BABEŞ-BOLYAI UNIVERSITY, CLUJ-NAPOCA, ROMÂNIA FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Data Modelling Techniques and Various Applications of Sentiment Analysis

PhD thesis summary

PhD Student: Sergiu Limboi
Scientific Supervisor: Prof. PhD Laura Dioşan

2024

Keywords: Sentiment analysis, Twitter, Text mining, Affective computing, Data preprocessing, Machine learning

Abstract

In recent years, social media has become a popular environment where everybody can express their ideas about events, celebrities, politics, or education. Step by step, the opinions and news posted here became an essential source of information for many people. Analyzing this data can be essential for predicting and defining marketing strategies, influencing political assessments or social views, or building relevant recommendations for potential customers.

Sentiment Analysis is a domain with several activities that can be applied to this analysis. One of these is called polarity detection, which focuses on determining the sentiment from a given input. This task is suitable for studying and evaluating the textual data provided by social media. Exploring and designing relevant features based on online texts is an important step during this process.

Therefore, an important goal of the thesis is to model interesting features from Twitter data, also with the help of sentiment lexicons. Moreover, the fusion between these features is used for detecting the sentiment of tweets. Further, some models are applied for the grouping process of defining clusters corresponding to sentiment labels.

Another target of the research is combining polarity detection with some Recommendation Systems. Enhancing a review about a hotel or a restaurant with the appropriate sentiment can be very useful for building meaningful user suggestions. Besides this, a new similarity measure is designed. The *ARP (Attractiveness-Relevance-Popularity)* measure considers the sentiment determined from reviews.

The experiments indicate that using the newly defined features improves the polarity detection task for text inputs. Also, the results highlight that exploring the review's polarities for making suggestions can be challenging but very helpful. To conclude, the data modelling part of Sentiment Analysis and its applications bring valuable information to understand better the amount of data that surrounds us daily.

List of Publications

- Petruşel, Mara-Renata and Limboi, Sergiu-George **A restaurants recommendation system: Improving rating predictions using sentiment analysis** In 2019 21st International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC) (pp. 190-197). IEEE. (Conference Category C-2 points)
- Limboi, Sergiu and Dioşan, Laura. **Hybrid Features for Twitter Sentiment Analysis.** In Artificial Intelligence and Soft Computing: 19th International Conference, ICAISC 2020, Zakopane, Poland, October 12-14, 2020, Proceedings, Part II 19 (pp. 210-219). Springer International Publishing. (Conference Category C-2 points)
- Deac-Petruşel, Mara and Limboi, Sergiu. **A sentiment-based similarity model for recommendation systems.** In 2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC) (pp. 224-230). IEEE. (Conference Category D-1 point)
- Limboi, Sergiu and Deac-Petruşel, Mara. **A Validation Framework for ARP Similarity Measure**. In 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 1266-1271). IEEE.(Conference Category C-2 points)
- Limboi, Sergiu and Dioşan, Laura. **An unsupervised approach for Twitter Sentiment Analysis of USA 2020 Presidential Election.** In 2022 International Conference on INnovations in Intelligent SysTems and Applications (INISTA) (pp. 1-6). IEEE. (Conference Category C-2 points)
- Sergiu Limboi and Laura Dioşan. **The Twitter-Lex Sentiment Analysis System.** In Proceedings of the 14th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2022) - KDIR, pages 180-187. (Conference Category C-2 points)
- Limboi, Sergiu and Dioşan, Laura. **A Lexicon-based Feature for Twitter Sentiment Analysis.** In 2022 IEEE 18th International Conference on Intelligent Computer Communication and Processing (ICCP) (pp. 95-102). IEEE. (Conference Category D-1 point)
- Sergiu Limboi. **Comparison of Data Models For Unsupervised Twitter Sentiment Analysis**. Studia Universitatis Babeş-Bolyai Informatica, 67(2), p. 65-80, May 2023. (Journal Category D-1 point)

Total points: 13

Contents

Chapter 1

Introduction

1.1 Motivation

Nowadays, information overload [15] is a real problem for people since they face a lot of news, advertisements, and offers. If a person wants to buy a car, they can choose from various models with many options or enhancements. So, the vast amount of information overwhelms a potential client, and he tends to postpone the decisions to analyze the possibilities in more detail. Also, from a psychological point of view, the bigger the offer, the more complex the decision.

The online gained ground in recent years and has become a real power for influencing people and replacing traditional communication channels like radio, newspapers, or television. Social media allows users to express their opinions about politics, health, weather, or celebrities. Moreover, users can have a social network, share views, and follow different users. Compared with the traditional way of writing, the texts on social media are represented by the free-way style, the use of abbreviations, or colloquial language.

The Sentiment Analysis domain [11] could be a powerful tool for analyzing text-based information. One of its primary goals is to determine the sentiment or polarity from different sources of information. Furthermore, the sentiment can be positive, neutral, or negative. Labeling a text posted on social media can properly indicate users' opinions regarding an event. Before the sentiment detection part, the preprocessing data is challenging due to the way of writing on social platforms. Also, modelling data to represent a proper input for a classification algorithm is essential. In the end, all these actions build a classification task that determines if a message has a particular sentiment.

Sentiment Analysis has many applications in different areas. One of them is the process of making recommendations for users. If a user wants to shop online, the reviews are essential for his decisions. A product with many bad reviews will not be attractive to a potential customer, but one with positive comments can be the choice. Analyzing other users' online reviews is an interesting part of the Sentiment Analysis. Based on the difficulties of interpreting a text (negations, sarcasm, or colloquial writing), an enhanced sentiment review can be very constructive. Thus, making a list of suggestions with positive feedback can improve the system's job of presenting recommendations to users.

In other words, there is a need to describe the data posted on social media to define trends and current views and predict future events. Hence, the presented thesis will apply the Sentiment Analysis techniques for textual information (reviews and tweets) to improve the quality of sentiment detection.

1.2 Original contributions

The original contributions of this thesis are emphasized by describing the defined approaches step by step.

The first designed direction [7] focuses on the feature extraction level for improving the classification of sentiments for tweets. The purpose of such attributes is to reveal hidden information from the original tweet. Considering these aspects, we propose combining the features extracted from both parts of a Twitter message (text and hashtags) to test if their fusion improves the classification process. So, four features are defined: Baseline-based Sentiment Analysis (BSA), Hashtag-based Sentiment Analysis (HSA), Fused-based Sentiment Analysis (FSA), and Raw-based Sentiment Analysis (RSA).

The following approach [8] defines a lexicon-based feature to check if the polarity of a tweet can be strengthened with a hint. Therefore, a given tweet is enhanced with a **sentiment indicator**. This sentiment indicator is determined based on the sentiment score computed by a lexicon. Four sentiment lexicons are used for the experiments: Text Blob [17] , Vader [4], Senti WordNet [3], and aFINN [12]. The results show similar values regarding used lexicons, so for future experiments, it is enough to use only one of the previously mentioned lexicons.

The next natural step is to check whether combining Twitter features and the sentiment indicator can further enhance the sentiment classification problem for tweet information. A system called **Twitter-Lex Sentiment Analysis** [9] merges the lexicon-based viewpoint with Twitter-specific features. So, four features are defined: Baseline Sentiment Analysis-Lexicon (BSA_{lex}), Hashtag Sentiment Analysis-Lexicon (HSA_{lex}), Fused Sentiment Analysis-Lexicon (FSA_{lex}) and Raw Sentiment Analysis-Lexicon (RSA_{lex}). Each contains the original Twitter feature (BSA, HSA, FSA, and RSA) and the sentiment indicator determined based on the Vader lexicon.

For the subsequent approaches, the focus is shifted to the unsupervised context for detecting the sentiment in collected tweets. Firstly, presidential tweets from the 2020 election are considered for analysis [10]. The main goal is to build corresponding clusters for the sentiments (positive or negative) and to check if several representations or models increase the quality of the grouping. A new model called **hash index** is defined considering the extracted hashtags from a tweet. Other representations are used during the process: TF-IDF and $TF-IDF_{hash}$.

The perspective from [5] introduces two data representations defined based on text and hashtag features extracted from tweets. The mentioned representations are applied in the unsupervised context for detecting two groups of messages: one with positive tweets and one with negative messages. These models correspond to **BSA** and **HSA** features from the supervised context. The contribution is represented by comparing the two representations to determine which fits best in the unsupervised scenario. The results show that hashtag-based representation is better than the text one, and the clustering techniques do not influence the whole process.

The last part of the contributions highlights a collaboration work that presents the use of Sentiment Analysis for reviews in the context of Recommender Systems [14]. Our work involves Machine Learning classifiers that are used for labeling collected reviews. An improved dataset with the resulting polarity is passed to a collaborative filtering technique, and only positive reviews are considered during the recommendation phase.

Next, the **ARP (Attractiveness-Relevance-Popularity)** similarity measure [2] is defined considering a **Sentiment Scoring** module. Therefore, each review has an associated sentiment score computed considering the score determined based on the Senti WordNet lexicon. The experiments show that the sentiment rating system produces very good results and can be used for the recommendation process. The entire proposal for this similarity is presented in [2], describing also the recommendation part. The subject of this thesis is only our contribution which is the Sentiment Scoring module.

Another interesting step is the defined validation framework for **ARP** similarity [6] to check if it can be used like the classical ones (e.g., Pearson Correlation Coefficient, Euclidean distance, Jaccard metric, etc.). The validation framework is evaluated based on five criteria, but this thesis explains the noise robustness criterion since this is our contribution. This condition checks how the system behaves with nosy data. In the end, based on the results mentioned in the previous article all these requirements validate the similarity.

Chapter 2

Sentiment Analysis for Textual Information

Sentiment Analysis can be modeled as a classification problem that implies *subjectivity* classification (classify opinions into subjective and objective ones), *polarity* classification (classify expressions into negative, positive, and neutral), or opinion spam detection.

Sentiment Analysis, visualized as a polarity classification task, implies three main stages: the initialization one, the learning, and the evaluation phase. The initialization step prepares the data for the classification algorithm. Data collection means retrieving data and analyzing its content. A manual annotation is required to validate the approach if the data still needs to be labeled. The pre-processing phase means transforming unstructured information into a clear one without misspellings, abbreviations, or slang words. Then, a feature extraction stage determines how data is modeled regarding relevant features. The data representation step uses a word embedding technique to convert the models into numerical representations. Then, the learning step describes when a training model is passed to a Machine Learning algorithm. The last step is the evaluation, when performance measures are computed to reflect how good the methodology is.

Twitter (starting with July 2023, rebranded to **X**) is a popular social media platform for communicating with others, expressing feelings and opinions, and broadcasting news. The advantages of such a powerful tool are the availability of different electronic devices, the opportunity to have a large friend pool, and the fact that you can send small and concise messages to other friends on various subjects [16].

The main concepts used in a Twitter environment are URL, mention, user, friend, follower, tweet, recency, hashtag, emoticon, re-tweet, and singleton [13]. A tweet is a simple text message of a maximum of 280 characters, and it can contain hashtags and hypertext-based elements: related media (maps, photos, videos) and websites. Hashtags [16] are keywords prefixed with the "#" symbol that can appear in a tweet. Twitter users use this notation to categorize their messages and enable or mark them to be more easily found in search.

Chapter 3

Features and Models For Short Messages

Even though a tweet has text, hashtags, and hypertext elements, we analyze only the first two because hypertexts offer diverse information from a type perspective (links, videos, etc.) and a semantic one (meaning). Besides this, the multimedia concepts are only sometimes present in each tweet, so the dataset would be considerably decreased for the analysis. Four perspectives are outlined and applied to the Twitter polarity classification task.

The **Baseline-Based Sentiment Analysis (BSA)** feature implies that the input is consisted of textual information without the keywords that define a Twitter message. The **Hashtag-Based Sentiment Analysis (HSA)** feature is built by extracting the hashtags from a tweet. It maintains a list of hashtags (indicators) for each message. The **Fused-Based Sentiment Analysis (FSA)** approach combines the previous ones. The input for a classification algorithm will be represented by the text (without hashtags) concatenated with the list of hashtags. The **Raw-Based Sentiment Analysis (RSA)** feature describes the input as a raw text where the # sign for the hashtags is removed. If the # sign is removed, then the word becomes an ordinary one and will be processed like the others in the preprocessing step.

Another essential step in our features-related experiments is designing a lexicon-based feature for a tweet polarity classification problem considering English messages. In the previous attempts, we detected positive and negative tweets, but now we also included neutral sentiment. In addition, a simple tweet message can express a specific polarity, but only a few words remain relevant due to the short amount of words and after some preprocessing steps. Even though the hashtags indicate the message's sentiment, we decided to strengthen the message by giving a hint. This hint is a rule-based technique that determines a sentiment score as the output of a lexicon. Concerning this score, we can extract a sentiment (e.g., positive, negative, or neutral). In other words, we start from the original input, pass it to a lexicon, and enhance the text with an indicator, like a hashtag, but, for this situation, it is called a **sentiment indicator**. In the end, the enhanced information is passed to a classifier that can decide the actual value of the tweet.

Next, we define a system combining hybrid features with lexicon-based ones. The designed system, **Twitter-Lex SA**, aims to explore the information offered by the Twitter platform in combination with the use of a sentiment-based lexicon. Analyzing only one set of features (e.g., lexical ones) is insufficient for a good classification. In most cases, textual input is not enough when discussing the Twitter platform since relevant features can highlight the message (e.g., hashtags, mentions). Moreover, context can be essential, and combining words within the sentence and other features can change the overall polarity of input. Bearing in mind all these things and starting from the previous features (hybrid and lexicon-based), four features are defined:

- Baseline Sentiment Analysis-Lexicon (BSA_{lex})
- Hashtag Sentiment Analysis-Lexicon (HSA_{lex})
- Fused Sentiment Analysis-Lexicon (FSA_{lex})
- Raw Sentiment Analysis-Lexicon (RSA_{lex})

Overview of the experiments

An overview of the defined approaches is presented in the **Table 3.1** considering the used datasets, Machine Learning classifiers, evaluation measure and the use of a sentiment lexicon.

Table 3.1: Overview of Twitter Sentiment Analysis Features.

Approach	**Dataset**	**ML Classifier**	**Lexicon**	**Evaluation measure**
Hybrid features (BSA, HSA, FSA, RSA)	Sanders	SVM, LR and NB	Not used	Accuracy, precision
Lexicon feature	Apple Twitter Sentiment, Sanders, Twitter US Airline	SVM, LR, and NB	Text Blob, aFINN, Vader and Senti WordNet	Accuracy, precision
Twitter-Lex features	Apple Twitter Sentiment, Sanders, Twitter US Airline and Twitter Climate Change	SVM, LR and NB	Vader	Accuracy, precision

Chapter 4

Data Representations for Unsupervised Sentiment Analysis

The United States 2020 presidential election was watched worldwide due to the high polarities between the fans of the finalist candidates, Joe Biden and Donald Trump. Social media was the most used environment to express their feelings and opinions about their favorite or offend and attack the opposite candidate. Twitter's social media tool encapsulated the different ideas or thoughts that built an overview of the whole period of the campaign and the presidential debates.

Based on these aspects, Sentiment Analysis can be applied to explore Twitter's world to define trends and directions regarding the pre and post-event 2020 election.

In the current approach, we design and implement a system that offers an unsupervised perspective for recognizing the sentiment dominant in the tweets posted about the last two presidential candidates. The goal is to define two groups of messages (positive and negative) that can show us their opinion about Donald Trump and Joe Biden.

Therefore, the summary of the work is represented by the next phases:

- we define a new tweet model called **hash index**, a representation that takes into account the hashtags of a message;
- we analyze the impact of data representation for the entire process. We apply the TF-IDF model to a hashtag representation for a tweet, respectively, on the extracted list of hashtags;
- we use clustering algorithms on unlabeled data sets for detecting the relevant groups of sentiments for each presidential candidate;
- we extend the validation process of the entire approach by adding the external validation step by using an automated labeling process based on the Vader lexicon.

The subsequent experiments are focused on determining if different models or data representations are suitable to identify the sentiment of tweets in the unsupervised context. Applying different clustering algorithms, we want to define two groups of messages (one positive and one negative) by using two new models. Bearing in mind that in Twitter's world, hashtags represent an essential feature since they are indicators of the message, we define a hashtag representation that will use this concept determined from the tweet. On the other side, a text-based representation is built based on the idea that the text (without hashtags) composes a relevant input for the sentiment detection problem. Furthermore, in the numerical experiments, we determine which model is better for sentiment classification in the unsupervised context.

Overview of the experiments

An overview of the previously described approaches is presented in the **Table 4.1** considering the used datasets, clustering algorithms and evaluation measure.

Table 4.1: Overview of Unsupervised Twitter Sentiment Analysis

Approach	**Dataset**	**Clustering algorithm**	**Internal evaluation measure**	**External evaluation measure**
Hash index word embedding	US 2020 Presidential tweets	K-Means, Agglomerative	Silhouette, Davies-Bouldin	Accuracy
Text-based and hashtag-based representations	Joe Biden, COVID-19 and Republican Presidential debate dataset	K-Means, Agglomerative, Spectral	Silhouette, Davies-Bouldin	Accuracy

Chapter 5

Applications of Sentiment Analysis

This chapter handles longer messages, mainly reviews, and explains how Sentiment Analysis can be combined with Recommender Systems to enhance the quality of a recommendation process. The main goal is to combine the Sentiment Analysis process with a user-based collaborative filtering technique to analyze if considering the sentiment of a review can increase the quality of recommendations. The *Yelp Restaurant Reviews* [1] is used as input for the system. Then, data is passed to the Sentiment Analysis module, which aims to classify the collected reviews into positive and negative ones. These labeled data will represent the input for the recommender module that handles the collaborative filtering part that will generate, in the end, a list of recommendations for the users. The final results are presented in [14]. In this chapter, only the sentiment analysis module is detailed since it is our contribution.

The next interesting phase in our process for combining Sentiment Analysis with Recommender Systems for reviews implies defining a sentiment score rating approach. These scores are used for a newly designed similarity, called **ARP (Attractiveness-Relevance-Popularity)** that is integrated into the **k-Nearest-Neighbors (kNN)** collaborative filtering approach [1] for the recommendation process. After collecting the data (in this case, reviews), the textual input is passed to the **Sentiment Scoring** module. This component aims to compute the sentiment value for the textual information provided as input. The collected data is preprocessed and passed to a scoring function. The scoring function uses the *Senti WordNet* tool[2] to calculate the sentiment for each word. The sentiment score is then mapped to a one-to-five rating value. The obtained sentiment ratings are to be used instead of the real user ratings.

The necessity of a similarity measure is really important in several contexts: clustering, recommendation systems, classifications, etc. In literature, there are a lot of well-known measures like Euclidean, Cosine, Pearson or Jaccard. Since the **ARP (Attractiveness-Relevance-Popularity)** similarity measure is a new one, we need to define some validation steps that prove this measure can be included in the set of the popular measures. If the collected data set contains textual input (e.g., product descriptions, text-based reviews), we can use the ARP measure. Otherwise, the classical measures can be applied to a specific context (e.g., clustering, recommender systems, etc.). The following phases will be used for validating the ARP measure: metric conditions check, usefulness, expressivity, correlation with other measures, and noise robustness condition. In other words, the first step is to check if ARP is already a metric or can be generated as a metric. The usefulness condition checks

[1] https://www.yelp.com/dataset

[2] http://ontotext.fbk.eu/sentiwn.html

if the measure can be applied to various data sets (with or without ratings, text input, etc.). The expressivity stage determines if ARP is more suitable for text-based collections than other measures. The next step computes the correlation between ARP and the other measures. The last step implies checking how ARP behaves regarding noisy data (noise can come from classifiers/ sentiment lexicons or preprocessing techniques).

Overview of the experiments

An overview of the presented approaches is reflected in **Table 5.1** considering the dataset, sentiment classifiers and evaluation measures.

Table 5.1: Overview of the Applications of Sentiment Analysis.

Approach	**Dataset**	**Sentiment classifier**	**Evaluation measure**
SA for determining polarities of reviews	Yelp Restaurants Reviews	SVM, LR, NB	Accuracy, precision, recall, f-measure
Sentiment scoring module for ARP similarity measure	Yelp Restaurants Reviews, Datafiniti Hotel Reviews	Senti WordNet lexicon	MAE, RMSE
Validation of ARP measure	Yelp Restaurants Reviews, Datafiniti Hotel Reviews	aFINN, Text Blob, Vader, Senti WordNet	MAE
Aylien API	Sanders	deep learning	Accuracy, precision, recall, f-measure

Chapter 6

Conclusions and Future Work

Sentiment Analysis is a challenging domain that can be applied to various inputs. The focus of this thesis is the polarity classification task for textual information. Nowadays, people express their ideas regarding an event, a politician, a movie, etc., on social media. The drawback of such messages is represented by writing (slang words, use of colloquial expressions, etc.). Therefore, the primary challenge is to preprocess and clean the noisy data that will be passed to a system that handles the sentiment detection of such messages. In addition to this phase, feature extraction and data modeling represent critical phases due to the diversity of attributes that can be extracted from texts. Even though we use the same Machine Learning classifiers, different features, and models can produce different results.

Therefore, the first focus of our experiments is on applying Sentiment Analysis techniques to Twitter messages, called tweets. The main work is represented by building new features and representations to increase classification quality. So, four features are designed based on Twitter-specific elements.

The next step in our approach is to define a lexicon-based feature for detecting the sentiment of tweets. Hence, several sentiment lexicons (aFINN, Vader, Senti WordNet, and Text Blob) are used throughout the process. The enhanced tweet is the preprocessed tweet combined with a **sentiment indicator**, which is the sentiment determined based on the sentiment score computed by a lexicon. The last phase of our experiments is to merge the **BSA, HSA, RSA** and **FSA** features and the lexicon-based one. This fusion is represented by the **Twitter-Lex Sentiment Analysis** system. The main idea is that every Twitter feature (BSA, FSA, HSA, and RSA) is concatenated with the Vader lexicon's sentiment indicator.

The next step of the experiments is to switch from supervised learning to unsupervised context. So, several clustering algorithms are applied to US 2020 presidential tweets to determine groups of messages that will indicate the corresponding sentiment. Moreover, a new word embedding model is defined, called **hash index**, a representation that considers the hashtags from a tweet. The whole clustering process is evaluated with both internal and external measures. Next, a detailed comparison between two representations is conducted for unsupervised learning. The text-based and hashtag-based features are used on several datasets where a clustering technique is applied.

The second important purpose of the presented approaches is to highlight Sentiment Analysis's applications in other areas, mainly on Recommender Systems. Consequently, the Sentiment Analysis process is applied to texts of longer dimensions than tweets and reviews posted by users. The rating feature from Recommender Systems (usually a one-to-five rating given by the user for a specific item/product) will be replaced or enhanced with the sentiment derived by applying Machine Learning classifiers. Next, a new similarity measure called **ARP (Attractiveness-Relevance-Popularity)** is defined. A sentiment rating function determines a sentiment score for a review based on the Senti WordNet lexicon. This sentiment is considered for building the measure. In the end, the ARP measure is validated based on five conditions.

The plan considers extending the existing experiments by applying the designed features and representations on multiple datasets, taking into account the size and the noise. Moreover, a detailed comparison with the existing approach in the literature (e.g., BERT models) and combining the features with these are needed. On the other side, we want to explore a transformer network-based classifier for determining the polarity of a tweet. Also, the features offered by Twitter are not mined enough so that we can consider several aspects, like retweets, mentions, or replies. Regarding the presidential tweets, we want to generalize the process to use the methodology for future elections and proceed with an analysis based on a specific time frame (e.g., the impact of USA elections over ten years) by defining some mathematical models for the entire system. A general framework for these features can be designed and implemented to determine general tendencies and improve the quality of Recommender Systems, clustering techniques, or Machine Learning classifiers. Finally, the Sentiment Analysis area for textual input is still challenging, with many opportunities to explore.

Bibliography

[1] R Baeza-Yates. Modern information retrieval. *Addison Wesley google schola*, 2:127–136, 1999.

[2] Mara Deac-Petruşel and Sergiu Limboi. A sentiment-based similarity model for recommendation systems. In *2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, pages 224–230. IEEE, 2020.

[3] Hussam Hamdan, Frederic Béchet, and Patrice Bellot. Experiments with dbpedia, wordnet and sentiwordnet as resources for sentiment analysis in micro-blogging. In *Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 455–459, 2013.

[4] Clayton Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225, 2014.

[5] Sergiu Limboi. Comparison of data models for unsupervised twitter sentiment analysis. *Studia Universitatis Babeş-Bolyai Informatica*, 67(2):65–80, 2023.

[6] Sergiu Limboi and Mara Deac-Petruşel. A validation framework for arp similarity measure. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1266–1271. IEEE, 2021.

[7] Sergiu Limboi and Laura Dioşan. Hybrid features for twitter sentiment analysis. In *Artificial Intelligence and Soft Computing: 19th International Conference, ICAISC 2020, Zakopane, Poland, October 12-14, 2020, Proceedings, Part II 19*, pages 210–219. Springer, 2020.

[8] Sergiu Limboi and Laura Dioşan. A lexicon-based feature for twitter sentiment analysis. In *2022 IEEE 18th International Conference on Intelligent Computer Communication and Processing (ICCP)*, pages 95–102. IEEE, 2022.

[9] Sergiu Limboi and Laura Dioşan. The twitter-lex sentiment analysis system. In *Proceedings of the 14th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, IC3K 2022, Volume 1: KDIR, Valletta, Malta, October 24-26, 2022*, pages 180–187. SCITEPRESS, 2022.

[10] Sergiu Limboi and Laura Dioşan. An unsupervised approach for twitter sentiment analysis of usa 2020 presidential election. In *2022 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, pages 1–6. IEEE, 2022.

[11] Walaa Medhat, Ahmed Hassan, and Hoda Korashy. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4):1093–1113, 2014.

[12] Finn Årup Nielsen. A new anew: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*, 2011.

[13] Kishori K Pawar, Pukhraj P Shrishrimal, and RR Deshmukh. Twitter sentiment analysis: A review. *International Journal of Scientific & Engineering Research*, 6(4):9, 2015.

[14] Mara-Renata Petrusel and Sergiu-George Limboi. A restaurants recommendation system: Improving rating predictions using sentiment analysis. In *2019 21st International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, pages 190–197. IEEE, 2019.

[15] Gloria Phillips-Wren and Monica Adya. Decision making under stress: The role of information overload, time pressure, complexity, and uncertainty. *Journal of Decision Systems*, 29(sup1):213–225, 2020.

[16] Jagan Sankaranarayanan and et al. Twitterstand: News in tweets. In *Proceedings of the 17th ACM SIGSPATIAL*, GIS '09, pages 42–51. ACM, 2009.

[17] Bhupender Singh Shekhawat. *Sentiment classification of current public opinion on brexit: Naïve Bayes classifier model vs Python's Textblob approach.* Phd thesis, Dublin, National College of Ireland, 2019.