Using Machine Learning to Analyze Sudanese Opinions for Political Decision Making

Fatima Salih Ibrahim, Eltyeb Elsamani, and Shazali Siddig

Abstract—**Sentiment Analysis (SA) is the application of text mining techniques for extraction and identification of subjective opinions from textual data. SA has many applications ranging from financial analysis to political decision-making domains. Although a lot of works have been made in SA of English tweets, very limited work focused on Arabic political colloquial tweets. In this paper, we focused on analyzing tweets during Sudanese revolution 2018 and leverage Term Frequency-Inverse Document Frequency and word embedding methods with machine learning classifiers to detect user's sentiments of colloquial Arabic political tweets instead of using modern standard Arabic. Experiments were conducted to evaluate our classifiers and the results showed superiority of the Ensemble learning model with an F-score of 83% compared to other machine learning classifiers.**

Index Terms—**Sudanese revolution, Arabic tweets, Sentiment Analysis, machine learning, word embedding.**

I. Introduction

The development of Internet technologies led to a significant increase in the number of users on social networks or social media platform, which in turn generated a huge amount of data such as unstructured texts in the form of conversations, messages and blogs. Besides the task of sharing information, social media has become an effective and convenient way to express people's opinions and ideas, which gain popularity when liked by a large set of users. This popularity may reflect the people' opinions toward that organization, person, or a place. The social networks, such as twitter, creates massive amounts of the text that contain political insights, which can be extracted to analyze the people's sentiments and predict the future trends in the election [1]-[3]. Similarly, Interest in political decision-making procedures has grown significantly over recent decades. Some recent studies have recommended that it is important to expand the study of public opinion from policy outcomes to decision-making processes, and that there are coherent patterns in citizens' expectations of the way in which political decisions are made [4]. It becomes a common practice for public users to post their expressions toward the political leader on social media and used SA techniques to understand their behaviors. [5], [6]. Although different reporters have been taking an interview with the political leader to know their opinions. However, It is very expensive and time-consuming to search people's opinions about specific political talk via surveys and polls for detecting their opinions [7]. SA is automated way of analyzing tweets, capture the political sentiments and predict the future trends in the political events such as election [1], [3], [8]-[10]. Sentiment analysis/opinion mining is considered as one of the main Natural language processing tasks [11] and is helpful for analyzing posts from various social media platforms like twitter, Reddit, and Facebook [12] to detect people feelings. SA is useful in judging the overall views regarding an anew released item, movie, song, book, etc., and it can also distinguish between positive, negative and neutral opinions from a text [13].

The research in SA area mainly focused on analyzing English language text to detect people's feelings, but there are only few studies on colloquial Arabic which spoken by more than millions of people, and is the fastest-growing language on the web [14]-[16]. SA in Arabic social media is challenging task because Arabic language is not a case-sensitive language and has no capital letters. In addition, Arabic is a high inflectional language, and often a single word has more than one affix [17]. Other challenges of analyzing Arabic text are that every country/ part of a county has its own version or dialect of Arabic. That means there are different dialects of Arabic text available online that could hold different meaning and resulting high complexities when analyzing sentiments with different dialects. Further, the root for Arabic words could have multiple forms based on the context such as (يتكلم, كلمات, كلام). The aim of this paper is to analyze Arabic political tweets collected during Sudanese revolution in 2018 and leverage machine learning classifiers to get their classification performance and determine which one is better in SA of Arabic colloquial tweets. To achieve this goal, we proposed a method consists of four steps namely data collection and labeling, data preprocessing, feature extraction, and classification step. In our method, we used two sets of features, including Term Frequency-Inverse Document Frequency (TF-IDF) and word embedding with five machine learning classifiers, including K-Nearest Neighbors algorithm, Random Forest, Support Vector Machine (SVM), Multinomial Naïve Bayes (NB) and Ensemble model. TF-IDF is a traditional feature extraction method and can represent the text statistically to get the feature vector while word embedding is a recent representation technique and it is useful in capturing the syntactic and semantic information of tweets.

The rest of the paper is structured as follows: Section II, we review the related works. In Section III, we describe the methodology for analyzing Arabic tweets and identify the users' sentiments. Section IV, presents the experimental settings and discusses the experimental results. Finally, we conclude the paper and highlight the future work in Section V.

Manuscript received November 7, 2020; revised March 20, 2021.

The authors are with the Department Computer Science, Shaqra University, Saudi Arabia (Corresponding author: Fatima Salih Ibrahim; e-mail: Fatimaasalih@su.edu.sa).

doi: 10.18178/ijmlc.2022.12.3.1085

II. Related Work

In this section, we review the existing approaches relating to SA and present some literature related to most important studies that have been conducted in Arabic opinion mining. We can say that most of the works on opinion mining focused on English language due to the absence of free lexical resources and the complexity of the automatic analysis of Arabic language.

In early 2000s, Pang *et al*. [18] stared SA of movies' reviews. Since then the researchers investigated different topics and fields in the area of SA using different text data such as social media content.

Machine learning classifiers have been used to analyze the text data. For examples, Elghazaly *et al*. [8] developed online political SA to predict the election' result by focusing on Egypt's presidential elections. The authors used TF-IDF to obtain a document vector with two algorithms. They compared between SVM and naïve bays techniques for Arabic text classification. They concluded that the Naïve Bayesian method gives the highest accuracy compare to other classifiers. Mesleh *et al.* [19] proposed a system based on SVM and by using CHI statistics as a feature extraction method for Arabic language text classification. Their dataset was collected from Arabic newspaper archives and some other websites and contained 1445 examples classified into nine categories. Their results indicate a high F-score compared to other classification algorithms [19]. Similarly, Rushdi-Saleh *et al*. [20] used SVM and NB methods for classifying movie reviews collected from different web pages. They applied different preprocessing techniques such as spelling correction, stop words removal, and stemming to clean the text. Their experimental results showed that using SVM yields the best classification accuracy of 90%. Further, in [21] the authors collected and analyzed Arabic comments from Yahoo social network. They collected 4625 Arabic reviews and comments, which contains the topic, comments, manual polarity, gender of the users. They applied SVM and Naïve Bayes classifiers on this dataset and compared between them. The result showed that the best accuracy achieved by using the SVM.

Furthermore, Elhawary *et al.* [22] focused on the SA of twitter using common machine learning techniques such as a Naïve Bayes and a Maximum Entropy Model for exploring and understanding people, their lives, potential, interests, and opinions.

Recently, several studies used word representations as features for common NLP tasks. For example, Zirikly *et al.* [23] explored the impact of using word embeddings in detecting named entities in Arabic text. In their work, the authors demonstrated that this novel representation scheme can replace the use of dictionaries and gave better performance although the authors used a small twitter corpus of tweets. The term "word embeddings" was first presented in [24] and The model was based on the idea of obtaining the values for word vectors or embeddings by training a neural language model. In [25], the authors presented their own SA dataset of sentiments on health care domain. They built a SA method and their experiments included the use of many ML algorithms. Further, [26] Dashtipour *et al*. developed a hybrid SA system by combing dependency grammar-based rules and deep neural networks in Persian text data.

However, most of the work is focused on English whereas colloquial Arabic did not receive much attention until recently, but it still lacks behind due to the many challenges of the Arabic language as mentioned in the previous section.

III. Proposed Methodology

In recent years, the widespread application of machine learning algorithms has been successfully utilized in opinion mining. In this paper, to analyze Arabic opinions (Sudanese political tweets), we present our proposed methodology that automatically analyzes opinions from tweets for political decision-making. The opinion of Sudanese's revolution content in the social media was chosen as a case study. To handle our problem, we proposed a method that consists of four steps namely data collection and labeling, data preprocessing, feature extraction and data classification step. In the classification step, we adopt widely used supervised machine learning classifiers such as SVM, multinomial Naïve Bayes, K-Nearest Neighbors (KNN), Decision Tree (DT) and ensemble learning model. The followings are the detailed description of the steps:

A. Data Collection and Labeling

The first step in the SA process is collecting tweets by specifying a keyword to retrieve all tweets that are related to our keyword. Tweets can be collected from different sources. In our work, we used twitter's API to collect Arabic tweets related to Sudanese revolution that was started on December 19, 2018. The data was collected from hash tags related to the Sudanese revolution, such as # تسقط_بس # مدن السودان تنتفض #اعتصام القيادة)). The collected dataset is highly imbalanced: 1822 positive, 243 negative and 420 neutral tweets as shown in Table I.

TABLE I: Distribution of Collected Tweets

Class Name	No of Tweets
Neutral	420
Negative (Military)	243
Positive (Civilian)	1822
Total	**2485**

TABLE II: Sample of Labeled Tweets

Tweet	In English	Class
مدن السودان تنتفض	Sudan cities revolt	Neutral
القيادة العسكرية قررت بحتمية التغيير	The military leadership decided that change was imperative	Positive
كلهم ما نافعين	They are all not useful	Negative

It is noticeable that the positive (civilian) category is much higher compared to the negative (military) and neutral categories, which clearly reflects the majority of people opinions may be against the Sudanese regime. This is expected due to the poor economic situation, and the people's desire for democratic civil rule.

The data labeling process is conducted manually by three annotators (graduate students with Computer Science background), and we gave them clear instructions to label the dataset into positive (we call it civilian), negative (military) and neutral classes as shown in Table I. Table II shows an example of labeling process. The first column shows Arabic tweet sentence, in the second column is the English translation of the tweet, and lastly is the class category.

B. Data Preprocessing

The aim of the data preprocessing step in this paper is to prepare the data in a form that can be analyzed efficiently by using machine learning classifiers. In addition, to improve the data quality by deleting irrelevant tweets. We followed [27], [28] and cleaned the data by removing irrelevant information, such as URLs, special characters, duplicated characters, and non-Arabic letters. Further, we remove stop words such as prepositions and articles that occur frequently and may not be useful in discrimination between data classes/labels. Finally, we reduce words to their stem or root from where morphological information is used to match different variants of words.

TABLE III: EXAMPLE OF PREPROCESSING OF TWEET

Preprocessing Step	Tweets After preprocessing
The original tweet	نطلع الشاااارع يا حكومة مدنية يا ثورة أبدية!؟
Data cleaning	نطلع الشاااارع يا حكومة مدنية يا ثورة أبدية
Removing Duplicated Characters	نطلع الشارع يا حكومة مدنية يا ثورة أبدية
Tokenization	نطلع‘ الشارع، يا‘ حكومة‘ مدنية‘ يا‘ ثورة ‘أبدية
Stopword Removal	نطلع الشارع حكومة مدنية ثورة أبدية
Stemming	طلع شرع حكم مدن ثور أبد
Normalization	طلع شرع حكم مدن ثور ابد

C. Feature Extraction

After cleaning and preprocessing steps, we transform our text data into numerical representation, which would be used as input to the machine learning algorithms. We use two feature extraction methods namely Term Frequency-Inverse Document Frequency (TF-IDF) and Word2Vec [29]. We transform the pre-processed opinions into a feature vector by using Word2Vec and TF-IDF techniques. Word2Vec is the word embeddings technique, which is widely used for converting words to feature vectors. Whereas, the term TF-IDF is a numerical statistic based on weighting metric to describe how the words in sentences are significant in the data.

Since word embeddings work on the word level and the classification problem works on sentence or document level, we propose to use two approaches for transforming word embeddings to sentence embedding. The first is by averaging the embeddings of all the words in a sentence which we denote by Mean Embeddings. The second approach uses the TF-IDF of each word as a weight when averaging the word embeddings which we denote by TF-IDF.

D. Classification Models

In this paper, our methodology employs five different type of classification techniques which are KNN, Random Forest, SVM, Multinomial Naïve Bayes (NB) and Ensemble learning voting. Ensemble methods seek to combine the prediction of multiple classifiers to obtain a classification model with better predictive performance. Combining machine learning classifiers is helpful to reduce the variance and bias of classification. An ensemble model works by creating multiple models and then combine them to produce improved results and usually produces more accurate solutions than a single machine learning model. In this paper, we used KNN and SVM as base models (inputs of ensemble method) and chose voting as ensemble method for classification. Every model makes a prediction for each test example and the final output prediction is the combination of the base models.

IV. EXPERIMENTAL SETTING AND RESULT ANALYSIS

In this section, we empirically investigate the performance of the methodology on our dataset and reported the results for political opinion mining on Arabic text data.

A. Experimental Settings and Evaluation Metrics

To validate the classification performance of our methodology, we performed experiments based on machine learning classifiers. We evaluate the performance of each classifier based on 10-fold cross validation. In our experiments, we divided the dataset into 70% training and 30% for testing. We performed a grid search to select best hyperparameters for evaluation. Once the best settings were found, the final classifiers were learned on the training sets and tested on the test set. Furthermore, Google Word2Vec and Sklearn open source libraries were used to implement the proposed methodology.

We employed the confusion matrix to visualize the algorithms' performance. The components of the confusion matrix can be used for finding the values of some evaluation metrics such as precision, recall, and F-score as formalized in the following:

Precision is the ratio of total number of the examples correctly labeled as positive to the total number of positively classified example.

$$\text{Precision} = \text{TP}/(\text{TP}+\text{FP}) \times 100 \quad (1)$$

Recall is the ratio of total number of the positive labeled examples to the total examples that are truly positive.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \times 100 \quad (2)$$

F-score is described as the harmonic mean or the weighted average for both the precision and recall obtained.

$$\text{F} - \text{score} = 2 \times \frac{\text{RcallxPrecision}}{(\text{Rcall+Precision})} \quad (3)$$

where, TP (true positive) is described as the examples which are positive and have been classified as positive by the classifiers, FN (false negative) represents positive examples, but classifier classifies them as negative. True negative (TN) represents the examples that are negative and classified as negative examples by the classifiers, FP (false positive) represents negative examples, but classifier classifies them as positive.

B. Experimental Result and Analysis

As mentioned earlier, the proposed methodology employs TF-IDF and Word2Vec to generate features. Consequently, Table IV depicts the Word2Vec results and Table V shows the TF-IDF results in terms of precision (P), recall (R) and F-score. From Table IV, we observe that, based on Word2Vec the ensemble learning model can accurately classify opinions with at least 83. % F-score. Whereas, based on TF-IDF (see Table V) the ensemble approach can accurately classify opinions with at least 76. % on F-score. Furthermore, Word2Vec outperforms TF-IDF with an average improvement of 6 %, 5 %, and 7% in terms of precision, recall, F-score respectively. We further note that, SVM outperforms all other classifiers in both TF-IDF and Word2Vec. KNN and random forest are other algorithms, which have shown to be effective as well in the classification. However, in this experiment, multinomial NB algorithm has achieved poor results comparing against other classifiers. It has achieved a F-score of 40% and 61% on TF-IDF and Word2Vec respectively. Since the reported results suggest that, Word2Vec outperforms TF-IDF, we thus report our further analysis based on Word2Vec.

Moreover, as presented in Fig. 1, the SVM results show that positive class has highest precision, recall, and F-score compare to other classes' results. This due to imbalance of our dataset as presented in Table I.

TABLE IV: CLASSIFIERS PERFORMANCE MEASURED IN PRECISION, RECALL AND F-SCORE ON FEATURES EXTRACTED USING MEAN EMBEDDING (CONTINUOUS BAG-OF-WORDS (CBOW))

Method	Precision	Recall	F-score
SVM	0.85	0.81	0.82
Multinomial NB	0.55	0.82	0.61
KNN	0.84	0.76	0.78
Random forest	0.92	0.62	0.78
Ensemble learning	0.87	0.79	**0.83**

TABLE V: CLASSIFIERS PERFORMANCE MEASURED IN PRECISION, RECALL AND F-SCORE ON FEATURES EXTRACTED USING WEIGHTED EMBEDDING (TF-IDF))

Method	Precision	Recall	F1-score
SVM	0.62	0.79	0.70
Multinomial NB	0.27	0.76	0.40
KNN	0.59	0.79	0.68
Random forest	0.92	0.38	0.54
Ensemble learning	0.81	0.72	0.76

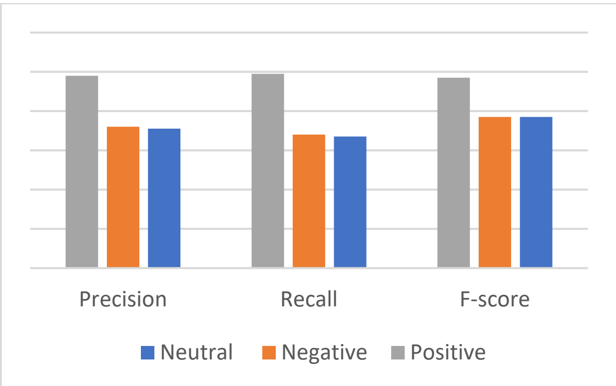

Fig. 1. Example of SVM detailed classes' results.

Finally, ensemble learning with Mean Embedding features performed the best compared to other embedding based classifiers. Due to the complexity of embedding features, which requires more sophisticated classifiers that can better model non-linearities such as deep neural network approaches, deep learning can capture the syntactic and the semantic information of the text unlike TF-IDF, which is purely dependent on the statistics to represent the text. In addition to that, the used dataset is imbalance. Therefore, our model requires incorporating imbalance techniques to boost the classification performance.

V. CONCLUSION AND FUTURE WORK

In this study, we leverage machine learning to analyze Arabic tweets' data during Sudanese revolution to help in the process of political decision-making. In this study, new dataset was collected from Sudanese revolution tweets. The proposed methodology utilized common machine learning algorithms with TF-IDF and word embeddings techniques as feature extraction. Word2vec was used to extract the feature because it is very useful in capturing the syntactic and semantic information of the sentences in the text data.

Experiments were conducted and the performance of the classification techniques showed that ensemble learning model with Word2Vec gave the best result compare with other machine learning classifiers.

For future, we intend to study sophisticated classifiers that can better model nonlinearities such as neural networks or deep learning approaches such convolutional neural networks. Since our dataset is small and imbalance, we plan to increase its size for future studies on Arabic political SA and incorporating imbalance techniques in our models to boost the classification performance.

CONFLICT OF INTEREST

The authors declare that no conflict of interest

AUTHOR CONTRIBUTIONS

First author collected the dataset and labeled it and also conducted the experiments and wrote the paper. The second and third authors guided the first author and helped in the whole process from preparing the dataset and conducting experiments to writing the paper. All authors have approved the final version.

ACKNOWLEDGMENT

The first author thanks his supervisor Dr. Eltyeb Elsamani for the help and support during the process of working on this paper. Also, thanks to my colleagues at Shaqra University for helping annotate the dataset and recommend some useful comments to improve the paper quality.

REFERENCES

[1] A. Bakliwal *et al*., *Sentiment Analysis of Political Tweets: Towards an Accurate Classifier*, Association for Computational Linguistics, 2013.

[2] A. Mahmoud and T. Elghazaly, "Using Twitter to monitor political sentiment for Arabic slang," *Intelligent Natural Language Processing: Trends and Applications*, Springer, pp. 53-66, 2018.

[3] A. Tumasjan *et al*., "Predicting elections with twitter: What 140 characters reveal about political sentiment," in *Proc. Fourth international AAAI Conference on Weblogs and Social Media*, Citeseer, 2010.

[4] Å. Bengtsson and H. S. Christensen, "Ideals and actions: Do citizens' patterns of political participation correspond to their conceptions of democracy?" *Government and Opposition*, vol. 51, no. 2, pp. 234-260, 2016.

[5] T. Gadekallu *et al*., "Application of sentiment analysis in movie reviews, in sentiment analysis and knowledge discovery in contemporary business," *IGI Global*, pp. 77-90, 2019.

[6] M. Kaya, G. Fidan, and I. H. Toroslu, "Sentiment analysis of Turkish political news," in *Proc. 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, 2012.

[7] X. Zhou *et al*., "Sentiment analysis on tweets for social events," in *Proc. the 2013 IEEE 17th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 2013.

[8] T. Elghazaly, A. Mahmoud, and H. A. Hefny, “Political sentiment analysis using twitter data,” in *Proc. the International Conference on Internet of Things and Cloud Computing*, 2016.

[9] R. Bose *et al*., “Analyzing political sentiment using twitter data,” *Information and Communication Technology for Intelligent Systems*, Springer, pp. 427-436, 2019.

[10] R. Parikh and M. J. C. N. F. R. Movassate. (2009). Sentiment analysis of user-generated twitter updates using various classification techniques. [Online]. Available: https://nlp.stanford.edu/courses/cs224n/2009/fp/19.pdf

[11] B. Liu, “Opinion mining and sentiment analysis,” *Web Data Mining*, Springer, pp. 459-526, 2011.

[12] G. Paltoglou and M. Thelwall, “Twitter, Myspace, Digg: Unsupervised sentiment analysis in social media,” *ACM Transactions on Intelligent Systems and Technology*, vol. 3, no. 4, pp. 1-19, 2012.

[13] A. Yousif *et al*., “A survey on sentiment analysis of scientific citations,” *Artificial Intelligence Review*, vol. 52, no. 3, pp. 1805-1838, 2019.

[14] P. Gamallo and M. Garcia, “Citius: A Naïve Bayes strategy for sentiment analysis on English tweets,” in *Proc. the 8th International Workshop on Semantic Evaluation*, 2014.

[15] P. Tripathi, S. K. Vishwakarma, and A. Lala, “Sentiment analysis of english tweets using rapid miner,” in *Proc. 2015 International Conference on Computational Intelligence and Communication Networks*, 2015.

[16] P. Basile and N. Novielli, “Uniba: Sentiment analysis of English tweets combining micro-blogging, lexicon and semantic features,” in *Proc. the 9th International Workshop on Semantic Evaluation*, 2015.

[17] A. Ghallab *et al*., *Arabic Sentiment Analysis: A Systematic Literature Review*, 2020.

[18] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up? Sentiment classification using machine learning techniques,” in *Proc. EMNLP*, 2002.

[19] M. A. Mesleh, “Chi square feature extraction based svms arabic language text categorization system,” *Journal of Computer Science*, vol. 3, no. 6, pp. 430-435, 2007.

[20] M. Rushdi-Saleh *et al*., *OCA: Opinion Corpus for Arabic*, vol. 62, no. 10, pp. 2045-2054, 2011.

[21] M. N. Al-Kabi, N. A. Abdulla, and M. Al-Ayyoub, “An analytical study of Arabic sentiments: Maktoob case study,” in *Proc. 8th International Conference for Internet Technology and Secured Transactions*, 2013.

[22] M. Elhawary and M. Elfeky, “Mining Arabic business reviews,” in *Proc. 2010 IEEE International Conference on Data Mining Workshops*, 2010.

[23] A. Zirikly and M. Diab, “Named entity recognition for Arabic social media,” in *Proc. the 1st Workshop on Vector Space Modeling for Natural Language Processing*, 2015.

[24] Y. Bengio *et al*., *A Neural Probabilistic Language Model*, vol. 3, pp. 1137-1155, 2003.

[25] A. M. Alayba *et al*., “Arabic language sentiment analysis on health services,” in *Proc. 2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR)*, 2017.

[26] K. Dashtipour *et al*., “A hybrid Persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks,” vol. 380, pp. 1-10, 2020.

[27] L. M. Q. Abualigah, *Feature Selection and Enhanced Krill Herd Algorithm for Text Document Clustering*, Springer, 2019.

[28] L. M. Abualigah and A. T. Khader, “Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering,” *The Journal of Supercomputing*, vol. 73, no. 11, pp. 4773-4795, 2017.

[29] T. Mikolov *et al*., *Efficient Estimation of Word Representations in Vector Space*, 2013.

Fatima Salih received her bachelor degree in information technology from Computer Man College, Sudan. She received her masters’ degree in computer science from the University of Gezira, Sudan. Currently, she is a lecturer at Shaqra University, Saudi Arabia. Her research interest includes machine learning and deep learning.

Shazali Siddig received his bachelor degree in computer science and his master in information technology from AL-Neelain University, and received his PhD in information technology from AL-Neelain University, Sudan. Currently, he is an assistant professor and researcher at Data Science Group at the Department of Information Technology, Al-Neelain University. His research interest includes machine learning, database systems, data science.

Eltyeb Elsamani received his bachelor degree in computer science from Al-Neelain University, master in computer science from Al-Neelain University, and PhD from Al-Neelain University, Sudan. Currently he is an associated professor in the Faculty of Computer Science and Information Technology, Al-Neelain University Sudan. His research interest includes machine learning, text analysis, database systems, and sentiment analysis.