REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1S (2024)
http://www.veterinaria.org
Article Received: Revised: Accepted:

*w*CaWESA - weighted Capitalized Word Enhanced Sentiment Analysis Model

Ahongsangbam Dorendro[1]* and Haobam Mamata Devi[2]

[1,2] Department of Computer Science, Manipur University, Imphal West, India.
ah.dorendro@gmail.com,mamata_dh@rediffmail.com

***Corresponding Author:** Ahongsangbam Dorendro
*Email: ah.dorendro@gmail.com

Abstract
Sentiment analysis is a crucial task in natural language processing (NLP) that aims to determine the emotional tone of text. Traditional sentiment analysis models often overlook the impact of word formatting, particularly capitalization, on sentiment intensity. In informal communication, such as social media platforms, users frequently employ all-capitalized words to emphasize their emotions, signaling a heightened sentiment. This study introduces a novel model, weighted Capitalized Word Enhanced Sentiment Analysis (wCaWESA), which addresses this gap by explicitly accounting for the sentiment strength of capitalized words. The model assigns specific weights to capitalized words based on their frequency and context, enhancing the overall accuracy of sentiment analysis. By incorporating this additional layer of sentiment intensity, wCaWESA provides a more nuanced understanding of the emotional content in text, particularly in environments where informal and emphatic language is prevalent. The proposed model demonstrates significant improvements over existing approaches, making it a valuable tool for applications ranging from social media monitoring to opinion mining.

Keywords: Sentiment Analysis, Sentiment Strength, Capitalized Word, *w*CW, *w*CaWESA

1. Introduction
Sentiment analysis is the process of determining the emotional tone or opinion expressed in a piece of text. It is widely used in various applications, including marketing, customer service, and social media monitoring (Kumar & Vadlamani, 2015). One of the challenges of sentiment analysis is how to account for the intensity or strength of sentiment expressed in the individual word due to differences in the individuals informal writing style (Wankhade, Rao, & Kulkarni, 2022). One of the informal writing style found in social medias are the use of all capital letters word. A word expressed intentionally as a capitalized word emphasizes on the intensity of the sentiment expressed in the word (Irina & Phoey, 2018) (McCulloch, 2019). Traditional approaches of sentiment analysis involving the simple assignment of positive or negative labels to words without considering the capitalization of words fail to capture the nuanced intensity of sentiment expressed in the individual words. A more sophisticated approach that considers the intensity of the sentiment can be based on the intensity of sentiments expressed by each individual capitalized word which is being considered as an additional feature for sentiment analysis.

1.1 Problem Statement
Existing sentiment analysis models typically assign a static positive or negative label to words based on predefined lexicons or through machine learning-based approaches. These models often fail to capture the nuances of sentiment strength that can be inferred from the way words are presented in the text, particularly when capitalization is used to signify emphasis (Kiritchenko S. Z., 2014). This limitation is especially problematic in social media analytics, where the informal and expressive nature of the content demands a more sophisticated approach to sentiment detection.

1.2 Contribution
This study explores the concept of *w*CaWESA i.e. weighted Capitalized Word Enhanced Sentiment Analysis by accounting for the intensity or strength of sentiment carried by the individual word expressed as all capital letters word. This research contributes to the advancement of sentiment analysis methodologies, offering a model capable of deciphering nuanced sentiment strength in diverse textual contexts where the intensity of expression profoundly influences the overall meaning.

2. Background and motivation
Sentiment analysis has emerged as a critical tool in natural language processing, enabling the extraction of emotional and opinionated information from text data. This has applications in various domains, including business intelligence, customer feedback analysis, and political forecasting. The traditional sentiment analysis methods primarily focus on detecting polarity (positive, negative, or neutral) and often overlook the intensity or strength of sentiment conveyed by specific words or phrases (Pang, 2008). The strength or intensity of sentiment in a text can significantly alter its overall meaning. For example, in social media platforms where users often employ informal writing styles, the use of all capital

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1S (2024)
http://www.veterinaria.org
Article Received: Revised: Accepted:

letters (capitalization) is a common way to emphasize words, conveying stronger emotions or opinions (Potts, 2011). Research has shown that capitalization is not merely a stylistic choice but a deliberate attempt to amplify the sentiment expressed by a word or phrase (Lin & Wu, 2016). Ignoring this aspect can lead to an inaccurate sentiment analysis, as the true intensity of the sentiment may be underestimated or overlooked entirely (Thelwall, 2013).
Existing sentiment analysis models typically assign a static positive or negative label to words based on predefined lexicons or through machine learning-based approaches. These models often fail to capture the nuances of sentiment strength that can be inferred from the way words are presented in the text, particularly when capitalization is used to signify emphasis (Kiritchenko S. Z., 2014). This limitation is especially problematic in social media analytics, where the informal and expressive nature of the content demands a more sophisticated approach to sentiment detection.
The motivation behind this study is to address the gap in traditional sentiment analysis methodologies by developing a model that considers the strength of sentiment conveyed by capitalized words. The proposed model, wCaWESA (weighted Capitalized Word Enhanced Sentiment Analysis), aims to enhance sentiment analysis by incorporating a weighting mechanism for capitalized words. This approach recognizes the importance of capitalization as a marker of sentiment intensity and seeks to improve the accuracy of sentiment classification by adjusting the sentiment score based on the presence of capitalized words.

The introduction of wCaWESA represents a novel contribution to the field of sentiment analysis, providing a more nuanced understanding of text data where sentiment intensity plays a crucial role. By assigning appropriate weights to capitalized words, the model can better capture the emotional depth and strength of opinions expressed in the text, leading to more accurate sentiment predictions (Liu B. , 2012).

3. Related Works

Hutto and Gilbert introduced VADER, a lexicon and rule-based sentiment analysis tool specifically attuned to sentiments expressed in social media contexts. VADER accounts for sentiment intensity by considering factors such as capitalization, punctuation, degree modifiers, and emoticons (Hutto & Gilbert, 2014). The model assigns sentiment scores to words and adjusts these scores based on contextual cues to reflect intensity more accurately. While VADER considers capitalization as an emphasis mechanism, it does not assign specific weights to fully capitalized words beyond simple amplification. The weighting mechanism in wCaWESA aims to provide a more nuanced and systematic approach to quantify sentiment strength associated with capitalized words, potentially leading to more precise sentiment intensity.
(Kim, 2014) pioneered the use of CNNs for text classification, which became a foundation for various sentiment analysis models. These models leveraged fixed word embeddings such as Word2Vec and GloVe to predict sentiment. Similarly, RNN-based models, including LSTMs (Long Short-Term Memory networks), were later introduced to capture sequential information in text. While these models achieved high accuracy in sentiment polarity detection, they often failed to account for sentiment strength, particularly when it comes to informal writing styles like capitalization. CNNs and RNNs did not inherently capture the emphasis conveyed by capitalized words. Though effective at detecting general sentiment, they ignored the intensity or emphasis of sentiments, especially in contexts like social media, where the use of capital letters is prominent.

Lin and Wu explored the detection of emphasis in text by identifying linguistic markers such as exclamation marks, repetitions, and capitalized words. Their model, while recognizing the importance of capitalization, focused on general emphasis detection rather than sentiment analysis (Lin & Wu, 2016). Lin and Wu's approach detected emphasis in a broad sense but did not specifically translate emphasis into quantifiable sentiment strength.

Attention mechanisms, introduced in transformer models like BERT, were later adapted to emphasize sentiment-related words. (Kiritchenko S. Z., 2018) introduced an attention-based model to detect sentiment intensity in tweets. Their approach used a weighted attention mechanism to focus on sentiment-laden words. However, this model did not incorporate capitalization as an explicit feature, thus failing to recognize the increased intensity associated with capitalized words.

In 2018, Wang et al. proposed a CNN-based model that incorporates attention mechanisms to capture sentiment intensity by focusing on significant words and phrases within the text (Wang, Huang, & Zhao, 2018).
Variants of BERT, including RoBERTa (Liu Y. O., 2019) and DistilBERT (Victor Sanh, 2019), further improved upon the model's efficiency and accuracy in downstream tasks like sentiment analysis. Despite the success of BERT and its variants in capturing semantic nuances, they too overlooked the importance of capitalization as an indicator of sentiment intensity. While contextual embeddings allowed BERT to better understand sentiment based on word context, they did not explicitly emphasize capitalized words, leading to potential underestimation of sentiment strength.
Deep learning models often require large amounts of labeled data and substantial computational resources. They may also struggle to explicitly account for stylistic features like capitalization unless specifically engineered to do so.

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1S (2024)
http://www.veterinaria.org
Article Received: Revised: Accepted:

wCaWESA, with its emphasis on weighting capitalized words, provides a more direct and interpretable method for capturing sentiment intensity, especially in resource-constrained settings.

4. *w*CaWESA - weighted Capitalized Word Enhanced Sentiment Analysis model

w*CaWESA - weighted Capitalized Word Enhanced Sentiment Analysis model** is a supervised Sentiment Analysis model that assigns weights to individual words based on their emotional intensity expressed as capitalized word. Understanding that all-capital words contribute to sentiment expression, the ***w*CaWESA** model consists of two components. Firstly, the weightage for Capitalized Word (*w*CW) parameter assigns weightage to words presented in all capital letters, recognizing the potential for heightened sentiment intensity and the feature strength (f***) of the word being considered in the given text. These weights are used to calculate an overall sentiment score for the text. The sentiment strength or intensity of the individual capitalized word will be derived word will be derived from using *w*CW and ***f***.
For example, same word "HAPPY" and "happy" expressed as all capital letter and all small letter word can have different meanings and sentiment strength in the sense that "HAPPY" conveys a stronger feeling of happy than "happy".
sentiment strength or intensity of HAPPY = *w*CW +***f***
sentiment strength or intensity of happy = ***f***
The final overall sentiment score of the given text is either 1 for positive and 0 for negative derived by aggregating the sentiment strength or intensity of all the individual words being considered as features for sentiment analysis.

4.1 weightage for Capitalized Word (*w*CW)

Definition: It is the sentiment strength or intensity assigned to a word expressed as an all capital letter word i.e. capitalized word. *w*CW is defined as the term frequency of a capitalized word in a document normalized over the distribution of the capitalized word in the entire corpus i.e. in all the documents containing the capitalized word.

4.2 Calculation of weightage for Capitalized Word (*w*CW)

- **Capitalized Words, WF_{cap}**: This focuses only on capitalized words. If w is a capitalized word, $WF_{cap}(w,d)$ is the frequency of w in d, normalized by the total number of words in document d, then

$$WF_{cap}(w,d) = \frac{\text{Count of capitalized w in document } d}{\text{Total number of capitalized words in document } d}$$

  This provides the normalized frequency of the capitalized word within a single document.

- **Adjusted DF for capitalized words**: The Adjusted Document Frequency (DFcap) is the proportion of documents that contain the capitalized word w.
  If, $DF_{cap}(w)$ represents the number of documents containing w,
  $D = \{d_1, d_2, d_3, \ldots d_n\}$ then

$$DF_{cap}(w, D) = log(\frac{|D|}{1 + DF_{cap}(w)})$$

Where $DF_{cap}(w)$ is the number of documents containing the capitalized word, w.

Now, $wCW = WF_{cap}(w, d) \times DF_{cap}(w, d)$

This product reflects the importance of the capitalized word w based on its frequency within a document and its occurrence across the entire corpus.

- **Example Illustration**

Let us consider 4 documents as given below:
d_1 = I LIKE cricket VERY MUCH
d_2 = I LOVE cricket with all my heart
d_3 = I LIKE apple very much.
d_4 = I am EXTREMELY HAPPY as I got the job.
Now,
$D = \{d_1, d_2, d_3, d_4\}$
$|D| = 4$

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1S (2024)
http://www.veterinaria.org
Article Received: Revised: Accepted:

$$WF_{cap}(\text{LIKE}, d_1) = \frac{\text{Count of capitalized "LIKE" in document } d_1}{\text{Total number of capitalized words in document } d_1} = \frac{1}{3}$$

$$DF_{cap}(LIKE, d_1) = log(\frac{|D|}{1 + DF_{cap}(w)}) = log(\frac{4}{2}) = 0.3$$

Then, $wCW_{LIKE} = WF_{cap}(w, d) \text{ x } DF_{cap}(w, d) = \frac{1}{3} \text{ x } 0.3 = 0.1$

Thus, the wCW for the capitalized word "LIKE" in d_1 is 0.1. This weightage is assigned to the capitalized word "LIKE" in addition to the normal feature weight determined by the feature extraction used in the sentiment analysis model.
i.e. sentiment strength of (LIKE, d_1) = feature weight, f_{LIKE} + wCW $_{LIKE}$
= feature weight, f_{LIKE} + 0.1

4.3 Implementation of wCaWESA model
wCaWESA model is implemented with different machine learning models viz. Naïve Bayes (NB) Classifier, Support Vector Machine(SVM), Logistic Regression with additional feature of *w*CW. *w*CW for each capitalized words in all the documents is calculated as described in the preceding section and a vector representation for the capitalized words is generated called wCWVectorizer.

In this implementation, we combine the outputs of text representation model/feature vector of different classifier and wCWVectorizer to enhance the representation of capitalized words in text data. For example, if NB classifier is using CountVectorizer for its bag of word representation of text data, then CountVectorizer captures the frequency of each word in a document, while the wCWVectorizer provides a measure of the importance or sentiment intensity of each capitalized word. The core idea is to enrich the text representation model with additional wCW values specifically for capitalized words, which are often used to convey emphasis or special significance.

4.3.1 Example Illustration
Let us consider a Classifier using CountVectorizer for its bag of word representation of text data.
The process begins by generating separate feature matrices for both CountVectorizer and wCWVectorizer. Next, a mapping is created between the wCW features (focusing on capitalized words) and their corresponding lowercase counterparts in the CountVectorizer matrix. For each capitalized word, the wCW value is added to the count value of the corresponding lowercase word in the CountVectorizer matrix. This combined matrix integrates the raw frequency counts with the enhanced wCW values, resulting in a more nuanced feature representation that captures both the occurrence and significance of words, particularly those with capitalization. Finally, the combined matrix is used with a classifier, ensuring that the final feature set reflects the importance of capitalized terms alongside their raw counts.

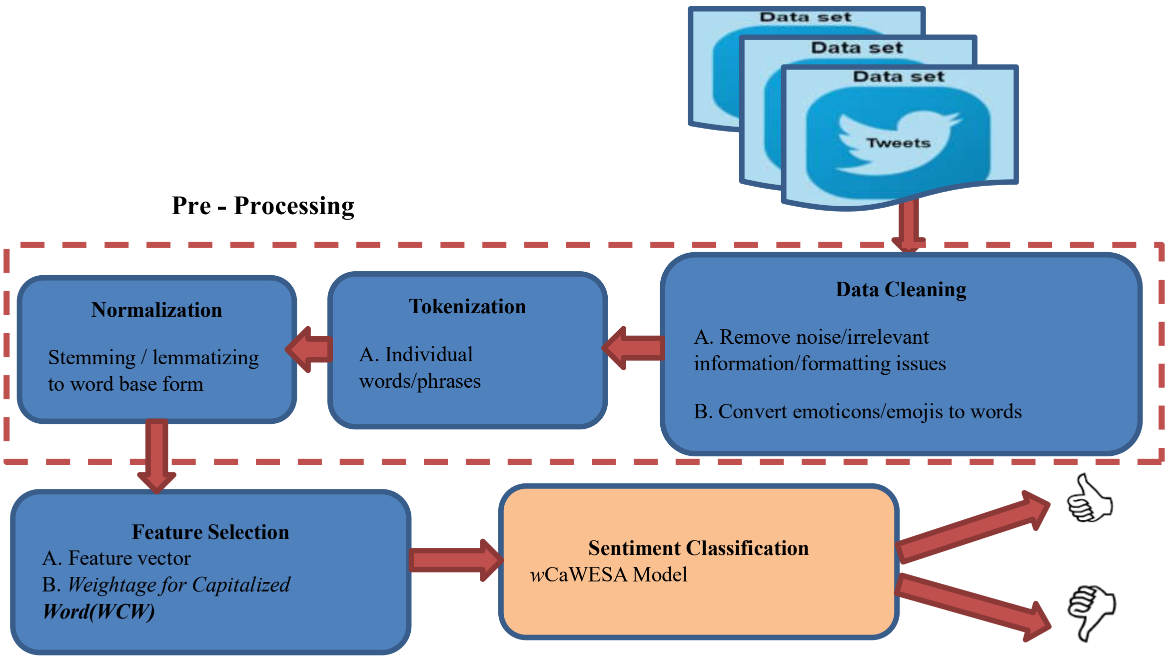

Fig 1: A high level diagram of the *w*CAWESA model

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1S (2024)
http://www.veterinaria.org
Article Received: Revised: Accepted:

5. Evaluation of the wCaWESA model:

Three wCaWESA models with NB classifier, Support Vector Machine (SVM) and Logistic regression are implemented that incorporates the wCW feature vector represented by wCWVector Matrix. The performance of these 3 models is evaluated against normal implementation of NB classifier, SVM and Logistic Regression with different feature selection viz. bag of Words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF) and Word2Vec word embedding.

5.1 Dataset:

Tweets for the Indian Loksabha Election 2019 were collected during the period 1/3/2019 to 30/4/2019. These tweets have been retrieved using various hashtags related to the Indian Loksabha Election 2019. These hashtags were trending during the events leading to Indian Loksabha Election 2019. Some of the hashtags used to collect the tweets related to the Indian election are

Table 1. Hashtags for tweet collection

#GoBackRahul	#VanakkamRahulGandhi	#LokSabhaElections2019
#GoBackPappu	#Congress(INCIndia)	#LokSabhaElections
#MyFirstVoteForModi	#ModiModiBhaiBhai	#Elections2019
#MissionModi2019	#ChowkidarChorHai	#IndianGeneralElections
#BJP(BJP4India)	#RahulTakingIndiaForward	#BattleOf2019
#MainBhiChowkidar	#RahulDemocracyDialogue	#MyVoteForIndia
#ChowkidarNarendraModi	#MainBhiBerozgar	
#ChowkidarPhirSe	#CongManifesto	
#NamoAgain	#INCIndia	
#WeWantChowkidar	#RahulGandhiWayanad	
#PappuDiwas	#AbHogaNYAY	
#PhirEkBaarModiSarkar	#AmethiKaRahulGandhi	
#IsBaarNaMoPhirSe	#Scared2Debate	

When it comes to political deliberation on social media platforms, the views and opinions expressed cannot be taken as it is. (Dorendro & Devi, 2024) has extensively discussed the challenges associated with the opinions and deliberation expressed over social media platform. In order to filter out likely manipulated opinions and to get a representation of general population, only tweets from accounts more than 1 month old at the time of tweet retrieval and having follower less than 20,000 at the time of tweet retrieval are considered for sentiment analysis. This study assumes that general population have on average less than 20,000 followers and tweets from newly created twitter accounts(less than 1 month) are less reliable.

5.2 Results

After the retrieving tweets from twitter (now x), duplicate tweets are removed and tweet filter mentioned in section 5.1 is applied. The remaining tweets are processed for the sentiment analysis. These tweets are annotated as 0 and 1 for binary sentiment classification. The dataset is divided into 80% training set and 20% test set. Three machine learning models namely Naïve Bayes (NB), Support Vector Machine (SVM) and Logistic regression are trained with different feature selection viz. bag of words (Bow), term frequency-inverse document frequency (tf-idf), word2vector (word2vec) and wCaWESA model purposed in this study. The evaluation result for the different classifier models are given below.

Table 2: Evaluation results of different classifier models

Feature selection	Naïve Bayes				SVM				Logistic regression			
	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
BoW	82.35	88.89	76.92	78.75	73.53	71.97	71.25	71.53	79.41	79.58	76	77
tf-idf	82.35	88.89	76.92	78.75	79.41	82.21	74.54	75.88	76.47	86.21	69.23	69.78
word2vec	52.94	60.87	58.97	52.28	64.7	62.06	61.17	61.36	67.65	65.86	62.09	62.11
wCaWESA	**88.24**	**88.93**	**86.08**	**87.12**	**82.35**	**84.44**	**78.39**	**79.84**	**79.41**	78.4	**77.47**	**77.86**

The results above show the performance of different feature selection techniques (BoW, TF-IDF, Word2Vec, and wCaWESA) across three classification models viz. Naïve Bayes, SVM, and Logistic Regression, evaluated using accuracy, precision, recall, and F1 score.

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1S (2024)
http://www.veterinaria.org
Article Received: Revised: Accepted:

Bag of Words (BoW) - Naïve Bayes achieves a relatively high accuracy (82.35%) with strong precision (88.89%) but slightly lower recall (76.92%). The F1 score (78.75) suggests that this model has a bias toward precision, meaning it predicts positive cases accurately but may miss some positive instances. For SVM, accuracy drops to 73.53%, with a balance between precision and recall. While both are around 71%, the lower F1 score (71.53) indicates that the model performs less well than Naïve Bayes using BoW. Logistic Regression, accuracy (79.41%) improves compared to SVM but is still lower than Naïve Bayes. The precision (79.58%) and recall (76%) suggest the model is more balanced, though the overall F1 score (77) indicates performance similar to Naïve Bayes but not superior.

TF-IDF - Naïve Bayes model performs similarly to BoW, with the same accuracy (82.35%) and precision (88.89%). The recall and F1 scores are also consistent, showing that Naïve Bayes benefits equally from both BoW and TF-IDF in terms of performance. SVM sees a significant improvement with TF-IDF, increasing accuracy to 79.41%. Precision (82.21%) also increases, and while recall (74.54%) is still lower, the F1 score (75.88) suggests the model is making more balanced predictions. Interestingly, Logistic Regression sees a drop in performance with TF-IDF compared to BoW. Accuracy (76.47%) decreases slightly, and the F1 score (69.78) reflects a larger imbalance between precision (86.21%) and recall (69.23%).

Word2Vec - Across all three models, Word2Vec shows significantly lower performance compared to BoW and TF-IDF. For Naïve Bayes, accuracy (52.94%) and F1 score (52.28) are low, indicating that Word2Vec features are not well-suited for this model. The precision (60.87%) is higher than recall (58.97%), but the model struggles overall. SVM with Word2Vec performs better here, with accuracy (64.7%) and a more balanced precision-recall distribution. However, the F1 score (61.36) still reflects mediocre performance. Logistic Regression model has similar accuracy (67.65%) and precision-recall levels as SVM, but the low F1 score (62.11) again points to weak performance when using Word2Vec.
wCaWESA - Naïve Bayes with wCaWESA model shines here, with the highest accuracy (88.24%) and balanced precision (88.93%) and recall (86.08%). The F1 score (87.12) indicates that the model performs very well in predicting both positive and negative instances. SVM with wCaWESA also performs strongly with SVM, reaching an accuracy of 82.35%, the highest for this model across all features. Precision (84.44%) and recall (78.39%) are close, producing a solid F1 score (79.84). Logistic Regression with wCaWESA model performs reasonably well, with accuracy (79.41%) comparable to BoW and TF-IDF, but slightly lower than its performance with Naïve Bayes and SVM. Precision (78.4%) and recall (77.47%) are balanced, producing a good F1 score (77.86).

5.3 Discussion

- Naïve Bayes consistently performs best when using the wCaWESA feature selection method, significantly improving accuracy, precision, recall, and F1 score compared to traditional methods like BoW and TF-IDF. This suggests that the emphasis on capitalized words in wCaWESA provides a useful signal for Naïve Bayes, likely improving its ability to distinguish between positive and negative sentiments.
- SVM shows marked improvement with wCaWESA compared to other features. While it performs well with TF-IDF, the higher recall and F1 scores with wCaWESA suggest that this method enhances SVM's ability to capture nuanced sentiment expressed through capitalization, making it more effective for sentiment analysis.
- Logistic Regression shows consistent performance across all features, but it doesn’t seem to fully exploit the advantages of wCaWESA like Naïve Bayes or SVM. Nonetheless, the balanced precision and recall scores suggest that Logistic Regression can still be a reliable model when using wCaWESA.
- Word2Vec underperforms across all models. This is likely because Word2Vec focuses on word embeddings and semantic meanings, which are not necessarily aligned with the signal provided by capitalized words in sentiment analysis tasks.

Overall, the results demonstrate that wCaWESA is the most effective feature selection technique, significantly boosting performance, especially for models like Naïve Bayes and SVM, which benefit from the emphasis on capitalized words. This confirms that the wCaWESA method effectively captures nuances in sentiment often missed by machine learning feature selection methods.

6. Conclusion

The study finds that wCaWESA generally enhances the performance of machine learning models like Naïve Bayes, SVM, logistic regression by making them more sensitive to sentiment cues conveyed through capitalization. These findings suggest that while capitalized words do add value in sentiment analysis, the extent of this value can vary depending on the underlying model architecture and its sensitivity to textual nuances.

Weighted sentiment analysis is a more sophisticated approach to sentiment analysis that takes into account the intensity of sentiment expressed in an individual word. The approach assigns weights to words based on their emotional intensity

REDVET - Revista electrónica de Veterinaria - ISSN 1695-7504
Vol 25, No. 1S (2024)
http://www.veterinaria.org
Article Received: Revised: Accepted:

and calculates an overall sentiment score for the text. This approach can also be adapted for repetition of letters to emphasis sentiment strength in textual data. The idea is presented a possible future work.

Weighted sentiment analysis has many applications in social media monitoring, customer service, and political analysis. As data continues to grow and become more complex, the need for sophisticated sentiment analysis techniques will continue to increase. Weighted sentiment analysis is an essential tool for businesses and organizations that want to understand the sentiment of their customers and stakeholders. This research contributes to the advancement of sentiment analysis methodologies, offering a model capable of deciphering nuanced sentiment strength in diverse textual contexts where the intensity of expression profoundly influences the overall meaning. Further refinement and evaluation of the wCaWESA model present opportunities for its application in real-world scenarios, ensuring a more accurate and comprehensive understanding of sentiment in textual data.

References

1. Dorendro, A., & Devi, H. (2024). Challenges in Determining the Authenticity, Honesty, and Intentions of Opinions Expressed on Twitter and Sentiment Analysis. *Journal of Electrical Systems, 20*(10), 1627-1631.
2. Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Proceedings of the Eighth International Conference on Weblogs and Social Media (ICWSM-14).*
3. Irina, P., & Phoey, L. T. (2018). Value of Expressions behind the Letter Capitalization in Product Reviews. *7th International Conference on Software and Computer Applications.* New york.
4. Kim, Y. (2014). Convolutional neural networks for sentence classification. *In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1746-1751.
5. Kiritchenko, S. Z. (2014). Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research, 50*, 723-762.
6. Kiritchenko, S. Z. (2018). Sentiment analysis of tweets: The role of capital letters. *Journal of Artificial Intelligence Research, 62*, 493-514.
7. Kumar, R., & Vadlamani, R. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems, 89*, 14-46.
8. Lin, T. &. (2016). Emphasis Detection for Expressive Text. *In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* , 1532-1541.
9. Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies, 5*(1), 1-167.
10. Liu, Y. O. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv preprint arXiv:1907.11692.
11. McCulloch, G. (2019, July 23). *The Meaning of All Caps—in Texting and in Life*. (Wired) Retrieved from https://www.wired.com/story/all-caps-because-internet-gretchen-mcculloch/
12. Pang, B. &. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval, 2*((1-2)), 1-135.
13. Potts, C. (2011). On the negativity of negation. *In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 1247-1252.
14. Thelwall, M. &. (2013). Topic-based sentiment analysis for the social web: The role of mood and issue-related words. *Journal of the Association for Information Science and Technology, 64*(8), 1608-1620.
15. Victor Sanh, L. D. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *https://arxiv.org/abs/1910.01108*.
16. Wang, Y., Huang, M., & Zhao, L. (2018). Attention-based LSTM for Aspect-level Sentiment Classification. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 606-615.
17. Wankhade, M., Rao, A., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artif Intell Rev, 55*, 5731–5780.