was combined with the Python library pycountry [17] to recognize countries. The snippet was adjusted so that it could identify countries by the common name, the official name and ISO (International Organization for Standardization) alpha-3 code. The recognition of ISO alpha-2 codes was disabled, because they can often match with both a federal state or with a country.

Once all samples for the test data set were prepared, the location value had to be associated with the correct cultural dimension group: individualist or collectivist. To achieve this for all of the 152 extracted countries, a dictionary was established and used to map countries to one of the two groups. The strength of the score for the dimension of individualism vs. collectivism (IDV) in each country was compared to the values of Hofstede [2]. If the value for a country was greater or equal than 50, the sample was assigned the tag “I”, for an individualistic culture. If the value was below 50, it was assigned the tag “C”, for a collectivist culture.

C. Pre-processing Pipeline

In order to evaluate text automatically, it first must be cleaned of unnecessary information and then transformed into a format which can be analyzed. Fig. 1 shows the pre-processing steps conducted before the data analysis, based on the recommendations of [18].

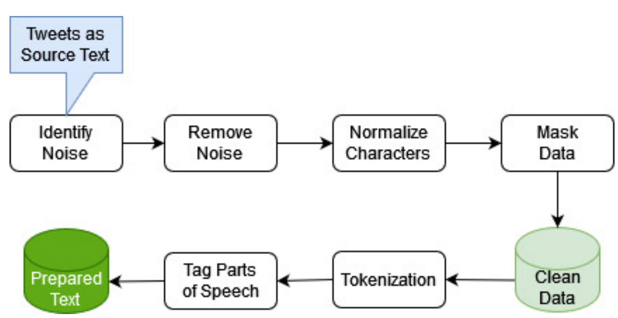

Fig. 1. Pre-Processing Pipeline

1) Data cleaning: One of the first steps for efficient noise removal is to correctly identify the noise in the given context. To achieve this, the length of each tweet was first calculated. A maximum number of 280 characters are allowed per tweet. If the length of a tweet exceeded this maximum, then it was flagged for closer analysis. Another measurement employed was the impurity score, which indicates the share of suspicious characters in a text. This enables recognition of noisy tweets and to measure improvements of data cleaning. For each tweet in the data set, its length and impurity score was calculated. If one particular tweet had a length longer than 280 characters and a high impurity score, this tweet was flagged for more detailed analysis. Tweets which didn’t exceed the maximum length but had an high impurity score were also flagged for further analysis. This approach did not always produce perfect results; not all of the noise could be effectively identified.

TABLE I
STOP WORDS

Before Removal	Polarity	After Removal	Polarity
The lockdown is good	pos	lockdown good	pos
The lockdown works	pos	lockdown works	pos
I did not like the lockdown	neg	like lockdown	pos
This lockdown is no good	neg	lockdown good	pos

Two successive cleaning approaches were implemented. First, artifacts of the extraction method are identified and removed. Next, a more specific cleaning method incorporates the insights of the analysis. These steps try to minimize tweet-specific patterns, such as URLs and user handles. These methods were implemented using functions, which were derived by Albrecht [18]. To summarize, these steps are necessary to remove unwanted patterns and simultaneously minimize word variants, in order to enable to learn a more precise language model [18] [19]. With all these measures, the tweets became cleaner and smaller, as measured by the impurity score.

An additional source of noise are so-called ”stop words” [20]. Stop words are parts of speech, such as prepositions, conjunctions or determinants, such as “and”, “or”, and “a”. Simply blindly removing these words from a text corpus is not ideal. If stop words such as “not” are removed, the sentiment of a tweet completely changes. This can cause problems, because the sentiment label is the opposite of what the tweet implied, as shown in Table I. In order to maintain the original sentiment, stop words which reverse the sentiment should not be removed from the tweets. They need to be removed from the default stop word list in the library SpaCy [20].

2) Tokenization: In order to use text in a machine learning algorithm, text must be correctly segmented into analyzable elements. This step is called tokenization and the results are referred to as n-grams. In the first step, each tweet gets segmented into one-word-tokens, called unigrams. Linguistic attributes are attached to each token, to achieve more precise n-grams in a later step. The process of segmenting text into smaller elements and attaching the linguistic attributes was done using a python library called SpaCy [21].

The linguistic attributes contain a boolean value, which indicates whether a given unigram is a stop word. When utilizing the stop word removal method described in Subsection III-C1, only non-sentiment changing words are removed. A further linguistic attribute is the Part-Of-Speech Attribute, which

 ISBN: 978-1-68558-002-5

can be used to extract grammatical insights of a tokenized tweet [18]. Here, it is used to construct meaningful n-grams. Building n-grams is important if context information plays a key role in the analysis.

The library Textacy contains a useful function which allows for POS-tag pattern search, similar to regular expressions [22]. With this pattern-search function, a phrase which starts with an adjective and ends with a noun, so-called adjective-noun-phrases, can be extracted. Capturing sentiment adjectives plays a key role. The adjective-noun-phrases and adjective-adverb-phrases were extracted according to methods described in [18].

One major drawback is that not every tweet is structured grammatically correctly and therefore may not yield good results. Thus, using only bigrams is not ideal. For this reason, the default unigrams were extended with the adjective-noun-phrases found to capture context information. These tokens could then be passed into the desired vectorization method, which will be described in the next subsection.

D. Feature Engineering

In order to use machine learning algorithms with text, words need to be transformed into a numerical representation. One approach often used is called the "Bag of Words" method. In this approach, every token represents a phrase or word learned in the vocabulary. For each tweet, the frequency of how often each of these words occur is calculated. Thus, the first thing this method does is to learn the vocabulary list of the data set and then to transform each tweet into word frequencies. If a word in a new tweet is not already in the vocabulary list, it is ignored. In this way, the majority of the words captured must be contained in the training data set [18]. However, this approach is prone to over-weighting frequently used words and under-weighting less frequently used words. To counter this trend, the Term Frequency - Inverse Document Frequency (TF-IDF) value is calculated. The TF-IDF calculation boosts words which are less frequently used and slightly lowers the weighting of frequently used words [23].

Two additional parameters were also used: maximum and minimum document frequency. These parameters restrict the algorithm from learning words which occur too often or too seldom. A maximal document frequency of 80% is used in this work. This means that tokens which appear in more than 80% of the tweets were not be used. Tokens which appear in less than five tweets were also omitted. By limiting the size of the vocabulary, the dimension of the vector is reduced, thus making learning more efficient. The disadvantage is that some information does get lost, because the discarded terms could potentially carry important meaning [18]. In this case, both learning time and the metrics used both improved.

E. Neural Network

As an initial experiment, a pre-trained model which uses the Bidirectional Encoder Representations from Transformers (BERT) [24] from the TensorFlow repository was used. This is a non-case sensitive, smaller version of BERT, which was pre-trained for English on Wikipedia and BooksCorpus. A notebook for the data set investigated was published on Kaggle [15] for free use. The model distinguishes between five classes of the sentiment of a text. It uses a special tokenizer for BERT and one layer of the pre-trained BERT model, positioned ahead the other three layers of the neural network. In total, the model has 109,533,701 trainable parameters, which requires high computing times. In an experiment with an average Windows laptop, only 15 samples could be used due to the processing capacity limitations. Because of the small number of samples, only a 40% accuracy rate could be achieved. This accuracy rate is lower than flipping a coin and was thus judged to be too low for further pursuit.

As an alternative, a simpler model of SciKit-Learn (sklearn) with a Multi-Layer Perceptron Classifier (MLPClassifier) offers was implemented. Using the module "neural network" from sklearn, it was easy to adapt the model to fit the research goal and to adjust its parameters. In order to analyze the sentiment of Twitter posts, it would have been possible to use a number of different machine learning algorithms, such as a random forest algorithm. Neural networks were selected for this implementation, because they can easily be easily trained to recognize complex patterns [25]. During the training phase, the network can be fine-tuned. For example, the selection of methods for weight optimization or value of the learning rate provide a high amount of flexibility and fine tuning potential. The architecture of the neural network, the number of hidden layers and the number of neurons in each hidden layer can also have an impact on the results [25].

For this prototype model, hyper-parameters were optimized to help improve accuracy. First, a subset of the data were used to try out different combinations of the random key for the initialization of the weights and different hidden-layer sizes. The best results were achieved with hidden-layer-sizes of 100, 35, 11 and 7. The Train-Test-Split random-seed was evaluated by testing different random-seeds.

IV. Results

Because this work describes a proof-of-concept prototype, results here are described in a step-by-step fashion. The first prototype designed was a simple model with just one hidden-layer, consisting of 100 neurons and an output-layer which contained five sentiments: very positive, positive, neutral, negative, very negative. The results were of this first prototype were extremely bad, because it was very difficult to distinguish between "positive" and "extremely positive" sentiments.

For this prototype it was sufficient to differentiate between positive and negative sentiments. Thus, the number of labels was reduced: Positive and very positive labels were combined to a single label, positive. Negative examples were combined in a similar way and the neutral examples were first temporarily omitted. With these simplifications, the model achieved an accuracy rate of 83%. After further improvements to the pre-processing, neutral labels were once again reintroduced.

Table II show the results of the final model configuration. The worst F1-Score was achieved when attempting to classify posts with a neutral sentiment.

 ISBN: 978-1-68558-002-5

TABLE II
PERFORMANCE METRICS

Metric	Score	Metric	Score
Accuracy	73.5%	Precision	71.0%
Recall	72.2%	F1	71.6%
Sensitivity	72.2%	Specificity	74.6%

Serrano et al. recommend measuring Roc-Auc curves to improve the performance of classification models [26]. The Roc-Auc-Curve is defined as the Area Under the Receiver Operator Curve. The value for the Roc-Auc-Curve in this result was 87%.

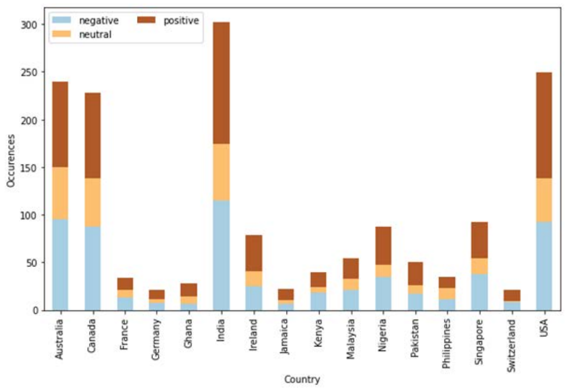

Fig. 2. Results of Sentiment Analysis by Country

Fig. 2 shows the preliminary results of the sentiment analysis for some of the individual countries investigated. Ghana has a score for individualism of 15 and would be classified as a collective country. The high percentage of positive (50%) to negative (25%) comments seems to support the hypothesis that collective countries show more positive sentiments. Singapore, with its low individualism score of 20 would be considered a collectivist society. Its percentages of positive (41%) and negative (39%)comments were almost identical. Switzerland, a country known neutrality and a high score of 68 for individualism, showed almost exclusively positive (52%) and negative (43%) sentiments, with very few neutral (5%) sentiments. Although Kenya, with an individualism score of 25 would be classified as a collective country, the percentage of negative comments (51%) is much higher than the percentage of negative comments for the U.S.A. (36%), a country with one of Hofstede's highest individualism scores (91%) [2].

The results comparing the sentiments recognized for individualist and collectivist countries is shown in Fig. 3. Against expectations, there are no statistically significant differences between the sentiment of collectivist and individualist cultures. Therefore the hypothesis that individualist cultures have a more positive sentiment about the coronavirus cannot be confirmed. Although the prototype model was able to achieve an accuracy of 73.5%, the validity of the location data and the labels remains uncertain. Another alternative explanation may lie in the binary cut-off value of 50 for Hofstede scores, which was used when assigning countries to either the individualist or

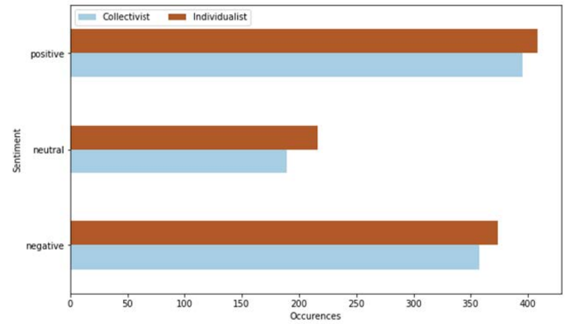

Fig. 3. Results Grouped by Cultural Dimension

the collectivist group. Some countries, such as India, with an score of 48 on individualism, were automatically be assigned to the collectivist group. An assignment based on mean or median values could possibly deliver more exact results.

A. Limitations

One drawback which may limit the validity of the results of this proof-of-concept prototype is that only a small subset of the data was used for this initial investigation. This subset is not necessarily representative. Furthermore, this data set was collected at the beginning of the pandemic, over a relatively short time period. Sentiments probably changed in different countries over the course of the pandemic. It would definitely be a good idea to investigate how sentiments in different countries changed over time.

The problem that the Twitter data is very noisy necessitated heavy pre-processing of the tweets, which may have introduced additional bias and thus skewed results. Uysal and Gunal [27] showed that the choice of pre-processing methods can have a significant effect on classification accuracy.

A further source of bias could have been introduced when assigning tweets to specific countries. Countries with more than one word, such as the "United Arab Emirates", are underrepresented in the test data set. Other countries were not included in the analysis, because Hofstede [2] did not provide a rating for 34 of the nations in the data set. It was also not possible to establish whether the location corresponds to the place where the tweet was posted or to the actual cultural background of the user who posted it. One method to improve the quality of the location data for the analysis would be to implement recognition of latitude an longitude coordinates.

A further potential problem is that the labeling process was done manually by the author of data set [15]. Errors could already have been introduced during this manual labeling. Differentiating between positive and negative sentiments is difficult, even for a human. The sentiment of a statement can be misinterpreted or experienced diversely, since it is a very subjective value. One suggestion would be to explore the use of crowd-sourcing to aid in the labelling of tweets.

V. CONCLUSION

An initial, proof-of-concept prototype to analyze tweets related to COVID-19 was described. The research questions posed at the beginning can now be answered:

- R1: Can a simple sentiment analysis model deliver meaningful insights into the opinions of the COVID-19 pandemic expressed in Twitter posts?
- R1: TRUE, the initial prototype described here was able to correctly classify the sentiments of 73.5% of tweets.
- R2: Do cultural dimensions of users from collectivist vs. individualist cultures affect their expressed opinions?
- R2: FALSE: No significant differences were found in the sentiments from collectivist vs. individualist countries.

Future work on the implementation of the complete model will include applying sentiment analysis to look at other Hofstede-Dimensions, such as power-distance, masculinity vs. femininity, uncertainty avoidance, long-term vs. short-term orientation and indulgence vs. restraint [2].

Methods to obtain better location data from each tweet will be explored. By using longitude and latitude data, it would be possible to get further information about the federal state or even the county, to analyze differences within a country. Potentially, countries grouped by other characteristics could be compared, such as the G7-States to the rest of the world. A closer look at the demographics of the users who post the tweets could be warranted. Do age, gender or other demographics affect sentiment on the COVID-19 virus?

Further work will include use of larger data sets gathered over longer time periods. The feasibility of cloud-sourcing platforms to improve the quality of data labeling will be explored. The usage of a pre-trained model such as BERT [24] could be useful to improve the recognition of different sentiments. A comparison of different methods, eXtreme Gradient Boosting (XGBoost), Random Forest or linear regression algorithms should be explored. A final idea for future work would analyze whether cultural dimensions affect the sentiment of students during the pandemic.

This work is part of a larger research project to develop hybrid courses to teach global software engineering to geographically separated groups of students. Once the full model has been implemented, investigations will focus on how teaching methods can best be adapted to students who come from different cultural backgrounds during and after the pandemic.

ACKNOWLEDGMENT

This work was supported by a grant from the German Academic Exchange (DAAD) program for International Virtual Academic Collaboration (IVAC) and with support from the Ritsumeikan University in Japan and the Nuremberg Institute of Technology in Germany.

REFERENCES

[1] W. Hussain, "Role of social media in covid-19 pandemic," *The International Journal of Frontier Sciences*, vol. 4, no. 2, pp. 59–60, 2020.
[2] G. Hofstede, G. J. Hofstede, and M. Michael, *Cultures and Organizations: Software of the Mind*. McGraw-Hill, 2010.
[3] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends® in information retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.
[4] E. T. Hall, *Beyond culture*. Anchor Books, 1989.
[5] Y. Kim, D. Sohn, and S. M. Choi, "Cultural difference in motivations for using social network sites: A comparative study of american and korean college students," *Computers in human behavior*, vol. 27, no. 1, pp. 365–372, 2011.
[6] K. Chakraborty, S. Bhattacharyya, and R. Bag, "A survey of sentiment analysis from social media data," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 2, pp. 450–464, 2020.
[7] W. Strathern, M. Schoenfeld, R. Ghawi, and J. Pfeffer, "Against the others! detecting moral outrage in social media networks," in *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 2020, pp. 322–326.
[8] S.-F. Tsao, H. Chen, T. Tisseverasinghe, Y. Yang, L. Li, and Z. A. Butt, "What social media told us in the time of covid-19: a scoping review," *The Lancet Digital Health*, vol. 3, no. 3, pp. e175–e194, 2021.
[9] T. Kuchler, D. Russel, and J. Stroebel, "Jue insight: The geographic spread of covid-19 correlates with the structure of social networks as measured by facebook," *Journal of Urban Economics*, vol. 127, p. 103314, 2022.
[10] G. Barkur and G. B. K. Vibha, "Sentiment analysis of nationwide lockdown due to covid 19 outbreak: Evidence from india," *Asian journal of psychiatry*, vol. 51, p. 102089, 2020.
[11] K. Chakraborty, S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag, and A. E. Hassanien, "Sentiment analysis of covid-19 tweets by deep learning classifiers—a study to show how popularity is affecting accuracy in social media," *Applied Soft Computing*, vol. 97, p. 106754, 2020.
[12] K. Garcia and L. Berton, "Topic detection and sentiment analysis in twitter content related to covid-19 from brazil and the usa," *Applied soft computing*, vol. 101, p. 107057, 2021.
[13] A. Kruspe, M. Häberle, I. Kuhn, and X. X. Zhu, "Cross-language sentiment analysis of european twitter messages duringthe covid-19 pandemic," *arXiv preprint arXiv:2008.12172*, 2020.
[14] A. S. Imran, S. M. Daudpota, Z. Kastrati, and R. Batra, "Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on covid-19 related tweets," *Ieee Access*, vol. 8, pp. 181 074–181 090, 2020.
[15] A. Miglani, "Coronavirus tweets nlp - text classification," Sep 2020, accessed on 18.07.2022. [Online]. Available: https://www.kaggle.com/datatattle/covid-19-nlp-text-classification
[16] Monga, "Names of countries," Aug 2018, accessed on 18.07.2022. [Online]. Available: https://www.kaggle.com/datatattle/covid-19-nlp-text-classification
[17] Flyingcircusio, "Flyingcircusio/pycountry: A python library to access iso country, subdivision, language, currency and script definitions and their translations." accessed on 18.07.2022. [Online]. Available: https://github.com/flyingcircusio/pycountry
[18] J. Albrecht, S. Ramachandran, and C. Winkler, *Blueprints for Text Analytics Using Python*. O'Reilly Media, 2020.
[19] D. Forsyth, *Applied Machine Learning*. Springer, 2019.
[20] J. Nothman, H. Qin, and R. Yurchak, "Stop word lists in free open-source software packages," in *Proceedings of Workshop for NLP Open Source Software (NLP-OSS)*, 2018, pp. 7–12.
[21] Y. Vasiliev, *Natural Language Processing with Python and SpaCy: A Practical Introduction*. No Starch Press, 2020.
[22] B. DeWilde, "textacy documentation," 2021, accessed on 18.07.2022. [Online]. Available: https://textacy.readthedocs.io/en/
[23] S. Qaiser and R. Ali, "Text mining: use of tf-idf to examine the relevance of words to documents," *International Journal of Computer Applications*, vol. 181, no. 1, pp. 25–29, 2018.
[24] S. Ravichandiran, *Getting Started with Google BERT: Build and train state-of-the-art natural language processing models using BERT*. Packt Publishing Ltd, 2021.
[25] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems.* " O'Reilly Media, Inc.", 2019.
[26] A. J. Serrano, E. Soria, J. D. Martin, R. Magdalena, and J. Gomez, "Feature selection using roc curves on classification problems," in *The 2010 international joint conference on neural networks (IJCNN)*. IEEE, 2010, pp. 1–6.
[27] A. K. Uysal and S. Gunal, "The impact of preprocessing on text classification," *Information processing & management*, vol. 50, no. 1, pp. 104–112, 2014.