The Twitter-Lex Sentiment Analysis System

Sergiu Limboi and Laura Dioşan
Faculty of Mathematics and Computer Science, Babeş-Bolyai University, Cluj-Napoca, Romania

Keywords: Twitter Sentiment Analysis, Hashtag-based Features, Lexicon Features, Data Representation.

Abstract: Twitter Sentiment Analysis is demanding due to the freestyle way people express their opinions and feelings. Using only the preprocessed text from a dataset does not bring enough value to the process. Therefore, there is a need to define and mine different and complex features to detect hidden information from a tweet. The proposed Twitter-Lex Sentiment Analysis system combines lexicon features with Twitter-specific ones to improve the classification performance. Therefore, several features are considered for the Sentiment Analysis process: only textual input from a tweet, hash-tags, and some flavors that combine them with the feature defined based on the result produced by a lexicon. So, the Vader lexicon is used to determine the sentiment of a tweet. This output will be appended to the four perspectives we defined, considering the features offered by Twitter. The experimental results reveal that our system, which focuses on the role of features in a classification process, outperforms the baseline approach (use of original tweets) and provides good value to new directions and improvements.

1 INTRODUCTION

Nowadays, social media is gaining ground due to the high availability, variety of ways to express your feelings, opinions, and fast access to much information. The Twitter platform is one of the most used environments for posting messages (Antonakaki et al., 2021). The critical concept around Twitter is represented by tweets which are messages consisting of a maximum of 280 characters. In addition, a message can be characterized by hashtags which are words prefixed with the # symbol and indicate the concepts or relevant keys of the message. Furthermore, a Twitter user can build connections with other users who share the same interest, developing a friend-follower relationship.

Considering these aspects, there is an increased need to analyze users' opinions to define future predictions, observe trends regarding an event or famous star, etc. Therefore, a Sentiment Analysis can be approached by following and depicting text polarities (positive, negative, and neutral) from tweets to build an overview of a relevant topic.

Various approaches use Sentiment Analysis in literature to classify textual information, especially tweets. An interesting phase of the sentiment detection task is determining the features involved in the process. According to Koto (Koto and Adriani, 2015) there are several categories regarding the features that can be defined for the classification of messages:

- punctuation: number of ”?”, ”!” or another special character;
- lexical features: the size of a tweet, number of lowercase words, number of hashtags;
- Part-of-speech (POS): number of nouns, verbs, adverbs, or adjectives;
- emoticon scores;
- sentiment scores from sentiment lexicons.

Another classification of features is done by Carvalho (Carvalho and Plastino, 2021) based on three main categories:

- n-gram features;
- meta-level features: POS, linguistic features, or emoticon;
- word embedding-based features: DeepMoji, Emo2Vec, fastText, GloVe-TWT.

The proposed **Twitter-Lex Sentiment Analysis** system uses a mix of Twitter-specific features (hashtags) and lexicon features for determining the sentiment (positive, negative, or neutral) of tweets. Consequently, the original contributions of this paper are the following:

- new features applied for Twitter Sentiment Analysis by using the knowledge provided by a lexicon

Limboi, S. and Dioşan, L.
The Twitter-Lex Sentiment Analysis System.
DOI: 10.5220/0011379900003335
In *Proceedings of the 14th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2022) - Volume 1: KDIR*, pages 180-187
ISBN: 978-989-758-614-9; ISSN: 2184-3228

and mining different aspects of a tweet. The features derived from tweets are text-based, hashtag-based, text concatenated with hashtags, and text with hashtags without the # indicator. All of them are merged with the sentiment label provided by Vader lexicon;

- a complex study of the enhancement consisted of applying the whole process on several datasets and considering multiple scenarios.

The remainder of the paper is organized as follows. Section two presents various approaches that use several features or analyze the impact of features for Twitter Sentiment Analysis. The overview of the system is highlighted in section three, followed by the fourth one that describes in detail the proposed approach. The numerical experiments and the results are reflected in the fifth section. Then, a comparison between viewpoints is made in the sixth one. Conclusions and future work are drawn in the last area.

2 RELATED WORK

Developing systems for Sentiment Analysis and evaluating the high impact of feature sets or different types of inputs/ data models was approached in various ways in the literature. Therefore, this section presents several perspectives that focus on the features and models used for the Twitter polarity classification problem.

The approach from Chiong (Chiong et al., 2021) aims to detect the depression hidden in tweets. The posted messages are analyzed based on a combination of features. This mix consists of components resulting from the sentiment lexicon and content-based Twitter-specific features. The data sets from Shen's and Eye's perspectives (Shen et al., 2017) are used for the methodology. Tweets are marked as indicating "Depression" (negative sentiment) or "Non-depression" (positive view). Six feature groups are defined for the depression detection task. Three groups contain features based on the sentiment lexicons, and three use platform-specific features. So, the first three groups (A, B, and C) have attributes from SentiWordNet and SenticNet libraries (e.g., number of positive, negative, or neutral words). The remaining groups have basic tweet information (e.g., the number of words, the number of links), part-of-speech (POS) features, and linguistic attributes (e.g., the ratio of adverbs and adjectives, school-level indicator for text understanding, etc.). After the feature extraction process, data is split into training and test and passed to four different classifiers: Support Vector Machine, Logistic regression, Decision Tree, and Multilayer Perceptron. The best classifier is detected based on evaluation measures (accuracy, precision, recall, and f-score).

The aim of Rani's perspective (Rani et al., 2021) is to analyze the impact of features' size on a sentiment classification for the Twitter US Airline dataset. Moreover, the feature selection technique is examined to see what method best fits a polarity detection problem. The designed system collects the messages and applies cleaning and preprocessing techniques. After this phase, Chi-Square and Information Gain are used as feature selection techniques for defining feature sets with different dimensions. In addition, a sentiment score is added to each feature set by using a sentiment lexicon. The enhanced model is passed to various machine learning classifiers (Naïve Bayes, SVM, or decision trees), and the results are evaluated via accuracy and Kappa metric.

The approach from (Ayyub et al., 2020) applies Sentiment Analysis to determine the "relative frequency" of a sentiment label, called "sentiment quantification." This methodology is divided into two main phases: sentiment classification task and computing the frequency of the target class, also known as the class of interest. The analysis aims to determine the impact of linguistic features on the whole process and compare different classification techniques. The designed system handles three types of feature extraction methods. Firstly, the bag of words is converted into TF-IDF values. The second approach uses n-grams (here, words have assigned probabilities). The last experiment involves the combination of the two methods. Standford Twitter Sentiment, STS-Gold, and Sanders are used as datasets. The experiments handle different feature sets based on the previously mentioned techniques and use different classifiers such as traditional machine learning approaches or deep learning. Moreover, absolute error or relative error are determined as evaluation measures.

Onan et al. (Onan, 2021) explores the sentiment classification issue for Turkish tweets. In addition, it analyzes different word embedding-based features using supervised learning algorithms (e.g., Naïve Bayes, SVM) and ensemble learning techniques (e.g., AdaBoost, Random Subspace). The proposed system defines nine weighting schemes. Two are unsupervised (TF-IDF or term frequency). Seven are supervised: odds ratio, relevance frequency, balanced distributional concentration, inverse question frequency-question frequency-inverse category frequency, short text weighting, inverse gravity moment, and regularized entropy (Onan, 2021). Tweets were collected for two months via Twitter API to build the data set. A manual annotation phase determines if a message is positive or negative. After the preprocessing stage, the list of words that compose the tweet is passed to the weighing scheme, and a sentiment classifier handles the result. The system is evaluated using accuracy, precision, recall, and f-score.

The **Twitter-Lex Sentiment Analysis** system we propose combines the information offered by a lexicon with Twitter-specific features for enhancing the quality of the polarity classification problem. The overview and methodology will be presented in the following sections.

3 PROPOSED APPROACH

3.1 System Overview

The **Twitter-Lex Sentiment Analysis** system, or simply Twitter-Lex SA, has the goal to do a polarity classification process for tweets, in terms of three classes: positive, negative, and neutral. The following phases, illustrated in **Figure 1**, compose the entire architecture:

- data collection
- data preprocessing
- tweet enhancement by using the Vader lexicon
- data representation
- classification by using a Machine learning algorithm

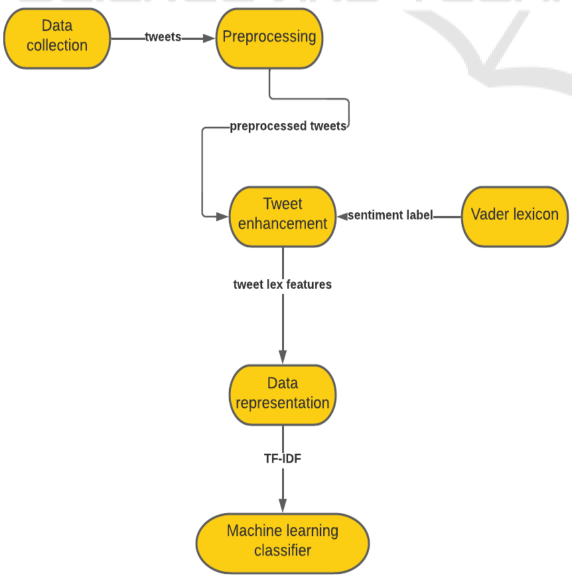

Figure 1: System Overview.

The data collection step means researching the datasets containing labeled tweets. These labels represent the ground truth in the learning process and are not used for the testing phase of tweet classification. The preprocessing stage is critical because the system handles textual information. Due to abbreviations and free-word style, some cleaning mechanisms are needed. The following techniques are used in this phase: lowercasing, removing punctuation and stop words, and stemming. Even though emojis are essential elements on social media and are potential indicators for sentiments, the preprocessing phase does not handle them since the used datasets have only a few emojis.

The tweet enhancement phase means defining four new features by using the well-known sentiment lexicon, Vader (Hutto and Gilbert, 2014), and Twitter-specific elements. This process is described in the following sub-sections. On the other hand, the tweet enhancement approach can be used in combination with other lexicons by adapting the way of defining the sentiment scores, considering the lexicon's features. The enhanced $tweet_{lex}$ is converted into a numerical representation needed for a classifier. Therefore, a TF.IDF representation (Baeza-Yates et al., 1999) is used due to its easy way of computation.

As Machine learning classifier, three algorithms are used, very popular in the literature, Logistic regression (LR) (Kleinbaum et al., 2002), Support Vector Machine (SVM) (Suthaharan, 2016) and Naïve Bayes (NB) (Webb et al., 2010). The goal is to classify the enhanced tweets into positive, negative, or neutral. As evaluation measures, accuracy and precision are computed. In addition, 95% confidence intervals are determined to define the interval of values for the model's performance.

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 to provide a good classification. In most cases, textual input is not enough when we talk about the Twitter platform since there are relevant features that can highlight the message (e.g., hashtags, mentions, etc.). Moreover, the context can be essential, and the combination of 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 approach from (Limboi and Dioşan, 2020), four features are defined as follows:

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

3.2 Vader Sentiment Lexicon

The **Vader (Valence Aware Dictionary and Sentiment Reasoner)** lexicon (Hutto and Gilbert, 2014) is used for the Twitter-based perspectives defined for the process of tweets' classification. This lexicon uses a combination of words marked based on their semantic orientation. For each word, it gives a compound score that means the sum of a word's positive, negative, and neutral values, followed by the normalization of the sum in the range $[-1,1]$. A sentiment score closer to 1 highlights a good positivity of the word compared to a value closer to -1, which means vast negativity. Given a tweet *t*, its Vader sentiment score will be represented by the sum of the compound scores of its words.

$$s_{Vader}(t) = \sum_{i=1}^{m} s_{Vader}(word_i), \tag{1}$$

where

- m is the number of words of a tweet
- $s_{Vader}(word_i)$ is the sentiment score of the i^{th} word, score provided by Vader sentiment lexicon.

The next step is determining the tweet's sentiment according to the sentiment score. The threshold *0.05* was set considering the search from (Hutto and Gilbert, 2014) that states best values are achieved by using this value.

$$sent_{Vader}(tweet) = \begin{cases} positive, & \text{if } s_{Vader}(t) > 0.05 \\ negative, & \text{if } s_{Vader}(t) < 0.05 \\ neutral, & otherwise \end{cases} \tag{2}$$

where $s_{Vader}(t)$ is the sentiment score of the tweet provided by the Vader lexicon (from Eq. 1).

3.3 Baseline Sentiment Analysis-Lexicon Feature

The Baseline-Lexicon Sentiment Analysis (BSA_{lex}) feature implies the fact that the input is represented by the tweet where the hashtags are removed and concatenated with the sentiment derived from the Vader lexicon. It contains only textual information, without the keywords that define a Twitter message. For example, if there is the tweet "#Beautiful #goodconcert this concert was the best from my life #goodvibe feeling awesome," the enhanced message, considering the preprocessing step, will be "concert life feel awesome positive."

Therefore, considering a collection of tweets $T = \{tweet_1, tweet_2, ..., tweet_n\}$ and a set of labels $L = \{positive, negative, neutral\}$, where n is the number of tweets and $tweet_i$ is a message that contains textual information and hashtags, the *i-th* $tweet^i_{BSAlex}$ will be defined as follows:

$$tweet^i_{BSAlex} = \{word^i_1, word^i_2, ..., word^i_m, lex^i_{label}\}, \tag{3}$$

where

- m is the number of words for the *i-th* tweet
- lex_{label} is a value from the set *L*, value determined using the Vader lexicon over the words of a tweet.

3.4 Hashtag Sentiment Analysis-Lexicon Feature

The Hashtag Sentiment Analysis-Lexicon (HSA_{lex}) perspective defines as input only the hashtags that are extracted from a tweet, concatenated with the sentiment derived from the Vader lexicon. The input will be represented by a list of hashtags followed by the sentiment provided via Vader. Seeing the previously mentioned tweet "#Beautiful #goodconcert this concert was the best from my life #goodvibe feeling awesome", the enhanced one will be "beautiful goodconcert goodvibe positive". Therefore, considering a collection of tweets T and a set of labels L (similar to the previous ones), the *i-th* $tweet^i_{HSAlex}$ will be defined as follows:

$$tweet^i_{HSAlex} = \{hashtag^i_1...., hashtag^i_p, lex^i_{label}\}, \tag{4}$$

where

- p is the number of hashtags for the *i-th* tweet
- lex_{label} is a value from the set *L*, value determined using the Vader lexicon over the *HSA* representation of $tweet^i$

3.5 Fused Sentiment Analysis-Lexicon Feature

The Fused Sentiment Analysis-Lexicon (FSA_{lex}) 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 and the sentiment provided by the Vader lexicon. In other words, if there is the same tweet "#Beautiful #goodconcert this concert was the best from my life #goodvibe feeling awesome", the new one will be "concert life feel awesome beautiful goodconcert goodvibe positive". Generally, considering T and a set of labels L (similar to the previous ones), the *i-th* $tweet^i_{FSAlex}$ will be defined as follows:

$$tweet^i_{FSAlex} = \{word^i_1, .., word^i_m, hashtag^i_1, .., hashtag^i_p, lex^i_{label}\}, \tag{5}$$

where

- p is the number of hashtags for the *i-th* tweet
- m is the number of words for the *i-th* tweet
- $hash_i$ is the *i-th* hashtag of the $i-th$ tweet
- lex_{label} is a value from the set L, value determined using the Vader lexicon over the *FSA* tweet representation (words concatenated with hashtags)

3.6 Raw Sentiment Analysis-Lexicon Feature

The Raw Sentiment Analysis-Lexicon (RSA_{lex}) feature describes the input as a raw text where the # sign for the hashtags is removed. Additionally, the sentiment from the lexicon is appended to the raw text. If the # sign is removed, then the word becomes an ordinary one and will be processed like the others in the preprocessing step. So, for the tweet"#Beautiful #goodconcert this concert was the best from my life #goodvibe feeling awesome", the enhanced one will be "beautiful goodconcert concert life goodvibe feel awesome positive".

All in all, considering $T = \{tweet_1, tweet_2, ..., tweet_n\}$ and a set of labels $L = \{positive, negative, neutral\}$, the *i-th* $tweet^i_{RSAlex}$ will be defined as follows:

$$tweet^i_{RSAlex} = \{word^i_1, ..., word^i_m, hashname^i_1 ... hashname^i_p, lex^i_{label}\}, \quad (6)$$

where

- p is the number of hashtags for the *i-th* tweet
- m is the number of words for the *i-th* tweet
- $hashname_i$ is the *i-th* hashtag of the tweet without the # sign
- lex_{label} is a value from the set L, value determined using the Vader lexicon over the *RSA* representation of the tweet

3.7 Overview of the Features

All in all, considering the sentence "#Beautiful #goodconcert this concert was the best from my life #goodvibe feeling awesome", the following features can be extracted:

- BSA_{lex}: concert life feel awesome positive
- HSA_{lex}: beautiful goodconcert goodvibe positive
- FSA_{lex}: concert life feel awesome beautiful goodconcert goodvibe positive
- RSA_{lex}:beautiful goodconcert concert life goodvibe feel awesome positive

4 EXPERIMENTS

The experiments are conducted on the previous four features, considering multiple datasets.

4.1 Data Sets

For the experiments, four datasets are used: *Apple Twitter Sentiment* (Pandey et al., 2017), *Sanders* dataset (Sanders, 2011), *Twitter US Airline* (Rane and Kumar, 2018) and *Twitter Climate Change Sentiment* dataset [1] from the Canadian Foundation for Innovation, University of Waterloo. Bearing that the approach handles hashtags, all tweets that do not have hashtags are removed from the datasets. During the classification task, 60% of data is used for training and 40% for testing.

Apple Twitter Sentiment (Pandey et al., 2017) dataset has 782 tweets containing the tweet and the sentiment (71 positives, 142 negatives, and 562 neutral). It contains two attributes: the text (actual tweet about Apple) and the sentiment (-1 indicates a negative score, 0 means a neutral one, and one is for positive sentiment).

Sanders (Sanders, 2011) dataset contains tweets related to four topics (four big companies): Apple, Google, Facebook, and Twitter and four sentiments (positive, negative, neutral, and irrelevant). Since we handle only three polarities, the irrelevant tweets are removed. So, 2819 are used for the experiments: 519 positives, 572 negatives, and 1728 neutrals. The collection contains the following attributes: topic, sentiment, tweet id, tweet date, and tweet text (original message).

The *Twitter US Airline* (Rane and Kumar, 2018), has 2402 tweets related to messages posted in 2015 about United States airlines. It has 436 positive messages, 1551 negative, and 415 neutral ones. Furthermore, the dataset handles features like tweet id, airline sentiment (positive, negative, or neutral), text, tweet created (when the user posted the message), user timezone, or tweet location.

The last dataset, the *Twitter Climate Change Sentiment* has tweets collected between the 27th of April 2015 and the 21st of February 2018 containing four polarities: pro-climate change, anti-climate change, neutral, and links (the tweet is a link that only presents news about climate change). The links are removed from the dataset, so 6711 tweets are used for the experiments: 5005 positives (meaning that the message is "pro-climate change"), 599 negatives ("anti-climate

[1] https://www.kaggle.com/edqian/twitter-climate-change-sentiment-dataset

change"), and 1107 neutral. The dataset includes attributes like the message, tweet id, or sentiment (encoded as -1 for negative, 1 for positive, and 0 for neutral).

4.2 Results

Even though we applied three classifiers for the numerical experiments, the best results were obtained using the Logistic Regression (LR) algorithm. Therefore, the output is presented only for this technique. Tables 1,2, 3 and 4 present the average accuracy and precision along with the 95% confidence intervals, for accuracy, (CI_{acc}) for all features (BSA-Lex, HSA-Lex, FSA-Lex and RSA-Lex) and datasets (Apple Sentiment, Sanders, Twitter US Airline, and Twitter Climate Change Sentiment).

Table 1: Average Accuracy.

Dataset	BSA_{lex}	HSA_{lex}	FSA_{lex}	RSA_{lex}
Apple	84.08%	82.80%	**86.62%**	84.62%
Sanders	78.37%	74.11%	**79.79%**	79.61%
Airline	76.72%	73.60%	77.75%	**78.17%**
Climate	78.78%	76.92%	78.11%	**83.22%**

Table 2: Average Precision.

Dataset	BSA_{lex}	HSA_{lex}	FSA_{lex}	RSA_{lex}
Apple	83.36%	78.12%	**85.63%**	83.64%
Sanders	76.21%	67.51%	**76.81%**	75.17%
Airline	69.76%	67.46%	73.52%	**74.91%**
Climate	73.73%	69.19%	70.18%	**82.24%**

Table 3: 95% CI_{acc} for BSA_{lex} & HSA_{lex}.

Dataset	BSA_{lex}	HSA_{lex}
Apple	(0.815, 0.866)	(0.802, 0.854)
Sanders	(0.769, 0.799)	(0.725, 0.757)
Airline	(0.750, 0.784)	(0.718, 0.754)
Climate	(0.778, 0.798)	(0.759, 0.779)

Table 4: 95% CI_{acc} for FSA_{lex} & RSA_{lex}.

Dataset	FSA-Lex	RSA-Lex
Apple	(0.842, 0.890)	(0.821, 0.871)
Sanders	(0.783, 0.813)	(0.781, 0.811)
Airline	(0.761, 0.794)	(0.765, 0.798)
Climate	(0.771, 0.791)	(0.823, 0.841)

The results highlight some ideas drawn around two directions: features and datasets.

In terms of datasets, the following conclusions are depicted:

- Apple dataset produces the best results, but it is a tiny and un-balanced dataset;
- the worst values are obtained for the Airline dataset with many negative tweets. A potential cause for these results is that our system does not handle the special case of negation. So, further improvements are required for the preprocessing phase to reach better results;
- FSA_{lex} feature is the best for Apple and Sanders dataset, while the RSA_{lex} fits the remaining (Twitter US Airline and Twitter Climate Change).

From the feature point of view, the next outcome is described:

- HSA_{lex} feature produces the worst results, which illustrates that using stand-alone hashtags does not bring value since they are only indicators that lose their power without a tweet. Even though the hashtags are enriched with the lexicon feature, it seems that the sentiment feature from Vader is not enough to increase the classification's quality.
- BSA_{lex} feature has better results than HSA_{lex} but not as good as the FSA_{lex} and RSA_{lex}. This analysis leads us to the idea that cleaning the original tweet by removing the hashtags does not boost the polarity of the message.
- RSA_{lex} and FSA_{lex} have the best results. The values are very similar in 3 out of 4 cases (only for Twitter Climate Change, there are essential differences between features). Therefore, combining text, hashtags, and the lexicon feature is the best mix and produces valuable information. Due to the very close values, we cannot say that the order of hashtags and text played an important role. So, more complex experiments are needed to clarify this aspect. Although the # sign is removed for the RSA_{lex} feature, the hashtags still play an essential role within the message since they are indicators within the tweet.

5 COMPARISONS

5.1 Comparison with Original Feature

The best values reached by the four designed approaches will be compared with the original tweet (only the preprocessed tweet), without another kind of enhancements, by applying the Logistic Regression classifier.

The tables 5, 6, 7 and 8 present comparison between the four defined features and the original tweet. All features are better than the original one for the Apple Twitter dataset. For the remaining datasets, the original feature is better than the HSA_{lex}, strengthening the idea that stand-alone hashtags are not key

Table 5: Apple dataset.

Method	Acc	Prec
BSA_{lex}	84.08%	83.36%
HSA_{lex}	82.80%	78.12%
FSA_{lex}	86.62%	85.63%
RSA_{lex}	84.62%	83.54%
Original feature	77.66%	70.92%

Table 6: Sanders dataset.

Method	Acc	Prec
BSA_{lex}	78.37%	76.21 %
HSA_{lex}	74.11%	67.51%
FSA_{lex}	79.79%	76.81%
RSA_{lex}	79.61%	75.17%
Original feature	78.37%	74.48%

Table 7: US Airline dataset.

Method	Acc	Prec
BSA_{lex}	76.72%	69.76%
HSA_{lex}	73.60%	67.46%
FSA_{lex}	77.75%	73.52%
RSA_{lex}	78.17%	74.91%
Original feature	73.85%	65.46%

Table 8: Twitter Climate Change dataset.

Method	Acc	Prec
BSA_{lex}	78.78%	73.73%
HSA_{lex}	76.92%	69.19%
FSA_{lex}	78.11%	70.18%
RSA_{lex}	83.22%	82.24%
Original feature	78.03%	70.08%

players in the classification task. Then, we can conclude that the Twitter-Lex SA system outperforms in comparison with the original approach when a message is only preprocessed.

5.2 Comparison with Related Work

Even though there is a different setup for the experiments, our results are compared with the approach from (Rani et al., 2021) since both are using the Twitter US Airline dataset. Rani's perspective appends the sentiment score from a lexicon to each feature set (attributes from the dataset) defined based on different feature selection techniques (Information gain and Chi-square). Another difference is that the system uses several sizes of features. **Table 9** presents the best accuracy value for our features (RSA_{lex}) and the best results for the approach that uses Machine learning classifiers (302 features) defined from (Rani et al., 2021).

Table 9: US Airline dataset- Comparison with related work.

Twitter-Lex	Rani (302 features)
78.17%	78.41%

Because there are considerable differences in terms of used features, we added the smallest feature size reported in the related work. The results show similar values, indicating that our approach can be extended and validated against other directions from literature (using the same datasets, classifiers, etc.).

6 CONCLUSION AND FUTURE WORK

The **Twitter-Lex Sentiment Analysis** system defines four features: a combination of a lexicon feature from the Vader library and a Twitter-specific one (hashtags). The experimental results show that the proposed framework outperforms the views when the Sentiment Analysis process uses the original tweet. Also, the best achievements are obtained for RSA_{lex} and FSA_{lex} directions.

Our next plan implies considering multiple and diverse datasets and using more features from different lexicons (e.g., Senti Word Net or Text Blob). Another interesting viewpoint will be to explore hybridization and fusion techniques for extending the Twitter-Lex system. Moreover, we would like to compare our approach with other interesting methods from the more complex literature and evaluate the results from a statistical point of view. In addition, the process can be enhanced with more specific Twitter features like retweets, replies, or mentions. So, there are still many interesting things to do, but the designed features offer promising results.

REFERENCES

Antonakaki, D., Fragopoulou, P., and Ioannidis, S. (2021). A survey of twitter research: Data model, graph structure, sentiment analysis and attacks. *Expert Systems with Applications*, 164:114006.

Ayyub, K., Iqbal, S., Munir, E. U., Nisar, M. W., and Abbasi, M. (2020). Exploring diverse features for sentiment quantification using machine learning algorithms. *IEEE Access*, 8:142819–142831.

Baeza-Yates, R., Ribeiro-Neto, B., et al. (1999). *Modern information retrieval*, volume 463. ACM press New York.

Carvalho, J. and Plastino, A. (2021). On the evaluation and combination of state-of-the-art features in twitter sentiment analysis. *Artificial Intelligence Review*, 54(3):1887–1936.

Chiong, R., Budhi, G. S., and Dhakal, S. (2021). Combining sentiment lexicons and content-based features for depression detection. *IEEE Intelligent Systems*, 36(6):99–105.

Hutto, C. and Gilbert, E. (2014). 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.

Kleinbaum, D. G., Dietz, K., Gail, M., Klein, M., and Klein, M. (2002). *Logistic regression*. Springer.

Koto, F. and Adriani, M. (2015). A comparative study on twitter sentiment analysis: Which features are good? In *International Conference on Applications of natural language to information systems*, pages 453–457. Springer.

Limboi, S. and Dioşan, L. (2020). Hybrid features for twitter sentiment analysis. In *International Conference on Artificial Intelligence and Soft Computing*, pages 210–219. Springer.

Onan, A. (2021). Ensemble of classifiers and term weighting schemes for sentiment analysis in turkish. *Scientific Research Communications*, 1(1).

Pandey, A. C., Rajpoot, D. S., and Saraswat, M. (2017). Twitter sentiment analysis using hybrid cuckoo search method. *Information Processing & Management*, 53(4):764–779.

Rane, A. and Kumar, A. (2018). Sentiment classification system of twitter data for us airline service analysis. In *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, volume 1, pages 769–773. IEEE.

Rani, S., Gill, N. S., and Gulia, P. (2021). Analyzing impact of number of features on efficiency of hybrid model of lexicon and stack based ensemble classifier for twitter sentiment analysis using weka tool. *Indonesian Journal of Electrical Engineering and Computer Science*, 22(2):1041–1051.

Sanders, N. J. (2011). Sanders-twitter sentiment corpus. *Sanders Analytics LLC*, 242.

Shen, G., Jia, J., Nie, L., Feng, F., Zhang, C., Hu, T., Chua, T.-S., and Zhu, W. (2017). Depression detection via harvesting social media: A multimodal dictionary learning solution. In *IJCAI*, pages 3838–3844.

Suthaharan, S. (2016). Support vector machine. In *Machine learning models and algorithms for big data classification*, pages 207–235. Springer.

Webb, G. I., Keogh, E., and Miikkulainen, R. (2010). Naïve bayes. *Encyclopedia of machine learning*, 15:713–714.