Investigating Sentiment-Bearing Words- and Emoji-based Distant Supervision Approaches for Sentiment Analysis

Koena Ronny Mabokela[1,3] **Mpho Raborife**[2] **Turgay Celik**[3]

[1]University of Johannesburg, Applied Information Systems
[2]University of Johannesburg, Institute for Intelligent Systems
[3]University of the Witwatersrand, School of Electrical and Information Engineering
{krmabokela@gmail.com}

Abstract

Sentiment analysis focuses on the automatic detection and classification of opinions expressed in texts. Emojis can be used to determine the sentiment polarities of the texts (i.e. positive, negative, or neutral). Several studies demonstrated how sentiment analysis is accurate when emojis are used (Kaity and Balakrishnan, 2020). While they have used emojis as features to improve the performance of sentiment analysis systems, in this paper we analyse the use of emojis to reduce the manual effort in labelling text for training those systems. Furthermore, we investigate the manual effort reduction in the sentiment labelling process with the help of sentiment-bearing words as well as the combination of sentiment-bearing words and emojis. In addition to English, we evaluated the approaches with the low-resource African languages Sepedi, Setswana, and Sesotho. The combination of emojis and words sentiment lexicon shows better performance compared to emojis-only lexicons and words-based lexicons. Our results show that our emoji sentiment lexicon approach is effective, with an accuracy of 75% more than other sentiment lexicon approaches, which have an average accuracy of 69.1%. Furthermore, our distant supervision method obtained an accuracy of 77.0%. We anticipate that only 23% of the tweets will need to be changed as a result of our annotation strategies.

1 Introduction

South African population is widely diverse and highly multilingual (i.e. origins, cultures, languages, and religions) with distinct language groups including English and Afrikaans (Statista, 2022). The Nguni group is the largest group which includes Seswati, isiNdebele, isiXhosa, and isiZulu. In this instance, our study focuses on the second-largest group—the Sotho-Tswana group comprising Sepedi (*Northern Sotho*) (Mabokela and Manamela, 2013), Sesotho (*Southern Sotho*), and Setswana (Statista, 2022).

Sentiment analysis is a branch of natural language processing (NLP) that studies the emotion (opinions or attitudes) of text. This field has received a lot of attention which led to its numerous successful NLP technologies in various areas (Aguero-Torales et al., 2021; Mabokela et al., 2022a). For example, its popular application has been in social media monitoring, support management, customer feedback (Wankhade et al., 2022) and AI for social good (Mabokela and Schlippe, 2022a).

Emojis being used alongside text messages on social media has become increasingly popular (Jindal and Aron, 2021; Grover, 2021). In recent years there has been more work on sentiment analysis with the use of emojis or emoticons (Grover, 2021; Hakami et al., 2021; Haak, 2021). Emojis have recently become an alternative to emoticons but they differ from emoticons in that emoticons are typographical facial representations (Gavilanes et al., 2018). Emojis are used to express feelings, moods, and emotions in a written message with non-verbal elements (Kralj Novak et al., 2015; Gavilanes et al., 2018).

The use of emojis—as a standardised collection of tiny visual pictograms portrays everything from happy smiles to flags from around the world (Grover, 2021; Gavilanes et al., 2018). The modern-day emojis can be traced back to chatrooms in the 1990s. They were used in conversations to signal a smile, anger or to portray a joke or sarcastic statement (kwan Yoo and Rayz, 2021). According to Emoji Statistics[1], there were 3,633 emojis in total in the Unicode Standard as of September 2021. That means the sentiment lexicon has to be enriched with new emojis that are frequently on social media (Kralj Novak et al., 2015). Therefore, it is necessary to extend the existing emoji lexicons for sentiment labelling.

[1]https://emojipedia.org/stats/

Proceedings of the Fourth workshop on Resources for African Indigenous Languages (RAIL 2023), pages 115–125
May 6, 2023 ©2023 Association for Computational Linguistics

Many NLP systems require a labelled dataset for machine learning algorithms to produce good results. For this purpose, an annotation method that is not labour-intensive and time-consuming is required. Emojis have received much attention because of their widespread use and popularity in natural language processing (NLP) (Kejriwal et al., 2021). Emoji sentiment lexicons for other languages has been explored as an alternative method which then yielded a significant improvement in sentiment classification (Gavilanes et al., 2018; Haak, 2021; Kralj Novak et al., 2015). However, sentiment annotation and investigation of sentiment via emojis have received little attention for low-resource languages (Hakami et al., 2021).

Several emoji sentiment lexicons were produced by manual construction involving human annotators, automatically and semi-automatic with little human intervention (Grover, 2021). According to our knowledge, there is no published work on analysing the sentiment of emojis in the Bantu languages. In addition, since emojis are perceived as an important part of social media communications, incorporating them is likely to yield a higher-quality sentiment classification (Kejriwal et al., 2021).

Interestingly, emojis have been able to provide more information towards an accurate sentiment of the texts (Ayvaz and Shiha, 2017). Related work has shown that emojis help to detect and determine the sentiment label of the tweets (Go et al., 2009; Wang and Castanon, 2015; Ayvaz and Shiha, 2017; Singh et al., 2019). For this reason, it is interesting that we adopt an automatic approach that employs emoji information to reduce manual effort.

The main objective is to investigate the impact of emojis in the context of sentiment analysis. This comprises two tasks: (i) the usage of emojis to lower the manual effort in creating training data for sentiment analysis systems and (ii) the impact of emojis on the final accuracy of final sentiment analysis systems. For the pre-labelling, we even investigate a novel distance supervision approach to use emoji-based tweets to build a sentiment lexicon from scratch completely language-independent. We evaluate and compare our pre-labelling strategies with frequently used emoji sentiment lexicons provided by (Kralj Novak et al., 2015; Haak, 2021; Hakami et al., 2021). We contribute the following through our study:

- We collected a new sentiment analysis corpus for Sesotho (i.e. 6,314 tweets and 3,168 Sotho-English tweets) and added it to the SAfriSenti corpus.
- We investigate the usage of emoji sentiment lexicons in sentiment labelling strategies to reduce manual annotation effort.
- We leverage the existing emoji sentiment lexicons (Kralj Novak et al., 2015; Hakami et al., 2021; Haak, 2021) to generate a suitable sentiment lexicon for our target languages and provide a solution for the newer emojis which are not yet in the emoji sentiment lexicon.
- To cover tweets without emojis, we leverage sentiment lexicons in a cross-lingual way.
- Since, specific morphemes indicate a mood in our target languages, we also built and analyse morpheme-based language-specific sentiment taggers.

The structure of this paper is as follows: The related work will be discussed in Section 2. In Section 3, we will describe our SAfriSenti sentiment corpus and data collection strategies, as well as quality assurance. In Section 4, we describe the different sentiment annotation strategies. In Section 5, we will present the experiments and evaluation. Section 6, presents the conclusion and future work.

2 Related Studies

Recent efforts have been made to address the challenges of sentiment analysis for under-resourced languages (Mabokela et al., 2022a; Abdullah and Rusli, 2021). For example, a small number of African languages, including a few Nigerian languages (i.e. NaijaSenti Corpus) (Muhammad et al., 2022; Alabi et al., 2022) Swahili (Martin et al., 2021), Tunisian dialects (Medhaffar et al., 2017) and Bambara (Diallo et al., 2021) have been studied for sentiment analysis. Recently, SAfriSenti corpus (Mabokela and Schlippe, 2022b; Mabokela et al., 2022b) was created —it is a multilingual sentiment corpus for South African under-resourced languages. SAfriSenti Corpus is the largest Twitter sentiment corpus to date, with the goal of addressing the challenges of 11 South African languages.

Many current NLP applications are employed for social media data which solely rely on the labelled dataset, preferably manual annotation (Chakravarthi et al., 2022). However, no work has been done for these low-resource languages in the aspect of utilising emoticons or emojis for sentiment labelling. Moreover, high-resourced languages such as English, Spanish, and Arabic explored the emoji- or emoticon-based sentiment analysis with promising progress (Gavilanes et al., 2018; Hakami et al., 2021).

Many studies investigated emojis and word-based sentiment lexicons for sentiment analysis (Cortis and Davis, 2020; Grover, 2021). Additionally, some researchers used sentiment-bearing emojis to collect tweets from Twitter (Go et al., 2009; Pak and Paroubek, 2010). However, emojis for sentiment analysis of low-resource languages have received little research attention (kwan Yoo and Rayz, 2021) and a combination of emojis with sentiment lexicon for sentiment labelling is still an area for investigation. Moreover, only a few studies explored emoji sentiment lexicons for multilingual sentiment analysis (Kralj Novak et al., 2015; Gavilanes et al., 2018). A previous study by (Kralj Novak et al., 2015) created sentiment lexicons by involving 83 annotators to rate each of the 840 emoji as *positive*, *neutral* and *negative*.

Similarly, (Wang and Castanon, 2015) further analysed the impact of emoticons in constructing sentiment lexicons and also training the sentiment classifier. Although it was observed that the performance of the sentiment model increased by 15%, the findings support their claim that a small number of emoticons are powerful and accurate indicators of sentiment polarity. But according to (Guibon et al., 2016), the usage of the emojis can be expanded to numerous additional avenues such as sentiment enhancement and sentiment modification and are not only limited to sentiment expression. Similarly, (Ayvaz and Shiha, 2017) explored the impact of emojis in sentiment analysis but only focused on positive and negative sentiment polarity for the English language. (Kimura and Katsurai, 2017) investigated an automatic construction of an emoji sentiment lexicon. This technique takes the sentiment words from WordNet-Affect and determines how often they occur alongside each emoji.

Consequently, (Gavilanes et al., 2018) created an emoji sentiment lexicon using an unsupervised method based on the emoji descriptions [2]. Based on the analyses of the sentiment of informal texts in English and Spanish, they automatically created sentiment lexica with 840 emojis using the unsupervised system with sentiment propagation across dependencies (USSPAD) approach.

Recently, (Haak, 2021) demonstrated a technique that accurately and quickly identifies the emotions conveyed by emojis without manual annotation. However, a study by (kwan Yoo and Rayz, 2021) examined how emojis are used in tweets and how they can affect the tone and the sentiment of a sentence in the tweets and improved the sentiment analysis accuracy using machine learning techniques. (Hakami et al., 2021) examined the consistency of contextual emoji sentiment analysis in Arabic and European languages. They created the Arabic emoji sentiment lexicon and then compared the sentiment expressed in each of the two language families and cultures.

Some studies attempted to learn emoji embeddings to complement text word embeddings for sentiment classification tasks (Grover, 2021). First, (Eisner et al., 2016) employed a pre-trained emoji embedding strategy using positive and negative, randomly selected Unicode emoji descriptions. (Chen et al., 2018) learnt bi-sense emoji embeddings and train an attention-based LSTM for sentiment classification. By considering only the positive and negative descriptions for each Unicode emoji, the fine-tuning of emoji embeddings can be expedited. However, (Singh et al., 2019) proposed a straightforward method for processing emojis by replacing emojis with their descriptions in tweets and using a pre-trained word embedding strategy that is similar to that of the standard words. Furthermore (Liu et al., 2021) examined and evaluated the impact of supplementing emojis as additional features to improve the sentiment analysis performance. They developed an improved emoji-embedding model based on Bi-LSTM which in turn achieved the best sentiment analysis accuracy on online Chinese texts.

Our study is similar to (Gavilanes et al., 2018), and (Hakami et al., 2021) in that they utilised emoji sentiment lexicons to perform sentiment analysis on tweets with emojis. In this research, we adopt their approach, but we provide some additional steps to construct our emoji sentiment lexicon. Comparing the above-mentioned studies to our study, we utilised the existing emoji sentiment lexicons by (Kralj Novak et al., 2015; Haak, 2021) to construct the initial emoji sentiment lexicon. Simply put, we translate emojis found in the tweets into their textual descriptions and leverage existing

[2] http://emojipedia.org/

emoji sentiment lexicons to create a novel method for effective sentiment annotation of tweets.

3 Languages and Dataset

This section includes statistics regarding our *SAfriSenti*[3] corpus, from the initial collecting of raw data to the final tweets using Twitter API for Academic Research. *SAfriSenti* (Mabokela et al., 2022b; Mabokela and Schlippe, 2022b) corpus was manually annotated by 3 native speakers per target language following strict annotation guidelines. The annotators labelled tweets into 3-classes; positive, negative, and neutral. The corpus contains over 50,000 tweets. About 4% of the tweets were removed for various good reasons while 2% was retained after review. Positive tweets dominate negative tweets in *Sepedi*, *Setswana*, and *Sesotho* monolingual tweets by a higher margin. In addition, we evaluated our annotated sentiment corpus using the inter-annotation agreement metric (i.e., Krippendorff's average value of α=0.7695) which is deemed acceptable.

With a total of about 36% tweets alternating between *Sepedi* and English and 6% between Setswana and English, code-switching between native languages and English is common.

Table 2 shows an extract from the dataset with examples of tweets in *Sepedi*, *Setswana*, *Sesotho*, and English with emojis. We further provide an example for *Sepedi* and *English* code-switched tweets with their associated sentiments. Figure 1 shows examples of tweets with emojis.

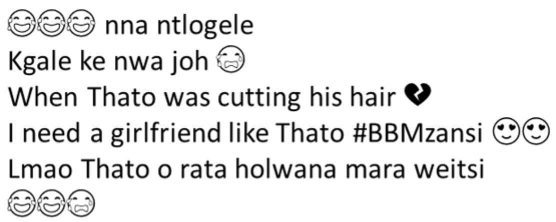

Figure 1: Examples of tweets with emojis

Table 2 presents a summary of the distribution of the tweets in this annotated subset that are monolingual and code-switched. The total monolingual tweets cover 64.4% (32,261) and 35.6% (18,223) of code-switched tweets. As demonstrated in Table 2, our corpus consists of a large number of code-switched tweets for Sepedi-English, Setswana-English and Sesotho-English. Code-switching is common between English and South African Bantu languages. 23.6% of those tweets contain code-switches of *Sepedi* and English. 5.7% of those tweets contain code switches of Setswana and English. 6.3% of those tweets contain code switches of Sesotho and English. *Sepedi*, *Setswana* and *Sesotho* share some common words since the languages are closely related. In our case, a tweet is considered a code-switched tweet if it has more than 3 English words in *Sepedi*, *Setswana* and *Sesotho* tweets.

Lang.	**Class**	#tweets	Percentage
Sepedi	**POS**	5,153	48%
	NEG	3,270	30%
	NEU	2,355	22%
	Total	10,778	
Setswana	**POS**	3,932	51%
	NEG	2,150	28%
	NEU	1,590	21%
	Total	7,672	
Sesotho	**POS**	3,050	48%
	NEG	2,024	32%
	NEU	1,241	20%
	Total	6,314	
English	**POS**	2,052	27%
	NEG	3,557	48%
	NEU	1,888	25%
	Total	7,497	

Table 1: Statistical summary of monolingual tweets for *Sepedi*, *Setswana*, *Sesotho* and English languages.

Lang.	**Class**	#tweets	Percentage
Pedi-Eng	**POS**	3,808	32%
	NEG	4,245	36%
	NEU	3,777	32%
	Total	11,830	
Tswa-Eng	**POS**	1,498	52%
	NEG	852	30%
	NEU	512	18%
	Total	2,862	
Sotho-Eng	**POS**	1,278	40%
	NEG	1,060	34%
	NEU	830	26%
	Total	3,168	

Table 2: Statistical summary of code-switched tweets for *Sepedi-English* (Pedi-Eng), *Setswana-English* (Tswa-Eng) and *Sesotho-English* (Sotho-Eng) languages.

4 Methodology

In this section, we will present the different sentiment annotation strategies that we will utilise for this study. For this, we describe how we employ our sentiment lexicons, and morphological sentiment taggers for the target languages and then also

[3]Our dataset will be made available here: https://github.com/Mabokela/SAfriSenti-Corpus

explain how we generate our novel emoji sentiment lexicons from the SAfriSenti corpus.

4.1 Words-Based Sentiment Lexicon

Numerous sentiment lexicons have been produced in various ways, including; manual creation—which is deemed to be a time-consuming and expensive process, and automatic and semi-automatic. Sentiment lexica are typically lists of words with values assigned to them that indicate the word's sentiment. Typically, these are integer values that express the polarity and intensity of the polarity as increasing or decreasing absolute values. For example, values usually range from -5: (very negative) to -1: (weakly negative) and +5: (very positive) to +1: (weakly positive). Sentiment lexicons have been used in many sentiment systems to help determine the semantic orientation of the texts (Nielsen, 2011; Hutto and Gilbert, 2015).

These sentiment lexicons have demonstrated that it is possible to combine the polarity values from a sentence and compute the sentiment on a continuous scale (Kaity and Balakrishnan, 2020). To generate word lexicon entries are chosen as a unit to associate opinion words more accurately. We used a cross-lingual approach by translating the existing English sentiment lexicons such as NRC[4] (Mohammad and Turney, 2013), VADER[5](Hutto and Gilbert, 2015) and AFFIN[6](Nielsen, 2011) to *Sepedi*, *Setswana* and *Sesotho*. Our sentiment lexicons for these targeted languages were constructed and verified by language experts. We still kept the English lexicon as our tweets contain English words. Additionally, some of the sentiment-bearing words were tagged by the annotators during the sentiment annotation process. Table 3 shows the distribution of our sentiment lexicons with words marked with sentiment polarity scores. The total number of words in the sentiment lexicon is 17,715.

Lexicons	#Words
Ours	1,250
AFFIN	7,520
VADER	2,477
NRC	6,468
Total	17,715

Table 3: Distribution of translated words in the sentiment lexicon for Sepedi, Setswana, and Sesotho.

[4]https://saifmohammad.com/WebPages/lexicons.html

[5]https://github.com/cjhutto/vaderSentiment.git

[6]https://github.com/fnielsen/afinn.git

Additionally, we used morphological sentiment taggers to tag the positive and negative tweets using morphemes with negative or positive moods. This is added to our sentiment lexicon. For example, the word *rata*, which means /love or like/ often ends with /-a/ but the ending /-e/ can signify a negation when used with the negative morphemes like /ga se/ and /rate/ which means /ga se rate/. However, the verbal ending with a vowel /-e/ can be the last component of the past tense forms as well as one of the markers of a negative mood (Prinsloo, 2020). In cases where tweets contain /ha se/ (e.g. *bona ha se wena fela motho waka /you are not my only person/*) from Sesotho (i.e. *Southern Sotho*) rather than /ga se/, our sentiment taggers presented limitations. We improved this by incorporating extra grammatical rules to compensate for this Sotho-Tswana language group scenario.

4.2 Emoji-Based Sentiment lexicons

Figure 2 shows the method for obtaining the emojis from tweets. To create the emoji sentiment lexicons, we leverage the information of the existing emoji sentiment lexicon created by (Kralj Novak et al., 2015), (Hakami et al., 2021), and (Haak, 2021). At this point, our emoji approach runs algorithms on both the tweets, emojis, and the description to automatically determine whether an emoji expresses a positive, negative, or neutral sentiment. To extract emoji-containing tweets from the SAfriSenti corpus, we only selected a subset of tweets with emojis by searching for any tweets with emojis. To create our unlabeled emoji sentiment lexicon, we follow the steps summarised below:

- We automatically extract emoji characters from the SAfriSenti corpus using regular expressions and then convert the emojis into a Unicode representation. A Unicode is a string encoding schema that translates characters into bytes.
- We create a list of unique emojis to avoid repetition in the emoji sentiment lexicons. That means emoji that repeats itself only appears once in the lexicon.
- Next, we retrieve the emoji descriptions from the python emoji translation and Emojipedia platform—an emoji dictionary in English with emoji images from different platforms. This is done by searching for the corresponding Unicode to match its description.

- To predict whether an emoji expresses a positive, negative, or neutral sentiment, we follow the approach by (Gavilanes et al., 2018) and also perform a lookup in the existing emoji sentiment lexicon and utilise the word sentiment lexicons to look up the words' polarities from their description without human intervention.

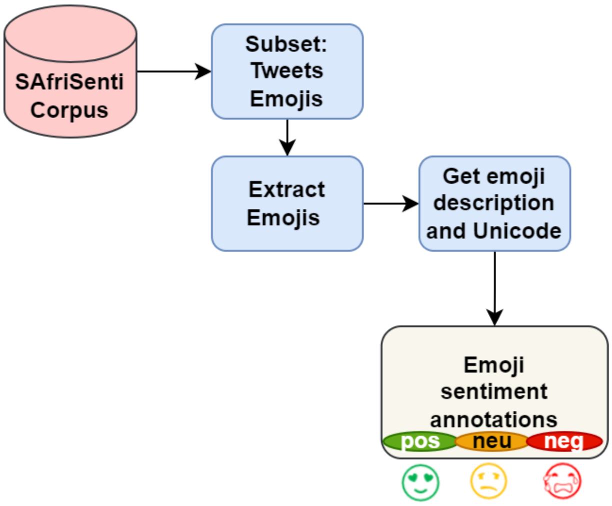

Figure 2: A process to obtain emojis and their description

In addition, we employ emoji sentiment lexicons which were obtained as follows:

- First methods are described in (Kralj Novak et al., 2015; Hakami et al., 2021). This emoji sentiment lexicon was obtained from 4% of 1.6 million tweets that were annotated (i.e. negative, neutral, positive) by 83 different native annotators for 13 European languages. It contains 751 most frequently used emojis on Twitter. The emoji sentiment lexicons were proposed as emoji sentiment rank language-independent resources for automatic sentiment analysis.

- The second method is in (Haak, 2021). This method is based on the intentions of the use of emojis for expressing the sentiment together with the methods used in (Kralj Novak et al., 2015). The emojis are statistically derived by occurrences in sentiment-bearing texts. In this case, the sentiment of the emojis is derived from the texts containing them and the sentiments are determined by using the English VADER lexicon.

4.3 Distant Supervision

In addition to our investigation, we looked at a simple and cheap process of developing an emoji and sentiment lexicon that is language-independent. As illustrated in Figure 3, 4, 5 and 6, we propose the following algorithm for sentiment labelling that leverages the information from emoji sentiment ranking[7] with sentiment-bearing emojis and words (Kranjc et al., 2015):

- $Step\ 1_{emoji_tweets}$: Use emoji unicode to identify and extract a subset of tweets with emojis (see Figure 3).

- $Step\ 2_{emojis}$: classify tweets with sentiment-bearing emojis into the classes *negative*, *neutral* and *positive* (Figure 3).

- $Step\ 3_{lists}$: create lists with sentiment-bearing words and assign a score from the translated word lexicon (Figure 4):
  1. collect all words from *negative*, *neutral* and *positive* tweets.
  2. Then remove words that occur in one or both other lists (Figure 5).

- $Step\ 4_{words}$: classify remaining tweets without sentiment-bearing emojis (e.g. tweets with no sentiment-bearing emojis) into the classes *negative*, *neutral* and *positive* based on highest word coverage with the lists of sentiment-bearing words.

- $Step\ 5_{words+emojis}$: classify all the tweets with and without sentiment-bearing emojis into the classes *negative*, *neutral* and *positive* based on highest word coverage with the lists of sentiment-bearing words and utilising the emoji sentiment lexicon scores (i.e. sentiment score [-1...+1]) from emoji sentiment lexicon (Kralj Novak et al., 2015; Hakami et al., 2021).

5 Experiments and Evaluations

In this section, we will describe our experimental setup, evaluation metrics, and the results. We also show that by acquiring the emoji sentiment lexicon from their descriptions, we then evaluate the proposed sentiment labeling framework in this section.

[7] https://kt.ijs.si/data/Emoji_sentiment_ranking/

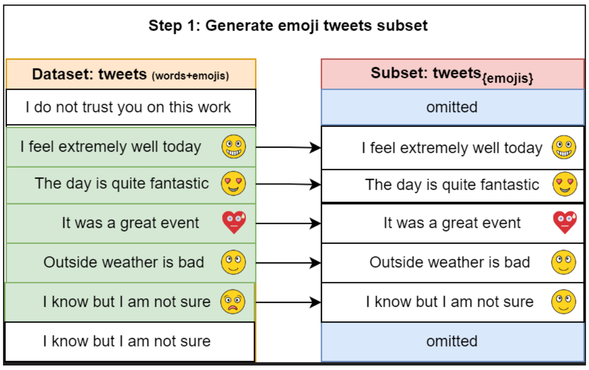

Figure 3: Use unicode to generate emoji tweets subset($step1_{emojisttweets}$).

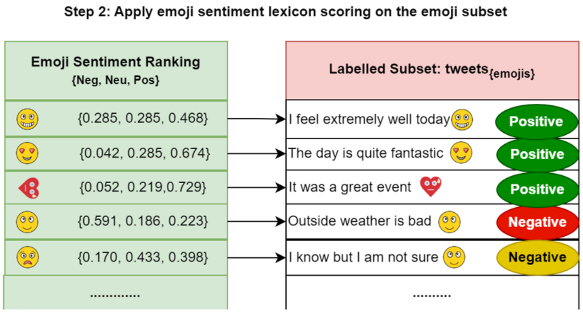

Figure 4: Classify tweets with sentiment-bearing emojis into the 3 classes ($step2_{emojis}$).

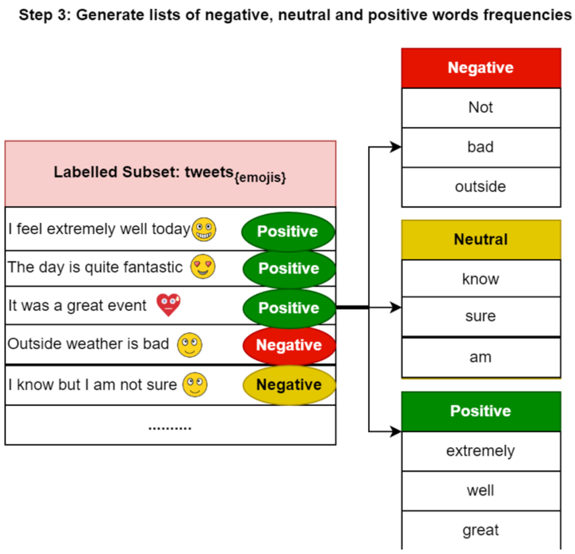

Figure 5: create lists with sentiment-bearing words ($step3_{words}$).

Our objective is to reduce manual annotation effort in creating training datasets for training NLP systems. Additionally, we investigated how to automatically create a sentiment lexicon using emoji scores from the existing emoji sentiment lexicon.

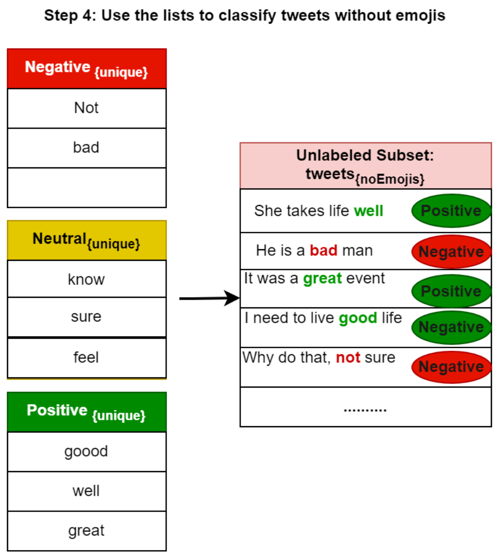

Figure 6: Sentiment-bearing words as indicators for remaining tweets' sentiment classes ($step3_{lists}$).

5.1 Experimental Setup and Metrics

We extracted the tweets with sentiment-bearing emojis for experimentation as in Figure 3. For quality assurance, all tweets have undergone a rigorous pre-processing step to remove noise, punctuations, and superfluous characters without any vital information (Mabokela and Schlippe, 2022b). In total, tweets with emojis are 34,269 which then constitute 72% of tweets in the SAfriSenti corpus. Only a small set of about 2.9% and 5.43% was found in the Setswana-English and Sesotho-English code-switched tweets while the rest of the monolingual tweets contain about 10%-24% of the tweets with emojis. Table 4 demonstrates the distributions of

Lang.	#Emojis	Freq.	Percent
Sepedi	214	7,723	22,1%
Setswana	103	5,260	15,4%
Sesotho	240	4,114	12,0%
English	287	6,103	17,8%
Pedi-Eng	370	8,200	23,9%
Tswa-Eng	78	1,008	2,94%
Sotho-Eng	64	1,861	5,43%
Total	686	34,269	

Table 4: Distribution of unique emojis and tweets with emojis per language.

tweets with emojis. A total of 686 unique emojis are identified from SAfriSenti Corpus across all the target languages. We also extracted emojis from code-switched tweets as well. We assessed our emoji tweets with the sentiment annotation methods defined in Section 4. Additionally, we use sentiment lexicon together with morpheme-based

Approaches	Features	Accuracy	Recall	Precision	F_1 score
Emoji Senti. lexicon (Kralj Novak et al., 2015)	emojis	68.7%	66.3%	65.0%	66.2%
Emoji Senti. lexicon (Hakami et al., 2021)	emojis	70.2%	72.6%	70.8%	71.7%
$SentiLexicon_{words+morph}$	words+morphemes	69.1%	67.1%	64.5%	68.4%
$SentiEmojiLex_{emojis}$	emojis	75.0%	72.6%	73.2%	74.5%
$CombSentiLex_{emojis+words}$	emojis+words	76.3%	72.6%	69.8%	73.9%
$DistSuper_{emojis+words}$	emojis+words	77.4%	76.9%	75.4%	76.3%

Table 5: Accuracy, recall, precision, and macro-F_1 score of the sentiment annotation methods.

sentiment taggers to accurately label the positive and negative moods.

We adhere to the metrics used in previous work (Gavilanes et al., 2018). We evaluated our sentiment labelling approach with accuracy, precision, recall, and macro-F_1. We evaluate our sentiment labelling strategies against manual annotations. We evaluated our approaches using accuracy, precision, recall, and macro-F_1 score using all labels as a multi-class task i.e. positive, neutral, and negative.

To obtain the tweet sentiment score associated with the sentiment label (i.e. negative, neutral, or positive), we used the discrete emoji distribution formula used in (Kralj Novak et al., 2015; Hakami et al., 2021). n emoji may appear in multiple tweets, each of which has been labeled with a sentiment. This creates a discrete distribution:

$\sum$N(c)= N, c $\in$ {-1, 0, +1}

which records the distribution of sentiment for the relevant set of tweets. The *N* denotes the number of all the occurrences of the emojis in the tweets, and *N(c)* are the occurrences in tweets with the sentiment label *c*. We considered the multiple occurrences of an emoji in a single tweet. From the above, we formed a discrete probability distribution: (P_-, P_0, P_+), $\sum$P(c) = 1.

The components of the distribution (i.e., P_-, P_0, P_+) denote the sentiment class (negative, neutral, or positive) of the emoji being identified. Then, we estimated the probabilities from relative frequencies:

$$P(c) = \frac{N(c)}{N}$$

Then, the sentiment score *S* of the emoji was calculated as the mean of the distribution:

$$S = (-1 \cdot P(-)) + (0 \cdot P(0)) + (+1 \cdot P(+))$$

In addition, the labels of the emojis are also determined from the existing emoji lexicons, and their agreement is then tested.

5.2 Results

Table 5 shows the percentages of the accuracies, recall, precision, and F_1 score measures for the 4 methods. Our results indicate that the emoji lexicon provided by (Hakami et al., 2021) performs slightly better as compared to the emoji lexicon developed by (Kralj Novak et al., 2015). Furthermore, this is because the emoji sentiment lexicon by (Kralj Novak et al., 2015) has few emojis than the one presented by (Hakami et al., 2021). The accuracy of 68.7% and F_1 score of 66.2% is considered comparable as per the previous work (Gavilanes et al., 2018). We used the sentiment lexicon ($SentiLexicon_{words+morph}$)) to classify the sentiments contained in the tweets based on words. Our $SentiLexicon_{words+morph}$ approach achieved an accuracy of 69.1% with a macro-F_1 score of 68.4%. Thus, the $SentiLexicon_{words+morph}$ performs slightly better with an increased margin of (+0.4%) compared to the emoji lexicon provided by (Kralj Novak et al., 2015).

As per the previous work (Hakami et al., 2021), emojis are classified according to categories namely; facial expressions, body language, human activity, hearts, nature, food, object and symbols, and flags. 50% of our emojis fall within the category of facial expressions having strong indicators for emotions. As shown in Table 5, Distant supervision ($DistSuper_{emojis+words}$) methods perform significantly better than the two existing emoji lexicons and word-based sentiment lexicon. Comparing our emoji sentiment lexicon ($SentiEmojiLex_{emojis}$) with the two existing emoji lexicons (Kralj Novak et al., 2015; Hakami et al., 2021), $SentiEmojiLex_{emojis}$ performs way better with an accuracy of 75%. This means that our $SentiEmojiLex_{emojis}$ approach is more effective in determining the sentiments of the tweets. Furthermore, we tested with $DistSuper_{emojis+words}$ approach—the combination of emoji-bearing sentiment and sentiment-bearing words. This $DistSuper_{emojis+words}$ approach achieved an accuracy of 77.4% as well as an F-score of 76.3%. Our results further show that our $SentiEmojiLex_{emojis}$ together with $DistSuper_{emojis+words}$ approach can be used to do language-independent sentiment labelling of tweets with and without emojis. Comparing $DistSuper_{emojis+words}$ with $DistSuper_{emojis+words}$, we obtained an increased margin of more than (+1.4%).

In addition, a combination of word-based sentiment lexicon and morphological sentiment tagger ($SentiLexicon_{words+morph}$) yielded an increase in accuracy. Our $DistSuper_{emojis+words}$ approach outperforms all the sentiment lexicon approaches used in the experiments. Our results show that obtaining the sentiment labels using emoji definitions performed better. However, using the $SentiEmojiLex_{emojis}$ approach achieved a good F_1 score of 73.9%. It is worth noting that combining emojis and words (i.e $CombSentiLex_{emojis+words}$) also improves accuracy by 1.3% compared to $SentiEmojiLex_{emojis}$. Furthermore, we were able to achieve superior results using $CombSentiLex_{emojis+words}$ compared to utilising the $SentiLexicon_{words+morph}$ approach. Additionally, there is no significant difference in the results obtained for F_1 score, Precision, and Recall in this corpus. This confirms the quality of the SAfriSenti corpus.

6 Conclusion

In this paper, we describe the different sentiment labelling strategies that involve the utilisation of emojis and words to automatically prelabel tweets for low-resource languages. Additionally, we utilised the $SentiLexicon_{words+morph}$ plus the sentiment taggers to perform sentiment labelling. We create our $SentiEmojiLex_{emojis}$ from the existing manually annotated tweets in the SAfriSenti corpus—a multilingual Twitter sentiment corpus for South African languages (i.e. *Sepedi*, *Setswana*, *Sesotho* and English) which will later be extended to other South African languages. We created our $SentiEmojiLex_{emojis}$ by extracting only the tweets that contain emojis and converting the emojis to their corresponding textual descriptions. Furthermore, we leverage the approach by (Gavilanes et al., 2018) to develop our emoji sentiment lexicon. We achieved better accuracy and F_1 score with our $DistSuper_{emojis+words}$. Comparing our $SentiLexicon_{words+morph}$ with the sentiment-bearing words lexicon, our results show that the $SentiEmojiLex_{emojis}$ strategy is more effective and reliable. In addition, by comparing our labelling strategies to existing emoji sentiment lexicons, we obtained comparable results with an accuracy of 75% for $SentiEmojiLex_{emojis}$ and 77% of accuracy for the $DistSuper_{emojis+words}$ approach. Furthermore, we used the $CombSentiLex_{emojis+words}$ and $DistSuper_{emojis+words}$ approaches to label the remaining tweets in the *SAfriSenti* corpus (i.e. 32% (16,215 tweets)). Therefore, developing an automatic sentiment annotation strategy for tweets with emojis is more likely to reduce human annotation effort. Additionally, these methodologies can be readily adapted to other under-resourced African languages, provided the data gathered contains emojis. Our future endeavors include leveraging emoji embedding to formulate context-sensitive sentiment labeling techniques for specialized systems. Moreover, we aim to enhance sentiment classification by incorporating various active learning approaches that incorporate emojis.

7 Acknowledgments

This research work is supported by the National Research Foundation (NRF). First, we acknowledge the NRF for the Black Academics Advancement Programme (BAAP) award (No: BAAP200225506825). We also want to express our deepest gratitude to the annotators who devoted their time and effort, the language experts, and the session facilitators. Finally, we thank the valuable feedback provided by the reviewers.

References

Nur Atiqah Sia Abdullah and Nur Ida Aniza Rusli. 2021. Multilingual sentiment analysis: A systematic literature review. *pertanika journal of science and technology*, 29.

Marvin M. Aguero-Torales, Jose I. Abreu Salas, and Antonio G. Lopez-Herrera. 2021. Deep learning and multilingual sentiment analysis on social media data: An overview. *Applied Soft Computing*, 107:107373.

Jesujoba O. Alabi, David Ifeoluwa Adelani, Marius Mosbach, and Dietrich Klakow. 2022. Multilingual Language Model Adaptive Fine-Tuning: A Study on African Languages. *arXiv e-prints*, page arXiv:2204.06487.

Serkan Ayvaz and Mohammed Shiha. 2017. The effects of emoji in sentiment analysis. *International Journal of Computer and Electrical Engineering*, 9:360–369.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Vigneshwaran Muralidaran, Navya Jose, Shardul Suryawanshi, Elizabeth Sherly, and John P. McCrae. 2022. Dravidiancodemix: sentiment analysis and offensive language identification dataset for dravidian languages in code-mixed text. *Language Resources and Evaluation*, 56:765 – 806.

Yuxiao Chen, Jianbo Yuan, Quanzeng You, and Jiebo Luo. 2018. Twitter sentiment analysis via bi-sense emoji embedding and attention-based lstm. *Proceedings of the 26th ACM international conference on Multimedia*.

Keith Cortis and Brian Davis. 2020. Over a decade of social opinion mining. *Artificial Intelligence Review*, abs/2012.03091:4873–4965.

Mountaga Diallo, Chayma Fourati, and Hatem Haddad. 2021. Bambara language dataset for sentiment analysis. In *Practical ML for Developing Countries Workshop. ICLR 2021, Virtual Event*.

Ben Eisner, Tim Rocktäschel, Isabelle Augenstein, Matko Bosnjak, and Sebastian Riedel. 2016. emoji2vec: Learning emoji representations from their description. In *SocialNLP@EMNLP*.

Milagros Fernández Gavilanes, Jonathan Juncal-Martínez, Silvia García-Méndez, Enrique Costa-Montenegro, and Francisco Javier González-Castaño. 2018. Creating emoji lexica from unsupervised sentiment analysis of their descriptions. *Expert Syst. Appl.*, 103:74–91.

Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. *Processing*, 150.

Vandita Grover. 2021. Exploiting emojis in sentiment analysis: A survey. *Journal of The Institution of Engineers : Series B*, pages 1–14.

Gaël Guibon, Magalie Ochs, and Patrice Bellot. 2016. From Emojis to Sentiment Analysis. In *WACAI 2016*, Brest, France. Lab-STICC and ENIB and LITIS.

Fabian Haak. 2021. Emojis in lexicon-based sentiment analysis: Creating emoji sentiment lexicons from unlabeled corpora. In *Proceedings of the LWDA 2021 Workshops: FGWM, KDML, FGWI-BIA, and FGIR*, pages 279–286. CEUR-WS.

Shatha Ali A. Hakami, Robert Hendley, and Phillip Smith. 2021. Arabic emoji sentiment lexicon (Arab-ESL): A comparison between Arabic and European emoji sentiment lexicons. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 60–71, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

C.J. Hutto and Eric Gilbert. 2015. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*, pages 218–225.

Kanika Jindal and Rajni Aron. 2021. A systematic study of sentiment analysis for social media data. *Materials Today: Proceedings*.

Mohammed Kaity and Vimala Balakrishnan. 2020. Sentiment lexicons and non-English languages: A survey. *Knowledge and Information Systems*, 62:4445–4480.

Mayank Kejriwal, Qile Wang, Hongyu Li, and Lu Wang. 2021. An empirical study of emoji usage on twitter in linguistic and national contexts. *Online Social Networks and Media*, 24:100149.

Mayu Kimura and Marie Katsurai. 2017. Automatic construction of an emoji sentiment lexicon. *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, page 1033–1036.

Petra Kralj Novak, Jasmina Smailovic, Borut Sluban, and Igor Mozetic. 2015. Sentiment of emojis. *PloS one*, 10.

Janez Kranjc, Jasmina Smailovic, Vid Podpecan, Miha Grcar, Martin Znidari, and Nada Lavrac. 2015. Active learning for sentiment analysis on data streams: Methodology and workflow implementation in the clowdflows platform. *Inf. Process. Manag.*, 51:187–203.

Byung kwan Yoo and Julia Taylor Rayz. 2021. Understanding emojis for sentiment analysis. In *FLAIRS Conference*.

Chuchu Liu, Fan Fang, Xu Lin, Tie Cai, Xu Tan, Jianguo Liu, and Xin Lu. 2021. Improving sentiment analysis accuracy with emoji embedding. *Journal of Safety Science and Resilience*, 2:246–252.

Koena Ronny Mabokela, Turgay Celik, and Mpho Raborife. 2022a. Multilingual Sentiment Analysis for Under-Resourced Languages: A Systematic Review of the Landscape. *IEEE Access*, 11:15996 – 16020.

Koena Ronny Mabokela and Madimetja Jonas Manamela. 2013. An integrated language identification for code- switched speech using decoded-phonemes and support vector machine. In *2013 7th Conference on Speech Technology and Human - Computer Dialogue (SpeD)*, pages 1–6.

Koena Ronny Mabokela, Mpho Raborife, and Turgay Celik. 2022b. Safrisenti: Towards a creation of multilingual sentiment corpus for south african under-resourced languages. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*

Koena Ronny Mabokela and Tim Schlippe. 2022a. AI for Social Good: Sentiment Analysis to Detect Social Challenges in South Africa. In *The South African Conference for Artificial Intelligence Research (SACAIR 2022)*.

Koena Ronny Mabokela and Tim Schlippe. 2022b. A Sentiment Corpus for South African Under-Resourced Languages in a Multilingual Context. In

The 1st Annual Meeting of the ELRA/ISCA Special Interest Group on Under-Resourced Languages (SIGUL 2022), page 70–77.

Gati Lother Martin, Medard Edmund Mswahili, and Young-Seob Jeong. 2021. Sentiment Classification in Swahili Language Using Multilingual BERT. *African NLP Workshop, EACL 2021*, abs/2104.09006.

Salima Medhaffar, Fethi Bougares, Y. Estève, and Lamia Hadrich Belguith. 2017. Sentiment analysis of tunisian dialects: Linguistic ressources and experiments. In *WANLP@EACL*.

Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3):436–465.

Shamsuddeen Hassan Muhammad, David Ifeoluwa Adelani, Sebastian Ruder, Ibrahim Said Ahmad, Idris Abdulmumin, Bello Shehu Bello, Monojit Choudhury, Chris Chinenye Emezue, Saheed Salahudeen Abdullahi, Anuoluwapo Aremu, Alipio Jeorge, and Pavel Brazdil. 2022. NaijaSenti: A Nigerian Sentiment Corpus for Multilingual Sentiment Analysis.

Finn Årup Nielsen. 2011. A new ANEW: evaluation of a word list for sentiment analysis in microblogs. *CoRR*, abs/1103.2903.

Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *LREc*, pages 1320–1326.

D.J. Prinsloo. 2020. Lexicographic treatment of negation in sepedi paper dictionaries. *Lexikos*, 30:1–25.

Abhishek Singh, Eduardo Blanco, and Wei Jin. 2019. Incorporating emoji descriptions improves tweet classification. In *Proceedings of NAACL-HLT*, page 2096–2101.

Statista. 2022. African countries with the largest population as of 2020.

Hao Wang and Jorge A. Castanon. 2015. Sentiment expression via emoticons on social media. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 2404–2408.

Mayur Wankhade, Annavarapu Rao, and Chaitanya Kulkarni. 2022. A Survey on Sentiment Analysis Methods, Applications, and Challenges. *Artificial Intelligence Review*, pages 1–50.

A Appendix: Language Information

Northern Sotho, also known as Sesotho sa Leboa, is a Sotho-Tswana language primarily spoken in the northeastern regions of South Africa. It is also commonly referred to as Sepedi or Pedi. The South African National Census of 2011 reports that it is the first language of over 4.6 million people, accounting for 9.1% of the population, thus ranking it as the 5th most spoken language in South Africa. The Sepedi language is most frequently used in the Mpumalanga, Gauteng, and Limpopo provinces.

Tswana, known by its indigenous name **Setswana**, is a Bantu language spoken in Southern Africa by approximately 8.2 million individuals. It belongs to the Bantu language family within the Sotho-Tswana branch of Zone S, and shares close ties with the Northern Sotho, Southern Sotho, Kgalagadi, and Lozi languages. Setswana is an official language in Botswana and South Africa and serves as a lingua franca in Botswana and certain parts of South Africa, particularly in the North West Province. Tswana-speaking ethnic groups can be found across more than two provinces in South Africa, mainly in the North West, where approximately four million people speak the language.

Sesotho, also referred to as Southern Sotho, is a Southern Bantu language belonging to the Sotho-Tswana ("S.30") group. It is primarily spoken in Lesotho, where it serves as both the national and official language, as well as in South Africa (particularly in the Vaal and Free State), where it is one of the 11 official languages. It is also recognized as one of the 16 official languages of Zimbabwe. As with all Bantu languages, Sesotho is an agglutinative language that utilizes numerous affixes, and derivational and inflectional rules to construct complete words.