Technological Forecasting & Social Change 202 (2024) 123326

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

An emoji feature-incorporated multi-view deep learning for explainable sentiment classification of social media reviews

Qianwen Ariel Xu [a], Chrisina Jayne [b], Victor Chang [a,*]

[a] Department of Operations and Information Management, Aston Business School, Aston Univegehrsity, Birmingham, UK
[b] School of Computing, Engineering and Digital Technologies, Teesside University, Middlesbrough, UK

ARTICLE INFO

Keywords:
Explainable sentiment analysis
Multi-view learning
High-stakes decision forecasting
Marketing analytics
Social media reviews

ABSTRACT

Sentiment analysis has demonstrated its value in a range of high-stakes domains. From financial markets to supply chain management, logistics, and technology legitimacy assessment, sentiment analysis offers insights into public sentiment, actionable data, and improved decision forecasting. This study contributes to this growing body of research by offering a novel multi-view deep learning approach to sentiment analysis that incorporates non-textual features like emojis. The proposed approach considers both textual and emoji views as distinct views of emotional information for the sentiment classification model, and the results acknowledge their individual and combined contributions to sentiment analysis. Comparative analysis with baseline classifiers reveals that incorporating emoji features significantly enriches sentiment analysis, enhancing the accuracy, F1-score, and execution time of the proposed model. Additionally, this study employs LIME for explainable sentiment analysis to provide insights into the model's decision-making process, enabling high-stakes businesses to understand the factors driving customer sentiment. The present study contributes to the literature on multi-view text classification in the context of social media and provides an innovative analytics method for businesses to extract valuable emotional information from electronic word of mouth (eWOM), which can help them stay ahead of the competition in a rapidly evolving digital landscape. In addition, the findings of this paper have important implications for policy development in digital communication and social media monitoring. Recognizing the importance of emojis in sentiment expression can inform policies by helping them better understand public sentiment and tailor policy solutions that better address the concerns of the public.

1. Introduction

1.1. Background

In the era of digital consumerism, electronic Word of Mouth (eWOM) plays a significant role in shaping customer opinions and influencing decision-making processes (Biswas et al., 2022; Stöckli and Khobzi, 2021). The online reviews from traditional customer review sites such as TripAdvisor, Yelp, and Amazon are common sources to inform decision-making. However, there is a risk of these reviews being manipulated (Sahut and Hajek, 2022), by companies creating artificial positive reviews or competitors creating malicious negative reviews. This can make decision-making, such as marketing analysis and preference prediction based on these reviews, even higher stakes. Therefore, businesses are shifting towards social media for a more genuine representation of customer sentiment. However, social media reviews also have their limitations. In particular, social media content, characterized by its informality and vast volume, demands an innovative approach to sentiment analysis. This is crucial in high-stakes business environments where accurate sentiment interpretation can significantly influence market predictions and strategic decisions, as demonstrated by studies such as Wołk (2020) in cryptocurrency price prediction, Mishev et al. (2020) in financial sentiment analysis, and Nguyen et al. (2023) in pharmaceutical demand forecasting during crises.

Within the informal language spectrum of social media, the usage of emojis has risen to prominence. Although the number of emojis is relatively small, the Unicode standard includes more than 3600 emojis as of September 2021.[1] Since the launch of emojis on Twitter in 2012,

* Corresponding author.
E-mail addresses: 220247906@aston.ac.uk, qianwen.ariel.xu@gmail.com (Q.A. Xu), C.Jayne@tees.ac.uk (C. Jayne), v.chang1@aston.ac.uk, ic.victor.chang@gmail.com (V. Chang).

[1] https://unicode.org/emoji/charts-14.0/emoji-counts.html

https://doi.org/10.1016/j.techfore.2024.123326
Received 26 April 2023; Received in revised form 14 December 2023; Accepted 5 March 2024
Available online 16 March 2024

usage has continued to rise and the upward trend has not changed. These pictorial symbols have become a crucial part of online communication, encapsulating an array of emotions and opinions that traditional text might fail to capture. This shift towards graphical expressions in online communication prompts our study to revisit the framework of Sentiment Analysis, as it is critically relevant in high-stakes sectors, where nuanced sentiment interpretation can impact financial markets, consumer behavior analysis, and even policy development, as indicated by the works of Hirata and Matsuda (2023) in logistics and Dehler-Holland et al. (2022) in assessing technology legitimacy. Traditionally, Sentiment Analysis, a natural language processing technique, extracts emotions and attitudes from the text (Agüero-Torales et al., 2019). It brings benefits to individuals, businesses, and governments by effectively identifying and classifying the emotions of people in written language to determine their opinions of things like events, services, products, etc., so as to develop timely and targeted strategies (Salur and Aydin, 2020). However, the popularity of emoji usage poses significant challenges (Hankamer and Liedtka, 2016).

In the process of data preprocessing, emojis are often removed, leading to the potential loss of sentiment information (Singla et al., 2022). While recognizing this gap, this study aims to develop a sentiment classifier for online reviews that incorporates emoji features from a multi-view learning perspective, enhancing the accuracy and comprehensiveness of high-stakes sentiment analysis, which will provide businesses with valuable insights into the sentiment of their customers. This approach is vital for high-stakes decision-making, aligning with the emerging needs in dynamic sectors like finance, healthcare, and logistics, as underscored by existing literature. In addition, to support business decision-making more efficiently, the proposed classifier is designed to operate without the need for additional preprocessing of the emoji features. Thus, the proposed classifier will also allow businesses to analyze large volumes of data more efficiently, thus enabling them to make better-informed decisions in a timely manner.

In recent years, while there have been studies on the application of advanced classification models, doubts about the practical adoption of these methods in the real world have also arisen. One of these concerns is the trust issue of these complex algorithms, as it is not easy for the decision makers to understand and comprehend the development process of the decisions output by the algorithms (Zytek et al., 2021). Therefore, they will be hesitant to act on their predictions, especially in high-stakes business, as misleading predictions may lead to significant financial loss. While aiming to address this issue, this study also applied one of the explainable artificial intelligence technology, LIME, to provide insights into the model's prediction process and foster trust among decision-makers. This transparency is critical in high-stakes environments, where understanding the nuances of sentiment analysis can lead to more informed and confident decision-making.

1.2. Contributions

The paper makes the following methodological and empirical contributions:

(1) It recognizes and addresses the gap in sentiment analysis that often overlooks the role of emojis, a key aspect of online communication. By doing so, it establishes a more realistic representation of sentiment in social media content.
(2) Experiments are conducted to test the impacts of different emoji handling methods on the effectiveness of sentiment classifiers individually or in combination using publicly available datasets. By comparing the performance of different algorithms with respect to the accuracy, F1-score and execution time, the results confirmed that emojis features can help to improve the effectiveness of the sentiment classifiers that use only textual features by almost 6.5 %.
(3) This study proposes an emoji feature-incorporated deep learning model for Twitter sentiment analysis. In this model, a more efficient Word_Emoji embedding layer is structured to generate both word and emoji embeddings instead of using separate embedding layers to generate them and then combine them. The performance of this classifier is evaluated and compared to the performance of classifiers using other emoji handling methods provided by existing studies, and the results show that this classifier has a comprehensive outperformance in terms of accuracy, F1-score, and execution time.
(4) The proposed multi-view sentiment analysis classifier is a powerful, intelligent business analytical tool that leverages the valuable information found in online reviews. On the one hand, the classifier requires minimal preprocessing of social media reviews to ensure its efficiency for businesses. By reducing the need for extensive preprocessing, the system can process large volumes of data more quickly and accurately, providing high-stakes decision forecasting with timely and actionable insights. On the other hand, it adopts interpretable artificial intelligence to visualize and explain the prediction results of the proposed classifier, supporting high-stakes decision-making.

1.3. Policy implications, utility, and applications

The findings of this study bear significant implications for policy formulation in digital communication and social media monitoring. By acknowledging the role of emojis in sentiment analysis and offering a new approach to incorporate them, this study potentially transforms how sentiment analysis is conducted, leading to a more accurate and comprehensive understanding of online sentiments. Recognizing the importance of emojis in sentiment expression can inform policies, promoting a more nuanced understanding of online communications.

Building on this, businesses and government agencies can utilize the proposed multi-perspective sentiment analysis approach to gain a deeper understanding of public sentiment. The application of this methodology spans a wide range of domains, including retail, hospitality, and public policy, and can contribute to policy development, policy communication strategies, and policy adjustments.

Emojis represent an emerging trend in digital expression, signaling a shift towards more graphical modes of online communication. This research embraces this change by proposing an innovative emoji feature incorporating sentiment analysis. As tools for understanding these graphical symbols must adapt to their evolution and complexity, this approach marks an advancement in this technological advancement. For the Technology Forecasting and Social Change (TFSC) audience, our research sheds light on the ever-changing digital communication landscape. By introducing a tool that can effectively process and interpret large amounts of multi-view social media content, this study paves the way for informed decision-making across industries. Our research, thus, aligns with the TFSC themes, providing insights into the future of sentiment analysis and its implications on online communication trends.

Section 2 introduces the existing literature on the research about sentiment analysis of online reviews and the role of emojis in sentiment analysis and summarizes the existing emoji handling methods. Section 3 presents the datasets and the data preprocessing process for this study. Section 4 describes the experimental steps, the research questions and how they can be answered through the experiments. The models, emoji handling methods and the evaluation metrics employed are also presented. Section 5 reports the experiment results and answers the research questions posed. Finally, Section 6 summarizes the study and discusses its contributions and limitations.

2. Literature review

This study focuses on high-stakes environments and aims to analyze the impact of incorporating emoji features on identifying sentiments of online reviews by sentiment classifiers. In this section, this study reviews the application of sentiment classifiers in high-stakes business environments, as well as the role of emoji features in multi-view sentiment analysis.

2.1. Sentiment analysis in high-stakes environments

The application of sentiment analysis in high-stakes business environments has gained significant traction, as evidenced by recent studies across various sectors.

Wołk's (2020) research demonstrates the pivotal role of sentiment analysis in cryptocurrency markets, particularly in predicting Bitcoin prices. Using Twitter and Google Trends, the study employs methods such as AdaBoost, Decision Tree, and Gradient Boosting, and reveals that cryptocurrency price fluctuations are predominantly influenced by public perceptions and opinions, rather than institutional regulation. This finding is crucial in high-stakes environments like cryptocurrency trading, where market sentiment can lead to rapid and significant financial impacts. Similarly, Mishev et al. (2020) explore sentiment analysis in finance, emphasizing the challenge posed by domain-specific language and the scarcity of large labeled datasets. Their evaluation of various sentiment analysis approaches, including lexicons and NLP transformers, showcases the effectiveness of advanced techniques in extracting actionable signals from financial news, which is vital for investment decision-making.

In the context of the pharmaceutical industry, Nguyen et al. (2023) highlight the utility of sentiment analysis in managing demand volatility during disruptive events like epidemics. Their development of a CamemBERT-based sentiment analysis model, which structures information from medicine-related news, exemplifies how sentiment analysis can enhance demand forecasting accuracy in times of crisis. This approach is particularly relevant for high-stakes decision-making in the pharmaceutical sector, where accurate predictions can have significant public health implications. Hirata and Matsuda (2023) focus on the logistics sector in post-pandemic Japan, utilizing sentiment analysis based on BERT algorithm of Twitter data to examine logistics trends. Their findings indicate a positive sentiment towards logistics and an increasing interest in the field. The study illustrates how sentiment analysis can serve as a powerful tool for understanding industry challenges and informing strategic decisions in logistics, a sector where efficient and timely operations are critical. Dehler-Holland et al. (2022) assess the legitimacy of wind power technology in Germany through lexicon-based sentiment analysis of newspaper articles. Their work demonstrates the broader implications of sentiment analysis, extending to policy development and public perception. By identifying the contexts and challenges faced by wind power, the study shows how sentiment analysis can influence policy decisions and maintain the legitimacy of technologies vital for sustainability.

The above studies highlight the versatility and importance of sentiment analysis in high-risk business environments. Whether in financial markets, crisis management in the pharmaceutical industry, logistics planning, or assessing the legitimacy of technology, sentiment analysis provides valuable insights for strategic decision-making.

2.2. Emojis in twitter sentiment analysis

Multiview data are a type of data that describe objects or phenomena through different feature sets or perspectives, such as combining text and image or web page and clickthrough data. These data are increasingly available in real-world applications, which can be used in conjunction with machine learning to yield more significant results compared to single-view representation learning (Zhang et al., 2022). For example, tweets are a form of multiview data that combines textual and visual elements like emojis, making them valuable for sentiment analysis. There are two types of facial expressions, including emoticons and emojis. Emoticons are made up of ASCII and are the predecessor to emojis, which are in image form. According to the Oxford Dictionary, emojis are facial expressions made up of various combinations of keyboard characters, such as smiles (:)), while emoticons are small digital images or icons used to express ideas or emotions, such as ☺ . A growing body of work has shown interest in considering emoji features as a way to enhance sentiment analysis on such data, particularly on social media platforms.

Emojis can alter the sentiment polarity of posts or tweets through subtle interactions with text. In the study by Lou et al. (2020), posts in which sentiment polarity changed and did not change as a result of emojis were investigated. They found in the data that the polarity of 4044 posts altered owing to emojis, representing 40.27 % of all posts.

Hankamer and Liedtka (2016) were the first researchers to take emojis into consideration in sentiment analysis studies after the widespread use of emojis on Twitter. Due to the lack of labeled tweet datasets that contain emojis, they collected the data themselves. Each sample in the dataset contains emojis and is labeled by VADER (Hutto and Gilbert, 2014), a lexicon-based approach. They used two methods to handle the emojis. The first method calculates the average "emoji score" per Tweet according to the occurrence information collated by Kralj Novak et al. (2015). This method was also employed in the study (Bansal and Srivastava, 2019) for the prediction of vote shares in the 2017 Uttar Pradesh legislative elections and was approved to decrease the prediction error of the lexicon-based approach significantly. They took the number of positive occurrences of each emoji, subtracted the number of negative occurrences, and then divided it by the number of total occurrences. The second way in the study of Hankamer and Liedtka (2016) is called "emoji substitution." They replaced each emoji with its alias (which is a word or several words) and averaged the GloVe embeddings of the alias to obtain an emoji embedding. In their case, the Shallow Neural Network performs better when adding an emoji score dimension, while the Recurrent Neural Network (RNN) significantly gains performance when using both emoji handling methods. Similar to Hankamer and Liedtka (2016), A. Singh et al. (2019) also used "emoji substitution" on the Twitter classification problem, although in their case, it is called the "emoji description strategy". Moreover, they also tried the direct use of pre-trained emoji embeddings, called the "emoji embedding strategy". The embeddings obtained by these two methods were learned by the BiSLTM model with an attention mechanism, respectively, and applied to the two classification tasks, i.e., irony detection and topic-based sentiment analysis. They compared the results and concluded that replacing emojis with their textual descriptions is more effective than using emoji embeddings.

Bansal and Srivastava (2019) integrated the method of adding emoji scores to lexicon-based approaches in their study of election prediction. They computed the overall sentiment of a tweet by adding up the sentiments of words provided by lexicon-based classifiers and the sentiment scores of emojis in each tweet. Then, they defined the vote share for each election party based on the overall score. To evaluate the effectiveness of the emoji scores, they evaluated their lexicon-based approaches by comparing their predicted vote share for each party to the true shares using mean absolute error (MAE). The results show that combining emoji sentiments reduces MAE for most lexicons, where the VADER lexicon performs the best (Hutto and Gilbert, 2014). However, the effect of this improvement is only more than 1 %, which may relate to the low number of emojis (1.45 %) discovered in the data.

Liu et al. (2021) presented two other ways of dealing with emojis in the text. Firstly, they defined two types of words to present emojis in their study. One is an emotion word, a word that directly indicates an emotion (e.g., 😊 happy), and the other is an emoji tag word, a word that describes an emoji (e.g., 😊 smiling face). One of their methods is to convert all emojis into corresponding sentiment words instead of the tag words, as they considered tag words to be ambiguous and could affect the sentiment recognition of the sentiment analysis algorithm. However, they compared the changes in algorithm performance and found that emoji tags' ambiguity did not show a negative effect on sentiment detection. They also considered the sentimental coherence between plain texts and emojis. According to Liu et al. (2021), the results show that posts where the emoji sentiment is inconsistent with the sentiment of the text, tend to compromise the performance of the SA algorithm. However, the dataset of this experiment only includes consistent samples. Therefore, the results may require further investigation.

In contrast to Liu et al. (2021), Lou et al. used the SkipGram mode of word2vec to train Chinese words and emojis simultaneously to obtain embedding representations (Lou et al., 2020). They trained the embeddings of words or emojis in a corpus of 3.5 million posts with a total vocabulary of 252,267. They proposed a deep learning model (EA-Bi-LSTM) to test the effectiveness of emoji embedding. Their model uses Bi-LSTM to read the text in both directions and then aggregates these informative word representations to create sentence representations using an attention mechanism. Their model proved to be the best performer, greatly outperforming all baseline models. Moreover, their experiments showed that both emojis and text performed an essential role in the sentiment recognition of microblog posts. While emojis had a stronger effect on the sentiment polarity of posts than text, the deep learning models that used both features performed better. However, all models performed extremely poorly in classifying neutral emotions.

To sum up, following a survey of the sentiment analysis literature related to emojis processing, this paper identifies the following types of emojis processing.

1) replacing an emoji for the corresponding descriptive words
2) replacing an emoji for the corresponding emotion words
3) adding an emoji score as an additional feature
4) transforming emojis into emoji embeddings using pre-trained emoji embeddings
5) manually annotating the sentiment consistency of the emoji with the plain text and using "sentiment consistency" as an additional feature
6) building own corpus and simultaneously training words and emoji embeddings
7) Employing BERT tokenizer with Transformer encoder

When using social media platforms like Twitter, people tend to express themselves in an effortless and quick way (de Barros et al., 2021), which is one of the reasons why the use of emojis is becoming increasingly popular. In sentiment analysis, the emoji processing approaches discussed above provide useful sentiment information for identifying the sentiments expressed by users through short texts, greatly improving the performance of sentiment classifiers.

2.3. Explainable AI and its role in sentiment analysis

Explainable AI (XAI), which focuses on enhancing the interpretability and transparency of AI models, has found applicability across a myriad of domains and has significantly influenced sentiment analysis, futures price series prediction, and even professional athlete scouting (Ghosh et al., 2022; Haque et al., 2023; Janssens et al., 2022a, 2022b). Recent advancements in this domain have further highlighted its value, especially when applied to the hospitality industry (Ghosh et al., 2023).

The study by Ghosh et al. (2022) underscores the vitality of XAI in deciphering the decision-making process of complex AI models. Their use of ensemble feature selection in combination with advanced AI-based predictive modeling elegantly illuminated the role of various explanatory features in predicting future price series, thus exemplifying the power of XAI. In a parallel pursuit, Chowdhury et al. (2021) delved into the interpretability of Bi-directional Long Short-Term Memory (LSTM) networks, a type of recurrent neural network known for its complexity, in sentiment analysis. Applying the Local Interpretable Model Diagnostics Explanation (LIME) framework, they successfully decrypted important features and their interactions during prediction.

The overarching domain of sentiment analysis has been particularly enriched by the advancements in XAI. This trend is evident in the works of Miron et al. (2023) and Dewi et al. (2022). Miron et al. (2023) used a unique sampling method to boost the performance of LIME for Aspect-Based Sentiment Classification (ABSC), thus offering deeper insights into the complex decision-making processes. Dewi et al. (2022) took a similar route by employing the SHAP method to explain a BERT model's decision-making in sentiment analysis of movie reviews, providing intuitive and meaningful explanations.

Expanding upon the traditional utilization of XAI, Moreira et al. (2021) and Lampridis et al. (2020) designed novel frameworks, LINDA-BN and xspells, respectively. These unique tools illuminated the underlying rationale behind predictions, either through local post-hoc interpretations or the generation of synthetic sentences, exemplifying the immense potential of XAI. This notion is further reiterated by Yang et al.'s (2023) study, where XAI was integrated with sentiment analysis, topic modeling, and Extreme Gradient Boosting (XGBoost) to predict customer ratings from online reviews. This comprehensive approach demystified complex prediction patterns and highlighted the crucial factors affecting predictions, displaying the prowess of XAI in deriving insights from unstructured data.

In contrast to these technical approaches, Kim et al. (2020, 2023) highlighted the critical aspect of user preferences in the development of XAI systems. Their findings emphasized that local explanations, visualizations, and transparency can lead to a more intuitive AI decision support system, thereby fostering user trust and acceptance. These studies illuminate the diverse applicability and significance of XAI, affirming the indispensable role of explainability in our proposed model.

2.4. Gaps and limitations of the existing studies

Upon thorough review of existing literature, it is evident that while sentiment analysis accuracy improves with expressive data processing, it also substantially complicates the data handling process. For example, it may not be practical to use the manually annotated sentiment consistency of an emoji with plain text (Liu et al., 2021) as an additional sentiment feature, as this feature is not an attribute value present in the text provided by the social media network. In addition, some approaches require the separation of text and expressions in order to process them separately and then merge them (Hankamer and Liedtka, 2016; Bansal and Srivastava, 2019), and some require the construction of their own corpus (Liu et al., 2021).

Another notable gap in the current research is the unrealistic usage of datasets, where all samples contain emojis. This significantly deviates from real-life scenarios, leading to an overemphasis on the impact of emojis in sentiment analysis. In the context of these limitations, our study proposes an emoji-incorporated deep learning sentiment classifier that minimizes the need for such exhaustive preprocessing. This study strategically aims to handle emojis in a practical manner by treating them as part of the input data without requiring separate processing. This approach substantially simplifies the data processing and makes the model more applicable to real-world data. Furthermore, to create a more representative study, this research employs a dataset with a realistic percentage of emojis, contrary to the often inflated representation in existing works, which we believe enhances the external validity of the proposed model.

In addition, while most existing studies on multi-view sentiment analysis overlook the importance of transparency, this study adopts Explainable Artificial Intelligence (XAI) techniques to improve model transparency. This feature adds significant value by enabling users to understand the rationale behind the model's predictions, thus building trust and facilitating better decision-making.

Fig. 1 illustrates the explainable proposed emoji-incorporated sentiment classifier as an intelligent high-stakes business analytical tool and compares the proposed classifier with other major approaches to highlight its advantages. Emojis have the potential to alter the entire meaning of a review. For example, the review "Thank you 😢" presented in the system shows that the emoji 😢 (sad crying face) implies that the

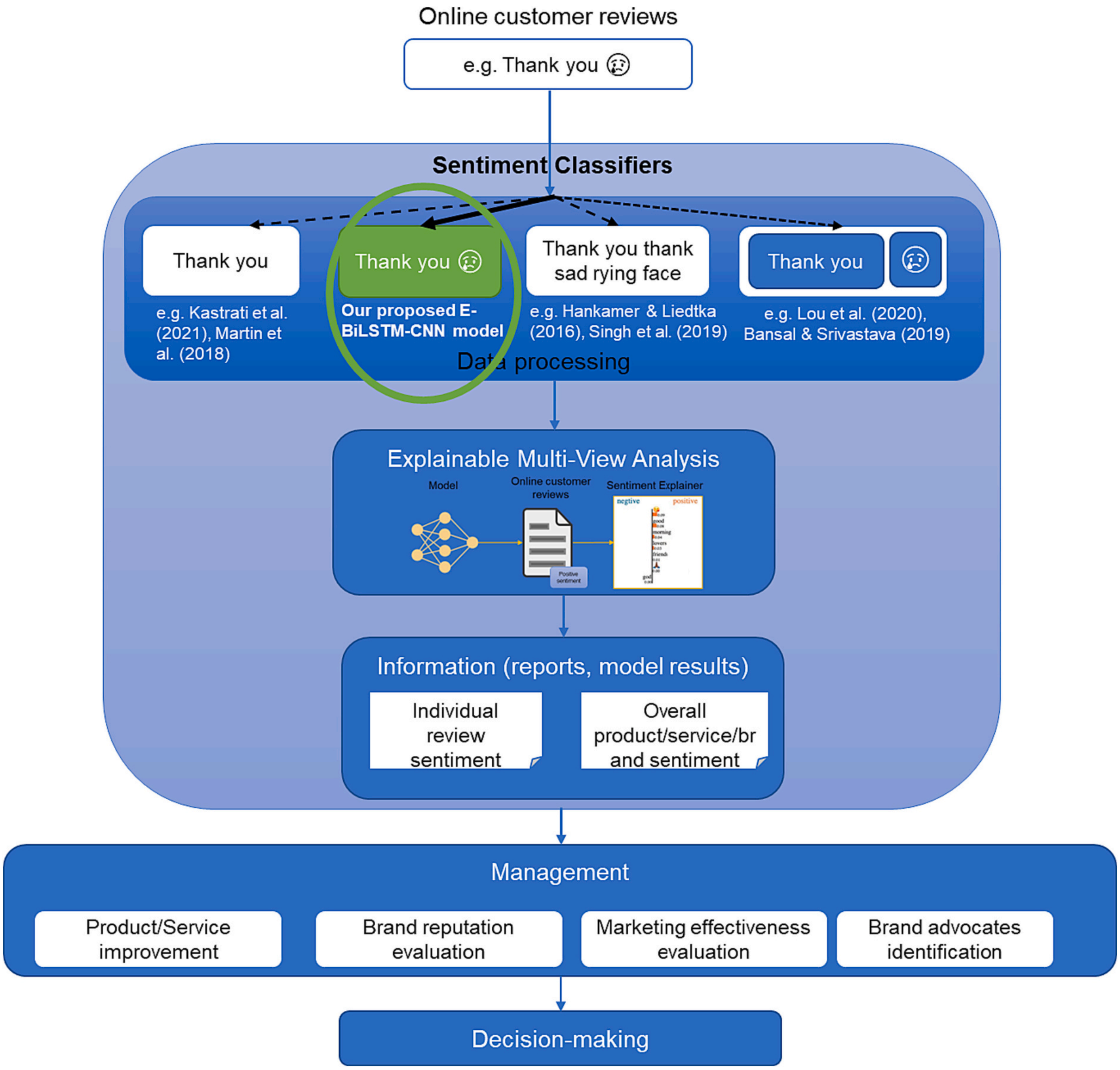


Fig. 1. The proposed explainable high-stakes business analytical system

customer may not be satisfied and has a negative sentiment. Without this emoji, the sentiment would be positive, expressing gratitude. Therefore, the classifiers used by Kastrati et al. (2021) and Martín et al. (2018) may fail to identify such sentiments, resulting in missing or misleading information for business. In addition, compared to classifiers used in studies such as Singh et al. (2019) and Lou et al. (2020), the online reviews imported to the proposed classifier do not require further processing of emojis, such as replacing emojis with text or separating emojis from text for transformation into scores or embeddings individually. In addition, in order to increase the transparency of the proposed system, it employs a LIME-based interpretable technique to visualize the factors or features on which the system's outputs are based, so as to maximize the ancillary functions of the system. Overall, the proposed classifier aims to provide a more efficient and accurate sentiment analysis of online reviews, which can help decision-makers in product/service improvement, brand reputation evaluation, marketing effectiveness evaluation, and identifying brand advocates.

3. Data

3.1. Data collection

In recent years, a growing body of work has also examined the role of emojis in sentiment analysis. While they proposed various methods to convert emojis to sentiment features, they ignored the issue of consistency between the dataset used and real-life datasets, for example, in terms of data distribution. This paper argues that it is important that the distribution of the data used for training the model is as close as possible to that used for testing the model in real life, which is also reflected in the study (Hankamer and Liedtka, 2016; de Barros et al., 2021). Aiming to construct an ideal dataset that can simulate a realistic distribution of tweets containing emojis, this study conducted an investigation into the ratio of tweets containing emojis to total tweets. Emojipedia (2022) reported that about 21.5 % of tweets contain emojis at the end of 2021. Therefore, this study determined the ratio is 20 %. Based on this, this study constructed a Modern Tweet Dataset for the proposed study by the use of a Sentiment 140 Dataset and an Emoji Tweet Dataset.

Sentiment 140 Dataset comprises 1.6 million tweets provided by Go et al. (2009). This dataset is class-balanced, with a 50/50 split between those labeled as positive and negative emotions. This study chose the Sentiment 140 Dataset because it is one of the most frequently used datasets in this domain, and its quality has been proven by many studies. In addition, it is less restricted to one specific domain compared to other datasets, covering various brands, products, or topics on Twitter. The Emoji Tweet Dataset is provided by Yan (2020). The dataset is also not limited to a specific domain. There are 16,011 pieces of data in total, each containing emoji. This dataset's class is balanced, with 8010 entries classified as negative and 8001 entries classified as positive.

Based on the Sentiment 140 Dataset and the Emoji Tweet Dataset, this study constructed a Modern Tweet Dataset that contains a total of 80,000 tweets, with a 20 % share of tweets containing emoji. This dataset is also balanced, with 40,000 samples labeled as positive and 40,000 samples labeled as negative.

3.2. Basic data preprocessing

Data from social networking sites are often non-structured and contain noisy information that is irrelevant and inefficient, and do not convey textual emotional meaning in the majority of cases (Priyadarshini and Cotton, 2021). Singla et al. (2022) claim that preprocessing is critical in identifying emotions or sentiments in non-uniform text input. To effectively conduct the classification tasks, a variety of data preprocessing techniques are required to convert text into an analyzable and predictable form and to derive relevant information from massive data.

Since one of the research purposes is to assess how emojis affect the effectiveness of machine learning algorithms, this study will conduct a basic preprocessing of the data beforehand. The preprocessing techniques consist of changing capital letters to lower case, removing web links, removing mentions (@), removing hashtags and punctuations, reducing consecutive repeated letters in the vocabulary, changing contractions to their full forms, removing stop words, and finally, removing extra spaces from the text. It is worth noting that, unlike other studies, this study has only removed the hash symbols (#) of the topic labels, leaving the topics that were carried. While exploring the data, this study found some topics containing sentimental messages, such as #lovethis and #funbutwrong. Therefore, these topics were left in this study. In addition, stop words are a group of frequently used terms in all languages, not just English, and removing them from the text corpus enhances model performance and makes the model more robust. This study employed the list of stop words provided by the NLTK package to remove stop words from the data samples. According to HaCohen-Kerner et al. (2020), however, the removal of stop words may also alter the meaning of the sentences, which has an impact on the accuracy of the classifiers. Therefore, this study remained the negatives in the text, including ‘but’, ‘no’, ‘nor’, and ‘not’. The finished text was processed by retaining the various emojis to conduct the following experiments. An example of a review before and after the preprocessing operation is shown as follows (Fig. 2):

4. Methodology and experiments

4.1. Experimental procedure

While aiming to assess the effects of handling emojis on the effectiveness of different classifiers and the effectiveness of the proposed model, this study first divided the review dataset into a training set (80 %) and a test set (20 %). Four types of experiments were carried out. The first type is to train sentiment classifiers with data, removing all emojis. The second type is to perform only one type of emoji processing method before training. The third type is to perform any two of the emoji processing methods, and the last one is to perform all three methods. All these experiments will be conducted using each classifier described in Section 4.2. This study trained the classifiers using the training set and evaluated them on the test set to see their ability to learn emoji features.

```
text = "I don't ❤️ flying @VirginAmerica. :D heyyyy ☺👍 ..&"
basic_preprocess_apply(text)
```

```
'not ❤️ flying laugh heyy ☺👍 '
```

Fig. 2. A review example before and after the preprocessing operation

To be specific, the purpose of the experiments is to address four research questions as follows:

1) Does the consideration of emojis as features facilitate sentiment recognition of online reviews by sentiment analyzers?

Rigorous contrast experiments were carried out to provide an answer to this question. To determine the effect of emojis on sentiment classification, this study compared the classifiers' performance for tweets with emoji features and text-only tweets.

2) Which of the methods proposed in this paper for transforming emojis into features, emoji replacement, adding emoji scores and creating emoji embeddings is the best for each algorithm?

This study put forward three methods to handle emojis and compare their impact on emotion recognition in their individual and combined forms, respectively. A detailed description of the three processing methods for handling emojis is given in Section 4.3

3) Does the emoji processing approach presented in this study outperform the others?

This study proposed a new method to handle emojis in the text by creating emoji embeddings along with word embeddings, which is presented in Section 4.3. The effectiveness of the new method *E*-BiLSTM-CNN is compared with the other classifiers.

4) Does the model, E-BiLSTM-CNN, outperform other sentiment classifiers when execution time is considered?

To answer this question, after exploring the performance of each algorithm with various emoji treatments and their combinations, this study extracted the performance data of the best classifiers for each algorithm and compared them. In addition, this study evaluates these classifiers by performing a weighted average of their performance in terms of F1-score and execution time.

4.2. Models

4.2.1. Baseline models

This study uses three classical machine learning algorithms as baseline models, including Bernoulli Naïve Bayes, Support Vector Machine, and Logistic Regression. All these algorithms will be employed and tested in each experiment, and their performance will be evaluated against each other and against the proposed model to address the research questions. This study denotes *Algorithm (T)* as the implementation of a selected algorithm in texts after removing emojis, *Algorithm (D)* as the implementation in tweets with emojis converted into their descriptions, *Algorithm (ES)* as the implementation in tweets with an additional feature of emojis score, *Algorithm (EB)* as the implementation in tweets with an additional feature of emoji embeddings, *Algorithm (D + ES)* as the implementation in tweets with emojis replaced and emoji scores added, *Algorithm (D + EB)* as the implementation in tweets with emojis replaced and emoji embeddings added, *Algorithm (ES + EB)* as the implementation in tweets with emoji scores and emoji embeddings added, *Algorithm (D + ES + EB)* as the implementation in tweets with all three emoji handling methods applied. This study presents detailed information on the settings for each algorithm as follows.

4.2.1.1. Bernoulli Naïve Bayes (BernoulliNB). Multinomial Naive Bayes, Gaussian Naive Bayes, Bernoulli Naive Bayes are three types of Naïve Bayes. By initially exploring their effectiveness in sentiment recognition, Bernoulli NB was finally chosen to conduct the experiments. This study denotes each experiment conducted by Bernoulli NB as *BernoulliNB (T), BernoulliNB (D), BernoulliNB (ES), BernoulliNB (EB), BernoulliNB (D + ES), BernoulliNB (D + EB), BernoulliNB (ES + EB)*, and *BernoulliNB (D + ES + EB)*.

4.2.1.2. Support vector machine (SVM). SVM is a powerful algorithm that has been proven to be useful in sentiment analysis (Chen et al., 2021). This research also tested its ability to learn emotional information from different emoji features. Each experiment conducted by SVM is named as *SVM (T), SVM (D), SVM (ES), SVM (EB), SVM (D + ES), SVM (D + EB), SVM (ES + EB),* and *SVM (D + ES + EB)*.

4.2.1.3. Logistic regression (LR). LR is a widely employed algorithm that serves to solve the binary classification problem (Xiao et al., 2021; Książek et al., 2021). In this research, the performance of LR in identifying text sentiment is also evaluated and compared when using different emoji handling methods. The experiments are named as *LR (T), LR (D), LR (ES), LR (EB), LR (D + ES), LR (D + EB), LR (ES + EB),* and *LR (D + ES + EB)*.

4.2.2. E-BiLSTM-CNN model

While exploring the influence of emojis on sentiment analysis from a multi-view perspective, this study presents an emoji-incorporated BiLSTM-CNN model (E-BiLSTM-CNN). To be specific, this model is built on a deep learning architecture that introduces emojis in tweets. It employs Bidirectional Long Short-Term Memory (BiLSTM) and Convolutional Neural Network (CNN) to extract key features from the text and learn their relationship with users' sentiments. As shown in Fig. 3, the proposed model has eight layers: the input layer, the embedding layer, the BiLSTM layer, the CNN layer, the max pooling layer, the concatenation layer, the dense layer, and the output layer.

Given an input Tweet T_i that consists of T elements s_t, any element of a tweet, including plain texts or emojis, are used as features. Then, a Tweet is described as $\{w_1, w_2, \ldots, w_a, e_1, e_2, \ldots, e_b\}$, where w_a refers to the word token and e_b refers to the emoji token, and $a + b = t \in [1,T]$. Another input is the normalized average “emoji score” of the tweet, es_i, and the calculation method is discussed in Section 4.3. Each word or emoji token is transformed to a vector representation, x_t, through the embedding layer to be the input of the Bi-LSTM layer and CNN layer to obtain a tweet representation. The emoji score feature is then concatenated with the features that were derived from the CNN layer. Dropout layers and dropout rates are employed to prevent the issue of overfitting in neural networks. Finally, the output layer applies the softmax activation function to compute a probability distribution of the tweet's sentiment polarity. Each layer of this deep learning architecture is introduced in the following sections.

A. Word_Emoji Embedding layer:

The Word_Emoji Embedding Layer serves as an initial layer in the *E-BiLSTM-CNN* model. Given an input Tweet *Ti* with elements (words and emojis) e_t, $t \in [1,T]$, the element e_i is transformed to a real-valued vector x_t, through an embedding matrix W_e. The conversion equation is shown below:

$$x_t = W_e e_t \tag{1}$$

$x_t \in Rd$, where d refers to the embeddings' dimension. The present study employed pre-trained word embeddings provided by GloVe and pre-trained emoji embeddings provided by Emoji2Vec, creating the Word_Emoji embedding layer. The output from this layer is a set of vectors $x = \{x_1, x_2, \ldots, x_t\}$.

B. Bidirectional LSTM layer:

BiLSTM is a variant of Recurrent Neural Network (RNN), which was proposed by Graves and Schmidhuber (2005). It was designed to address the drawbacks of the RNN model in terms of gradient explosion and

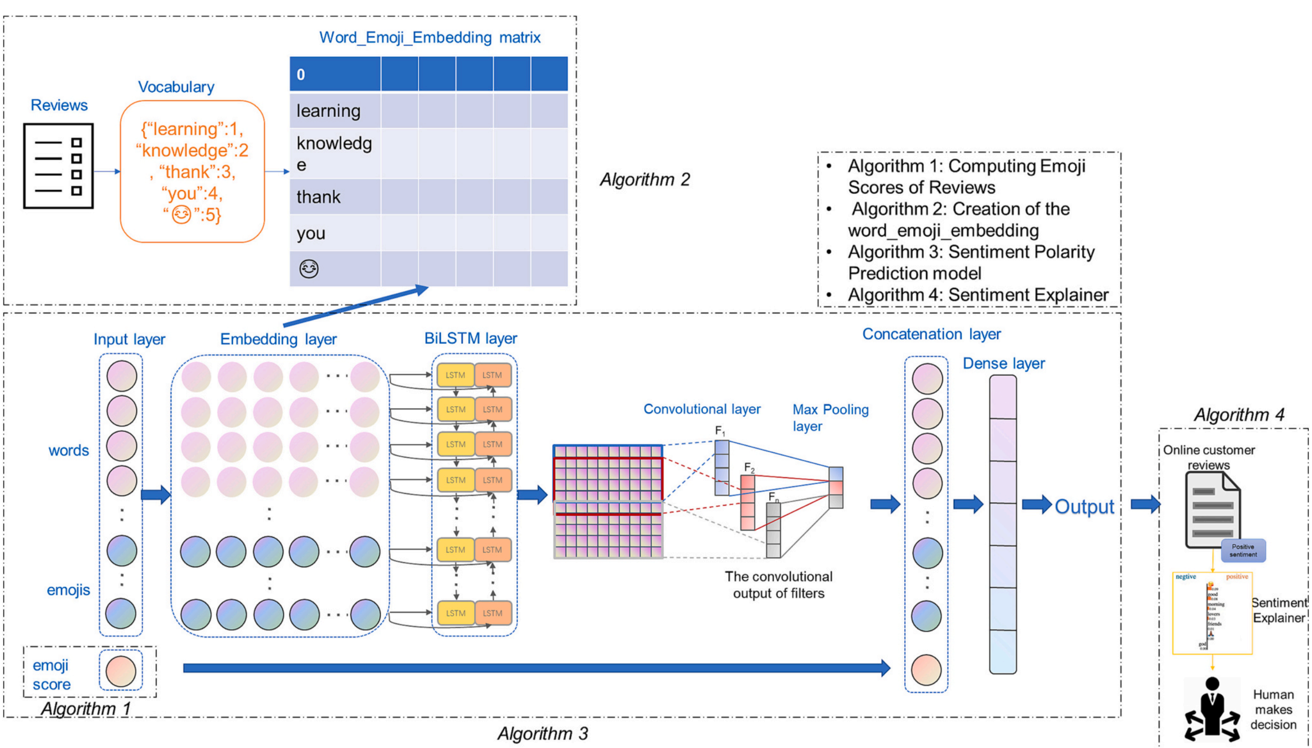


Fig. 3. The proposed emoji-incorporated deep learning model.

disappearance. Many researchers have employed BiLSTM models for text classification tasks and achieved excellent performance (Salur and Aydin, 2020; Zheng and Zheng, 2019). Abedin et al. (2021) constructed an exchange rate forecasting model based on BiLSTM and Bagging Ridge regression, which showed significant predictive performance and identified the currencies with the greatest impact on the US dollar. In this study, BiLSTM models are used in sentiment analysis to learn the sentence representations, which are subsequently utilized as features for sentiment classification.

The LSTM model consists of several LSTM units that are employed to capture long-range dependencies in a sequence. Each cell models memory in a neural network. The cell states are regulated by three gates, including input, forget and output gates, to enable the LSTM to store and access information over time (Lou et al., 2020; Efat et al., 2022).

First, by examining the input (x_t) and hidden state (h_{t-1}) values, the forget gate (f_t) decides whether to maintain or discard the information from the preceding cell state (c_{t-1}). The gate outputs a value of 0 or 1. In the same way, the input gate (i_t) determines the amount of information to be updated in the hidden state (h_{t-1}) and input text (x_t). A new candidate value vector G_t is also created through the tanh layer (Zheng and Zheng, 2019). The previous cell state c_{t-1} is updated with useful information retained by multiplying c_{t-1} and f_t, and new information from the new candidate value G_t by adding the product of i_t and G_t. The created cell state is represented by the value of c_t. The forget gate (f_t), input gate (i_t), new candidate value (G_t), and the created cell (c_t) are expressed as follows:

$$f_t = sigmoid\left(W_{fx}x_t + W_{fh}h_{t-1} + b_f\right) \tag{2}$$

$$i_t = sigmoid\left(W_{ix}x_t + W_{ih}h_{t-1} + b_i\right) \tag{3}$$

$$G_t = tanh(W_{cx}x_t + W_{ch}h_{t-1} + b_c) \tag{4}$$

$$c_t = c_{t-1} \odot f_t + i_t \odot G_t \tag{5}$$

The output gate (o_t) is responsible for managing the information flow from the current cell state (c_t) to the hidden state (h_t). It decides which part of the cell state is to be output by evaluating the hidden state (h_{t-1}) and input (x_t). Then, the output gate (o_t)'s output is multiplied by the current cell state (c_t) dealt with by the tanh gate to determine the current hidden state.

$$o_t = sigmoid\left(W_{ox}x_t + W_{oh}h_{t-1} + b_o\right) \tag{6}$$

$$h_t = o_t \odot tanh(c_t) \tag{7}$$

In this paper, the *E*-BiLSTM-CNN model uses a BiLSTM to read the text in both directions (Kamyab et al., 2021). BiLSTM contains a forward LSTM and a backward LSTM for reading text in the direction from x_1 to x_t and from x_t to x_1, respectively. The hidden state of the forward LSTM and backward LSTM are presented as $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$. A word can then be represented by concatenating the two states as h_t.

$$\overrightarrow{h_t} = \text{LSTM}\left(x_t, \overrightarrow{h_{t-1}}\right) \tag{8}$$

$$\overleftarrow{h_t} = \text{LSTM}\left(x_t, \overleftarrow{h_{t+1}}\right) \tag{9}$$

$$h_t = \left[\overrightarrow{h_t}, \overleftarrow{h_t}\right] \tag{10}$$

In this way, the representation of the text as $[h_0, h_1, h_2, \ldots, h_T]$ is obtained and fed to a convolutional layer to extract important features.

C. Convolutional layer:

CNN is another kind of neural network that is utilized to predict time series. They are biologically inspired variants of feed-forward neural networks. Because of their capacity to utilize spatially localized correlations in images, they are mainly employed for computer vision issues, but can also be applied to time-series problems such as sentiment analysis. In the CNN layer, the most significant higher-order features in the text are extracted (Khan and Niu, 2021). It first extracts local features over the matrix $h = [h_0, h_1, h_2, \ldots, h_T]$ output from the previous BiLSTM Layer. A group of k filters is applied, each for a window of *q* words, producing a new feature a_i from a window of vectors $h_{i:i+q-1}$. The new feature a_i can be represented as follows:

$$a_i = f\left(F \circ h_{i:i+q-1} + b\right) \tag{11}$$

where $F \in R^{l \times d}$ refers to the filter, b denotes the bias, and f refers to the activation function, which is ReLU in the present study. A feature map $c = [a_1, a_2, \ldots, a_{n-l+1}]$ is created by applying the filter to each window, resulting in k feature maps with k filters.

Only a few words and their combinations can provide relevant information about the meaning of a text in text classification tasks, while the max pooling layer allows for the discovery of the hidden semantic variables in the text (Rao and Yang, 2022). Therefore, after the convolutional operation, the max pooling operation is applied to feature maps to extract $m = \max\{c\}$, which refers to the maximum value. As a result, the output of the CNN layer is obtained by combing the maximum values from the pooling operation, which is $m = \{m_1, m_2 \ldots m_k\}$.

D. Concatenate layer, dense layer and output layer

The concatenate layer combines the features extracted by the CNN layer and the emoji score features into one layer, which is then passed on to the dense layer. In any neural network, a dense layer refers to a layer that is deeply linked to the previous layer (S. Wang et al., 2019). Each neuron in the dense layer is linked to each neuron in the previous layer. In this study, two dense layers will be employed. The reason for this is that convolutional layers attempt to extract features in a distinguishable way, while fully connected layers attempt to categorize the features. According to (Samala et al., 2017), there are more generic features in the early features of ConvNet that are useful for many tasks. At the same time, subsequent layers of the ConvNet become progressively more specialized to the characteristics of the classes contained in the original dataset. As a result, increasing the number of dense layers might help to perform a better classification of the extracted features (Suzuki et al., 2016; He et al., 2020). Dropout is a commonly employed regularization technique. It is employed to deal with the issue of overfitting. The dropout mechanism randomly drops some neurons to create a robust model, avoiding over-fitting. The dropout rate of 0.3 is employed in the proposed model.

The final layer of the model is the output layer. As this study addresses a binary sentiment classification task, binary cross-entropy is employed as the loss function. The equation of the binary cross entropy is presented as follows:

Binary cross entropy =

$$-\frac{1}{m}\sum_{i}^{m}\left(y_i * log(p(y_i)) + (1 - y_i) * log(1 - p(y_i))\right) \tag{12}$$

where m denotes the total number of text samples, y_i refers to the actual labels, $p(y_i)$ refers to the probability of actual labels.

As with the baseline model, the deep learning model will be executed in all experiments and are named as *E-BiLSTM-CNN (T), E-BiLSTM-CNN (D), E-BiLSTM-CNN (ES), E-BiLSTM-CNN (EB), E-BiLSTM-CNN (D + ES), E-BiLSTM-CNN (D + EB), E-BiLSTM-CNN (ES + EB),* and *E-BiLSTM-CNN (D + ES + EB).*

To visualize the process, think of a simple tweet consisting of three words and two emojis, say, "I love summer ☼😎". Here, 'I', 'love', and 'summer' are our word tokens, and '☼' and '😎' are our emoji tokens. Each of these is passed through the Word_Emoji embedding layer. Here, it is converted into a vector using the embedding matrix, which is created using the GloVe and Emoji2Vec embedding dictionaries. For each word and emoji token, the corresponding embedding vector is found in the relevant embedding dictionary. Now our tweet, "I love summer ☺😎", is represented as a sequence of vectors. Each word and emoji is now not just a numerical value, but a vector in high-dimensional space, containing rich information about its meaning. The sequence of vectors is then passed to a Bi-LSTM layer, CNN Layer, to obtain a tweet representation. The model then uses the tweet representation, along with the average emoji score, to compute the sentiment polarity of the tweet through the softmax activation function in the output layer.

4.2.3. Explainable multi-view sentiment analysis

Explainable AI (XAI) is particularly important in the context of sentiment analysis for high-stakes decision forecasting. While sentiment analysis can provide valuable insights into customer attitudes and behaviors, it is essential to understand the reasoning behind these predictions. However, sentiment analysis based on machine learning algorithms is one of the "black boxes" (Leung et al., 2021; Bussmann et al., 2021), which lacks transparency and interpretability (Zytek et al., 2021; Shin, 2021). As a result, decision-makers may be hesitant to act on its predictions. In addition, in high-stakes decision forecasting, the consequences of undetected incorrect predictions can be severe. For example, customer sentiment can have a significant impact on the success or failure of a product or service, and undetected inaccurate forecasting may lead to a decline in sales or even damage to the brand's reputation or waste resources on unnecessary product improvements or marketing campaigns.

Therefore, advanced artificial intelligent models must be transparent and interpretable. In order to realize this aim, Explainable Artificial Intelligence (XAI) methods offer explanations that make the functioning of AI comprehensible (Haque et al., 2023). This study will employ one of the most popular XAI methods, LIME, to visualize and explain the prediction results of the proposed multi-view sentiment analysis model. This will support high-stakes decision-making related to marketing or customer preference forecasting. Specifically, the *E*-BiLSTM-CNN model will serve as the baseline for LIME to demonstrate its interpretability and trustworthiness to end-users. The integration process involves several critical steps to enhance the model's interpretability. Initially, we initialize our model with pre-trained weights, setting it to evaluation mode for inference purposes. A specific inference method is then crafted to process text strings by tokenizing them with the same tokenizer used during training and padding them to a uniform length before feeding them into the E-BiLSTM-CNN architecture to compute output probabilities for different sentiment classes. Subsequently, we integrate LIME using the LimeTextExplainer, which requires class names and a splitter function that aligns with our tokenizer. To generate explanations for a specific instance, we utilize the LIME explainer's explain_instance method. Our model passes a sample text to this method, enabling LIME to produce explanations that highlight the most influential features (words or emojis) and their corresponding weights in the model's decision-making process. These explanations are visualized in a bar chart, offering a clear representation of the significance of each feature in the model's predictions. The algorithm of the sentiment explainer is shown in Algorithm 4 and the visualized process is shown in Fig. 4.

Algorithm 4. Location explanation plotting algorithm for multi-view sentiment analysis.

4.3. Features and embeddings

Both emojis and texts are employed as features. For all the algorithms, the words in each sentence are converted into 300-dimension word embeddings using Global Vectors for Word Representation (GloVe). It is an unsupervised learning algorithm that generates vector representations of words, which are trained over global word-word co-occurrence statistics (Pennington et al., 2014). As discussed in the literature review, emojis contain useful information that is related to the sentiment of a text. In this study, three methods were employed to handle the emojis, including emoji replacement, adding emoji scores and adding emoji embeddings. A detailed description of these methods is presented as follows.

4.3.1. Emoji replacement

This method employs the emoji package[2] to replace each emoji with its corresponding words or phrases. Then, the words or phrases become a part of the tweet and then are entered into the next step of word embedding.

4.3.2. Adding emoji scores

This approach computes the average "emoji score" of each tweet. Based on the calculation method provided by (Hankamer and Liedtka, 2016), an emoji score is calculated by taking the number of its positive occurrences, subtracting the number of its negative occurrences, and then dividing by the number of total occurrences (including neutral occurrences). As some tweets contain more than one emoji, for each tweet, this method takes the average score of the emoji appearing in the tweet and uses that score as an additional feature of the tweet. The pseudocode is shown in Algorithm 1. The occurrence information comes from the emoji sentiment lexicon provided by (Kralj Novak et al., 2015).[3] This lexicon contains occurrence information about 751 emoji characters.

$$\mathrm{es}_{\mathrm{eb}} = (N(e_{b+}) - N(e_{b-}))/N(e_b) \tag{13}$$

$$es_{\mathrm{i}} = \left(\sum_{1}^{b} \mathrm{es}_{\mathrm{eb}}\right) \Big/ b \tag{14}$$

Algorithm 1. Computing emoji scores of reviews.

4.3.3. Creating emoji embeddings

Unlike emoji scores, this method applies an embedding method to emoji and generates emoji representations directly. This study used the emoji2vec embedding approach provided by (Eisner et al., 2016), which includes 1662 emojis. This emoji embedding is pre-trained by the emoji's description in the Unicode emoji standard through the use of the Word2Vec embedding method. In the present study, the GloVe word embedding matrix was combined with the Emoji2Vec emoji embedding matrix to create a new Word_Emoji embedding (as described in Algorithm 2). Then, it was used as the weight in the model. As a result, the model is able to extract different emotional information from the tweets and then used to learn their relationship with users' sentiments (as discussed in Section 4.2.2 and outlined in Algorithm 3). In addition, this approach requires minimal preprocessing of the text as it does not require the removal of emojis or the calculation of emoji scores to add features.

Algorithm 2. Creation of the word_emoji_embedding.

Algorithm 3. Sentiment polarity prediction model.

4.4. Evaluation metrics

Accuracy and F1-score are the two most frequently used performance evaluation metrics in published studies. Accuracy is helpful because it helps us compute the number of correct predictions a model makes, but it does not take into account how the data is distributed. If most instances belong to the majority class, the accuracy score may be high

[2] https://pypi.org/project/emoji/
[3] https://www.kaggle.com/datasets/thomasseleck/emoji-sentiment-data

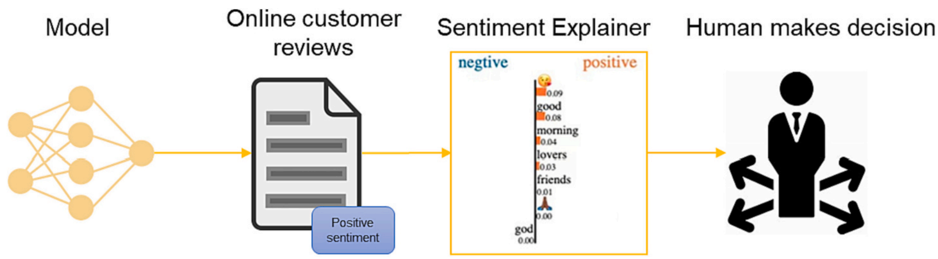


Fig. 4. Sentiment explainer.

Input: a processed headline text sample, and multi-view sentiment analysis pipeline
Output: A plot of the location explanation for the given sample

1. Instantiate the LimeTextExplainer to explain how the sentiment analysis model made its prediction for the headline
2. Use explain_instance() with appropriate parameters (text, predict_proba, num_features) to generate an explanation for the given text sample:
3. Extract the ordered dictionary of words and weights from the explanation
4. Plot a bar figure for the location explanation for the given sample
5. Output the location explanation figure

Input: Tweet T_i ={w$_1$, w$_2$, ..., w$_a$, e$_1$, e$_2$, ..., e$_b$}
Output: emoji score of each review (es_i)

1. Data preprocessing to obtain a cleaned version clean_ T_i
2. Extract emoji sequence EM for each clean_ T_i
3. **for** each emoji sequence**, do**
4. Obtain emoji scores for each emoji
   - emoji_scores = []
   - for emoji in EM:
     - es = compute_emoji_score(emoji) by $es_{e_b} = (N(e_{b+}) - N(e_{b-}))/N(e_b)$
     - emoji_scores.append(es)
   - end for
   Compute emoji score for the tweet
   - avg_emoji_score $(es_i) = (\sum_1^b es_{e_b})/b$
5. **end for**
6. Output the emoji score of each review (es_i)

---

even though it doesn't distinguish the classes very well. F1-Score accounts for both precision and sensitivity, it compensates for uneven class distribution in the training dataset (Chicco and Jurman, 2020). In this study, the dataset is class balanced, the accuracy score and F1-score are therefore both suitable for evaluating the classifiers. The following equations are the formulas of the metrics:

$$\text{Accuracy} = (\text{TP} + TN)/(\text{TP} + TN + \text{FP} + \text{FN}) \tag{15}$$

$$F1 - Score = 2^*(Recall^*Precision)/(Recall + Precision) \tag{16}$$

$$Precision = TP/(TP + FP) \tag{17}$$

$$Recall = TP/(TP + FN) \tag{18}$$

where TP is True Positive, refers to the sample size of positive labels correctly classified by the model. TN is true negatives and refers to the sample size of negative labels correctly classified by the model. FP is false positives and refers to the sample size of positive labels incorrectly classified by the model. FN is false negatives and refers to the sample size of negative labels incorrectly classified by the model. F1-Score is the weighted average of Recall and Precision.

As discussed in the previous study of the authors, many models in existing sentiment classification research have achieved high accuracy rates, F1-scores, or other statistical evaluation metrics. However, few studies have assessed these models from the practical perspective, e.g. execution time, which is critical for addressing real-world issues (Das et al., 2018). Therefore, the execution time is also employed to be the evaluation metric in this study. In addition, this study will compute a comprehensive score based on the F1-score and execution time for each classifier to evaluate their overall performance.

$$Final\ score = 0.6^*F1 - score + 0.4^*execution\ time \tag{19}$$

5. Results and discussion

5.1. The effect of emoji features on the performance of sentiment classifiers

Firstly, this study evaluates the effectiveness of handling emoji on sentiment recognition of online reviews by sentiment analyzers and the best handling method for each algorithm. The results are summarized in the following figures and tables. The figures show the scores of the evaluation metrics for each method, including accuracy, F1-score and execution time. Their improvement or reduction in each metric

Input: GloVe and Emoji2Vec embedding dictionaries, the dimensionality of embeddings, tweets
Output: word_emoji_embedding matrix and padded sequences

```
1.  Tokenize the tweets using a Tokenizer and build the vocabulary.
    -tokenizer = Tokenizer()
    -tokenizer.fit_on_texts(clean_T)
    -sequences = tokenizer.texts_to_sequences(clean_T)
2.  Pad the tokenized tweets to a fixed length using a padding function.
    -padded_sequences = pad_sequences(sequences, maxlen=max_tweet_length, padding='post')
3   Load the GloVe and Emoji2Vec embedding dictionaries.
4.  Define the dimensionality of the embeddings (e.g. 300)
5.  Initialize the embedding matrix with random values.
    -num_words = len(tokenizer.word_index) + 1
    -embedding_matrix = np.random.random((num_words, embedding_dim))
6.  for each token in the vocabulary, do
7.    Check if the token exists in the GloVe embedding dictionary or the Emoji2Vec embedding
8.  dictionary.
9.    If it does, use the corresponding embedding for this token in the embedding matrix.
      If it doesn't, check if the token exists in the Emoji2Vec embedding dictionary.
      -for word, i in tokenizer.word_index.items():
         if word in glove_embeddings:
           embedding_matrix[i] = glove_embeddings[word]
         elif word in emoji_embeddings:
10.        embedding_matrix[i] = emoji_embeddings[word]
11. end for
12. Output the embedding matrix and padded sequences.
```

Input: word_emoji_embedding matrix and padded sequences, emoji scores
Output: The sentiment polarity of each review polarity (T_i)

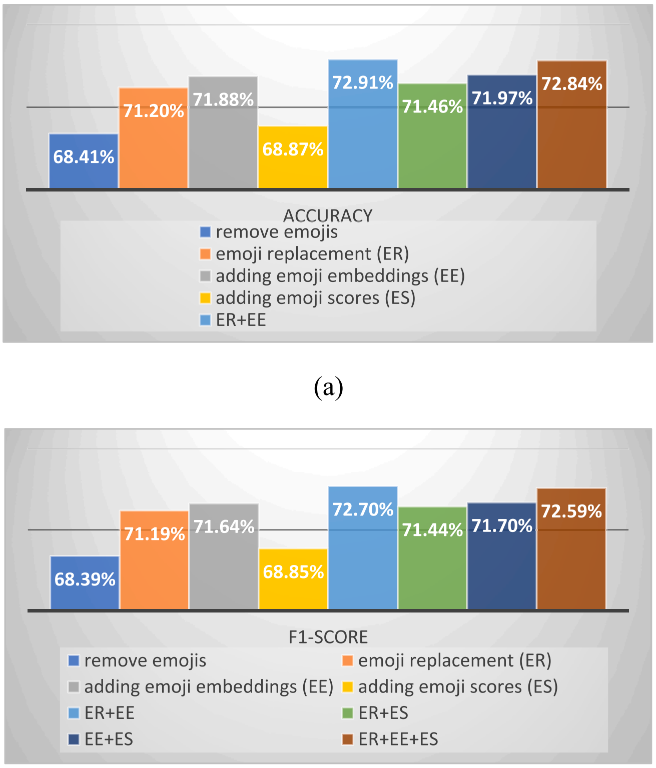
1. **for** each sequence, **do**
2. The element (word and emoji) vector xt, is obtained by transforming the element e_i to vector x_t through the word_emoji embedding matrix W: $x_t = W_e e_t$
3. h_i = BiLSTM(x_i): pass x to BiLSTM layer and obtain the element vector $h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}]$
4. a_i = Conv1D (h_i); pass h to 1D convolutional layer to extract higher-order features $a_i = f\left(F \circ h_{i:i+q-1} + b\right)$ and obtain a feature map $c = [a_1, a_2, \ldots, a_{n-l+1}]$
5. the maximum value $m = \max\{c\}$ is extracted by applying the max pooling operation to each feature map
6. Add emoji score feature to the features extracted by the CNN layer: $V_i = [m_i, es_i]$
7. The dropout rate of 0.3 to deal with the overfitting issue
8. **end for**
9. Using sigmoid activation function to compute the binary classification distribution p
10. Predict sentiment polarity: polarity (T_i) = (p > best_thresh).astype
11. Output the sentiment polarity of each review polarity (T_i)

compared to only considering word features in models and their rankings in each metric and the overall ranking are also listed in the tables. In addition, Emoji_less refers to using the method of removing emojis from the text, ER refers to emoji replacement, EE refers to creating emoji embeddings, and ES refers to adding emoji scores.

For Naive Bayes, handling emojis using any of the three approaches helped to increase the classifier's performance in sentiment recognition (Fig. 5 and Table 1). However, the effectiveness of the method, including employing emoji scores as an additional feature, is not significant. According to Table 1, it only achieved an accuracy/F1-score 0.67 % higher than only word features were considered. From the perspectives of accuracy and F1-score, the best emoji handling method for NB is using both emoji replacement and adding emoji embedding methods (6.58 % higher than EMOJI_LESS), while the method of removing emojis from the text takes the shortest execution time (Fig. 6). In order to evaluate the classifiers' comprehensive performance, this study considers the ranking of each classifier in the F1-score and the execution time evaluation metric. The findings confirm that the smaller the rank score, the higher the ranking and the better the classifier performs. Stacked bars are employed to visualize this ranking for decision-makers to better understand the outcomes; the shorter the bar, the better the corresponding classifier performs. According to Fig. 7, NB performs best when using both replacing emoji and adding emoji embedding methods, as it achieves the smallest ranking score.

For SVM, handling emojis using the emoji embedding method slightly improved the classifier's performance by 0.07 % accuracy compared to the Emoji_Less method, while replacing emojis with their description and adding an emoji score feature improved the performance of classifiers by 4.47 % and 3.82 % respectively (Table 2). From the perspectives of accuracy and F1-score, using emoji descriptions and emoji scores simultaneously could improve the classifier's overall performance in detecting the sentiment of tweets to a greater extent (Fig. 8). From the perspective of execution time, the best handling method is the emoji replacement method alone (Fig. 9). Suppose the performance of the classifiers in these two areas is considered together. In that case, the classifier employing SVM to conduct sentiment classification performs best when using the combination of emoji replacement and emoji scores is the best (Fig. 10).

Regarding Logistic Regression, applying the emoji embedding method slightly deteriorates performance compared to just considering



(a)

(b)

Fig. 5. Accuracy (a) and F1-score (b) of different classifiers using Naïve Bayes.

word features, while the methods of emoji replacement and adding emoji scores improved the performance of classifiers by 4.34 % and 3.72 %, respectively (Table 3). The results are similar to those when SVM is used. It indicates that SVM and LR are not able to derive meaningful information from emoji embeddings. In terms of accuracy and F1-score, the best handling method is using all of the three methods together (Fig. 11). From the perspective of execution time, the best handling method is employing emoji scores as an additional feature (Fig. 12). Taking two aspects into consideration, either emoji replacement (F1-score: 77.12 %; Time vs. Emoji_less: +0.4 %), adding emoji scores (F1-score: 76.66 %; Time vs. Emoji_less: −0.48 %), or use both of them (F1-score: 77.71 %; Time vs. Emoji_less: +0.57 %) can be chosen when using LR to construct the sentiment classifier (Fig. 13), as they achieved Accuracy and F1-score around 77 %, and spent execution time close to just removing all emojis. Although the classifier achieved the best result in accuracy when using the combination of the three methods (F1-score: 77.73 %), it took nearly four times (Time vs. Emoji_less: +428.62 %) as the time spent by other methods. Therefore, it was not the best choice for practical use when using LR to construct a sentiment classifier.

When using the BiLSTM-CNN model, even with emoji removed, its ability to recognize sentiment is comparable to any classical machine learning algorithm that takes emoji into account. Any of the three emoji handling approaches helped to increase the performance of classifiers and can contribute to inform decisions. Emoji replacement can make an improvement of 6.01 % in accuracy and f-score, adding emoji embeddings can improve by 5.11 %, and adding emoji scores can improve by 3.56 % (Table 4). From the results of using a combination of emoji processing methods, it appears that while emoji scoring and emoji embedding are effective methods, neither of them provides additional useful information when in a situation where emoji replace is already in use. Therefore, replacing emojis is the best handling method out of all combinations according to accuracy and F1-score values (Fig. 14). For

Table 1
Comparison and rankings of different classifiers using Naïve Bayes.

NB	Variation in accuracy (%)	Accuracy_ranking	Variation in F1_score (%)	F1_score_ranking	Variation in Time(%)	Time_ranking
EMOJI_LESS	0.00 %	8.00	0.00 %	8.00	**0.00 %**	**1.00**
Emoji replacement (ER)	4.08 %	6.00	4.09 %	6.00	0.86 %	2.00
Adding emoji embeddings (EE)	5.07 %	4.00	4.75 %	4.00	66.71 %	4.00
Adding emoji scores (ES)	0.67 %	7.00	0.67 %	7.00	62.33 %	3.00
ER + EE	**6.58 %**	**1.00**	**6.30 %**	**1.00**	143.17 %	6.00
ER + ES	4.46 %	5.00	4.46 %	5.00	139.42 %	5.00
EE + ES	5.20 %	3.00	4.84 %	3.00	225.08 %	7.00
ER + EE + ES	6.48 %	2.00	6.14 %	2.00	233.28 %	8.00

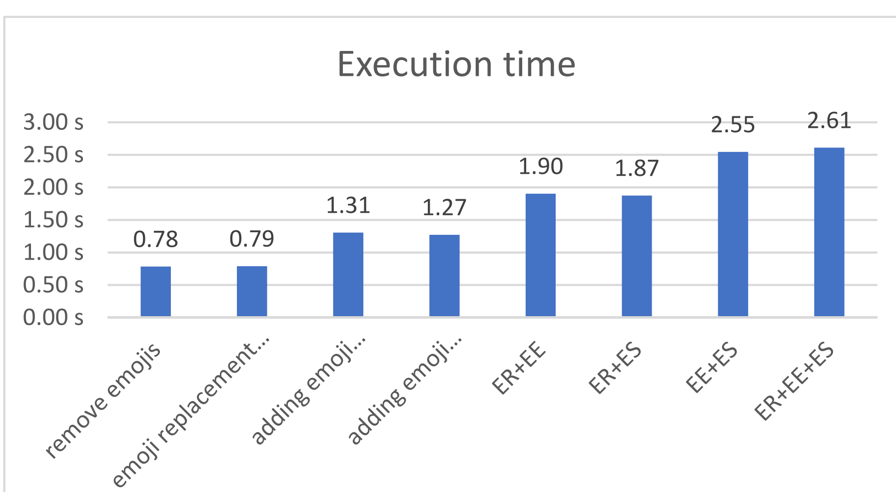


Fig. 6. Execution time of different classifiers using Naïve Bayes.

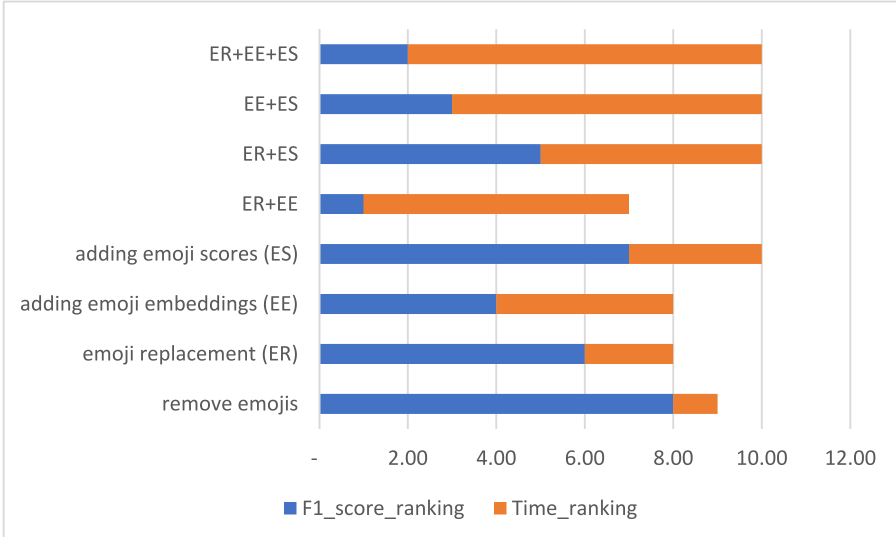


Fig. 7. The comprehensive ranking of different classifiers using Naïve Bayes.

Table 2
Comparison and rankings of different classifiers using Support Vector Machine.

	Variation in accuracy (%)	Accuracy_ranking	Variation in F1_score (%)	F1_score_ranking	Variation in Time(%)	Time_ranking
Remove emojis	0.00 %	8.00	0.00 %	8.00	0.00 %	3.00
Emoji replacement (ER)	4.47 %	4.00	4.47 %	4.00	**−2.60 %**	**1.00**
Adding emoji embeddings (EE)	0.07 %	7.00	0.07 %	7.00	70.16 %	8.00
Adding emoji scores (ES)	3.82 %	5.00	3.82 %	5.00	10.11 %	4.00
ER + EE	4.49 %	3.00	4.49 %	3.00	62.17 %	6.00
ER + ES	**5.50 %**	**1.00**	**5.50 %**	**1.00**	−0.67 %	2.00
EE + ES	3.42 %	6.00	3.42 %	6.00	64.26 %	7.00
ER + EE + ES	5.46 %	2.00	5.46 %	2.00	52.58 %	5.00

execution time, the best handling method is also the emoji replacement. The BiLSTM-CNN model performs best when using emoji replacement while considering the performance of the classifiers as a whole (Figs. 15 and 16).

5.2. The effectiveness of the word_emoji embedding matrix

The *E*-BiLSTM-CNN model this study proposed creates emoji features in a new method, which converts words and emojis simultaneously based on a new word_emoji embedding matrix. With the purpose of testing the effectiveness of the proposed model, this study compared its performance with other classifiers. As shown in Fig. 14 and Table 4, this method significantly enhanced the BiLSTM-CNN model's effectiveness (accuracy: 81.44 %; F1-score: 81.43 %; execution time: 448 s) by 5.11 % of accuracy/F1-score compared to the model using data samples of plain text (accuracy: 77.48 %; F1-score: 77.47 %; execution time: 307 s) or by 1.50 % of accuracy/F1-score compared to the model adding an additional feature of emoji scores (accuracy & F1-score: 80.24 %; execution time: 327 s). However, in terms of the time taken to train the BiLSTM-CNN model, the classifiers using this method took a longer time than others, which makes it fail to be the best method. The results that show the method "emoji replacement" shows better classification performance than adding emoji embeddings agree with A. Singh et al. (2019). The possible reason is that there are a large number of emojis (over 2800), some of which do not appear very often. Existing research has, therefore, focused on creating only the most frequently used emoji lexicon to provide emoji scores or pre-trained emoji embeddings, which is incomplete. However, words in their descriptions are much more common, so it is often more beneficial to utilize descriptions for sentiment analysis on current social networking platforms.

From the perspective of practical use for decision-making, the proposed method is more efficient when identifying the sentiment of new text than those provided by the existing literature (A. Singh et al., 2019; de Barros et al., 2021). It embeds words and emojis in each tweet at the same time rather than creating word embeddings and emoji embeddings separately and then combining them. Therefore, this technique has the advantage of requiring minimum preprocessing of the text as it does not require removing or separating emojis or computing emoji scores to add features.

5.3. Comprehensive performance comparison among best classifiers for each algorithm

This study also compared the performance of the best classifiers using each algorithm. Although naive Bayes was the fastest, it only achieved nearly 72.91 % accuracy, which was 5 % lower than the other classifiers. The most accurate classifier was BiLSTM-CNN using emoji replacement, achieving 82.14 % accuracy and F1-score, but it took a much longer time (15,400 times longer than NB(ER + EE)) due to the nature of deep learning (Figs. 17 and 18). This study performed a weighted average of their performance based on the F1-score and execution time, and the best classifier was Bi-LSTM using emoji replacement. Compared to the baseline models, the deep learning model can extract more meaningful information from emoji characteristics due to its powerful feature extraction capacity (Fig. 19).

5.4. Results of Explainable Multi-view Sentiment Analysis by LIME

For the purpose of improving the trust of decision-makers for the proposed multi-view deep learning sentiment analysis model, LIME is employed to help understand which features the model picks to make predictions. In addition, LIME is a local interpretation tool, which means

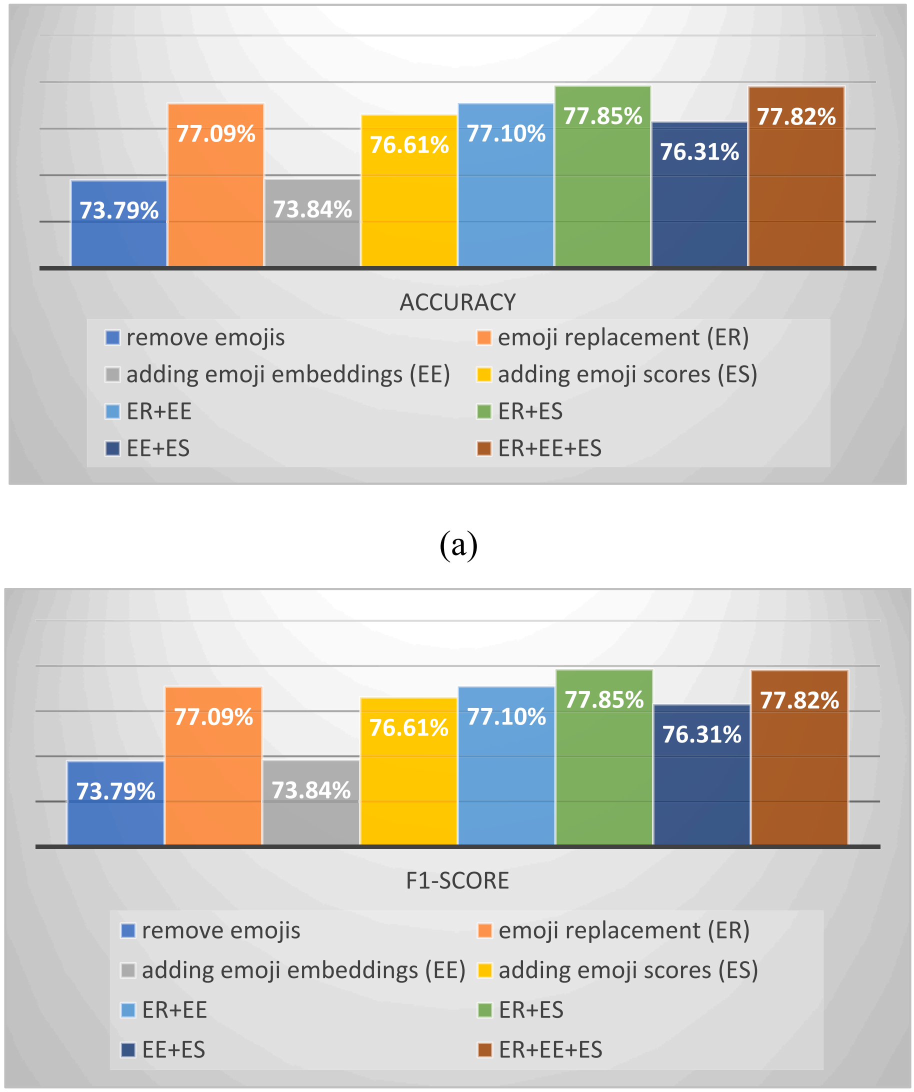


Fig. 8. Accuracy (a) and F1-score (b) of different classifiers using Support Vector Machine.

it is able to explain a specific instance according to the requirements of decision-makers.

Fig. 20 presents the local explanations of the proposed multi-view sentiment analysis model for three specific online review samples by the LIME technique. The sentiment predictions for these reviews are illustrated with their corresponding prediction probabilities. For instance, the first review is predicted as positive with complete certainty, as indicated by a prediction probability of 100 %. Conversely, the second review, focusing on the cost of living, is predicted as predominantly negative with a prediction probability of 0.91. The third review, discussing a flu vaccine, is predicted as predominantly positive with a prediction probability of 0.79. To improve the comprehension of the black box method, LIME is used to visualize the features on which the prediction is based. Two ways have been provided for decision-makers to reference. The first way is by the degree of color, which is shown in the 'Text with highlighted words' section of Fig. 20. The deeper the color, the more significant the feature. The second way is clearer, which is shown in the 'Prediction probabilities' section of Fig. 20. It uses a bar chart to rank the features in descending order according to their significance value, labeled on the chart. The features located on the right of the line are indicative of positive sentiment, whereas those on the left side represent negative sentiment. According to Fig. 20, the three most significant factors in the first review determined by the proposed model for the prediction on the given review are the '😊', 'good' and 'morning', which indicate positive sentiment.

5.5. Validation and ablation test

To assess the performance of the proposed *E*-BiLSTM-CNN model, this study compared it with other studies based on the F1-score and accuracy metrics, as these are the most commonly used and were available in the referenced papers. In addition, the performance of the proposed model was also compared to its building blocks, including LSTM, BiLSTM, and CNN. The following table is a summary of the findings (Table 5).

The proposed E-BiLSTM-CNN model (Model 4) demonstrated competitive results in terms of both F1-score and accuracy. Compared to its building blocks as (Model 1, Model 2 and Model 3), the proposed model outperformed with the highest accuracy and F1-score values, indicating the effectiveness of the integrated approach. The combination of LSTM, BiLSTM, and CNN components in the E-BiLSTM-CNN model synergistically enhances its ability to accurately interpret and classify

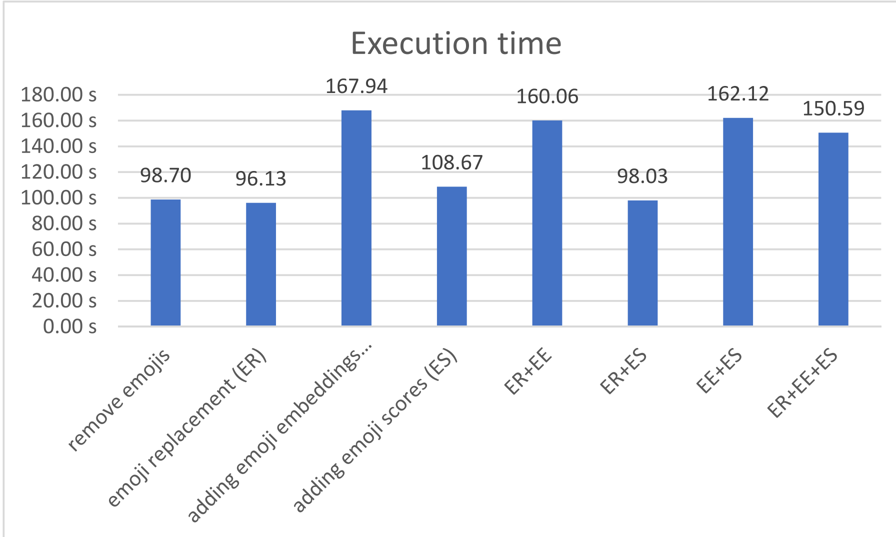


Fig. 9. Execution time of different classifiers using Support Vector Machine.

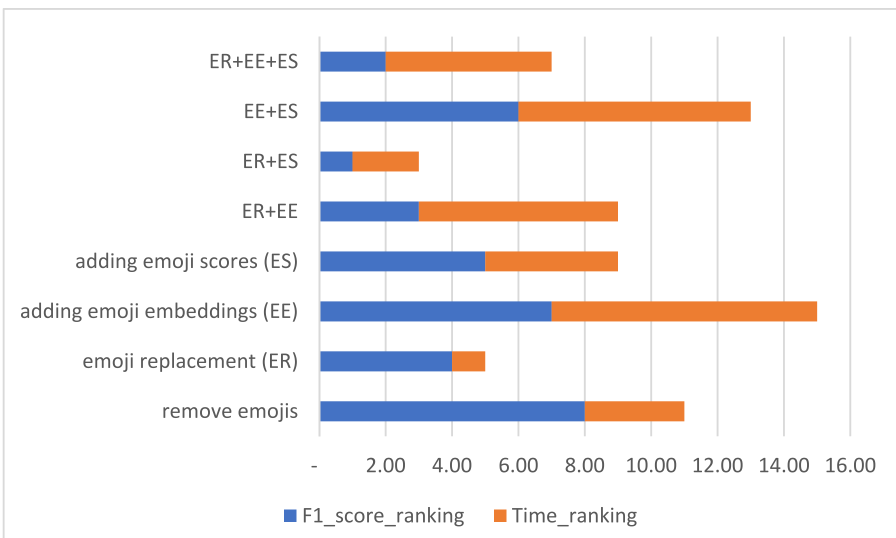


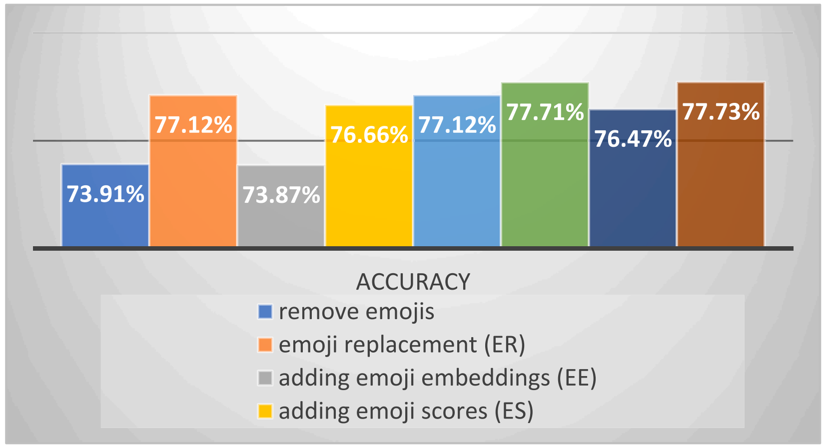
Fig. 10. The comprehensive ranking of different classifiers using a Support Vector Machine.

Table 3
Comparison and rankings of different classifiers using Logistic Regression.

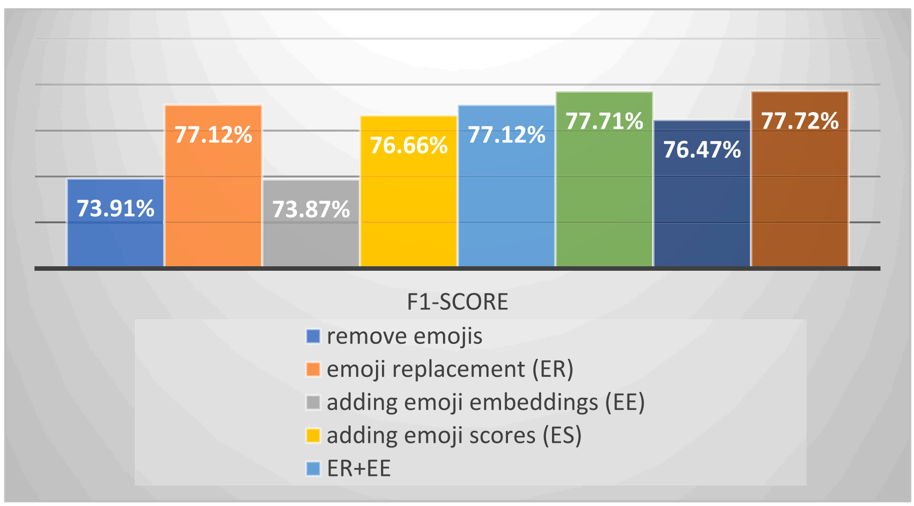
	Variation in accuracy (%)	Accuracy_ranking	Variation in F1_score (%)	F1_score_ranking	Variation in Time(%)	Time_ranking
Remove emojis	0.00 %	7.00	0.00 %	7.00	0.00 %	2.00
Emoji replacement (ER)	4.34 %	3.00	4.34 %	3.00	**0.40 %**	3.00
Adding emoji embeddings (EE)	−0.05 %	8.00	−0.05 %	8.00	408.56 %	6.00
Adding emoji scores (ES)	3.72 %	5.00	3.72 %	5.00	**−0.48 %**	**1.00**
ER + EE	4.34 %	3.00	4.34 %	3.00	414.48 %	7.00
ER + ES	5.14 %	2.00	5.14 %	2.00	**0.57 %**	4.00
EE + ES	3.46 %	6.00	3.46 %	6.00	329.38 %	5.00
ER + EE + ES	**5.17 %**	**1.00**	**5.15 %**	**1.00**	428.62 %	8.00

sentiment from social media text, including nuanced expressions conveyed through emojis. Despite the high accuracy of Lou et al.'s (2020) EA-Bi-LSTM model, the proposed model had a significantly higher F1-score, demonstrating a better balance between precision and recall. Moreover, the model outperformed Singh et al.'s (2019) EMJ-DESC model and de Barros et al.'s (2021) pre-trained BERT model (TweetSentBR version) in both respects.

When compared to the best-performing model from de Barros et al. (2021), the pre-trained BERT model-2000-tweets-BR, the F1 scores of the proposed model are almost comparable and only slightly less accurate. However, it is important to note that the proposed model was trained on a dataset with 80,000 tweets, much larger than the 2000-tweet dataset used in the pre-trained BERT models by de Barros et al. (2021). Despite the lack of F1-score for comparison with Liu et al.



(a)



(b)

Fig. 11. Accuracy (a) and F1-score (b) of different classifiers using Logistic Regression.

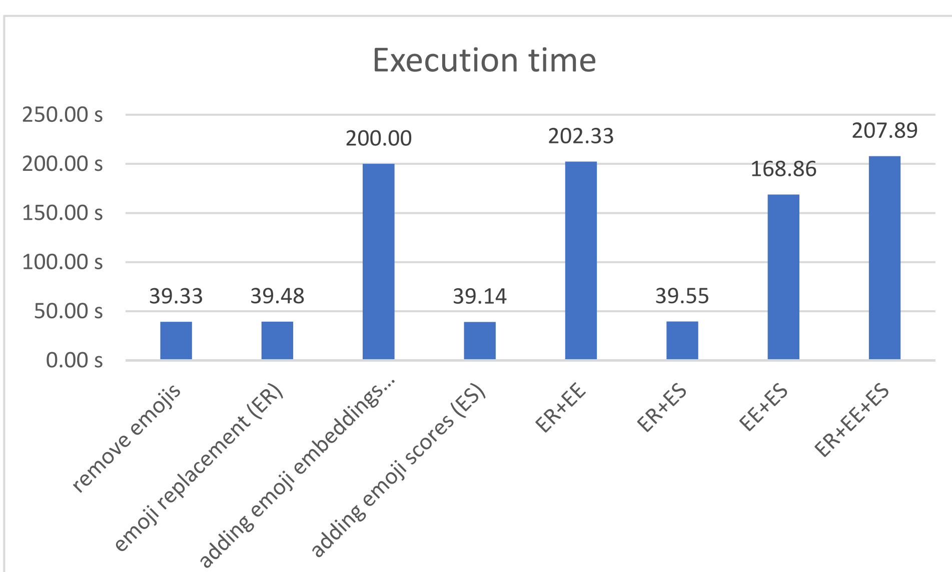


Fig. 12. Execution time of different classifiers using Logistic Regression.

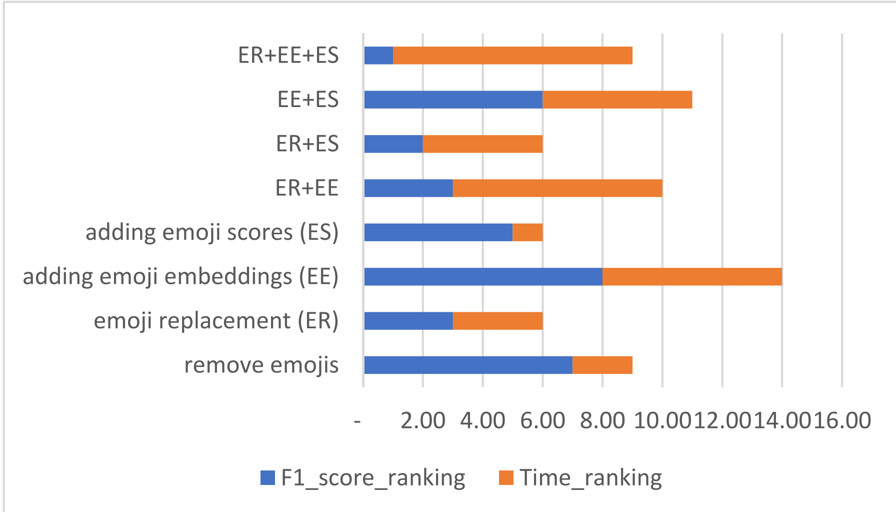


Fig. 13. The comprehensive ranking of different classifiers using Logistic Regression.

Table 4
Comparison and rankings of different classifiers using BiLSTM-CNN.

	Variation in accuracy (%)	Accuracy_ranking	Variation in F1_score (%)	F1_score_ranking	Variation in Time(%)	Time_ranking
Remove emojis	0.00 %	8.00	0.00 %	8.00	**0.00 %**	**3.00**
Emoji replacement (ER)	**6.01 %**	**1.00**	**6.03 %**	**1.00**	**−3.29 %**	**1.00**
Adding emoji embeddings (EE)	5.11 %	6.00	5.11 %	6.00	46.04 %	7.00
Adding emoji scores (ES)	3.56 %	7.00	3.58 %	7.00	6.63 %	4.00
ER + EE	5.18 %	4.00	5.19 %	4.00	45.42 %	6.00
ER + ES	5.76 %	2.00	5.77 %	2.00	−2.77 %	2.00
EE + ES	5.14 %	5.00	5.14 %	5.00	36.28 %	5.00
ER + EE + ES	5.20 %	3.00	5.21 %	3.00	46.46 %	8.00

(2021), it is found the proposed model's accuracy is similar.

Together with the benefits of our simplified preprocessing pipeline, the use of a realistic emoji proportion dataset, and the application of Explainable AI techniques, these results underscore the robustness and validity of our proposed *E*-BiLSTM-CNN model for sentiment analysis. Moreover, the larger dataset used in this study further contributes to the robustness and generalizability of the results.

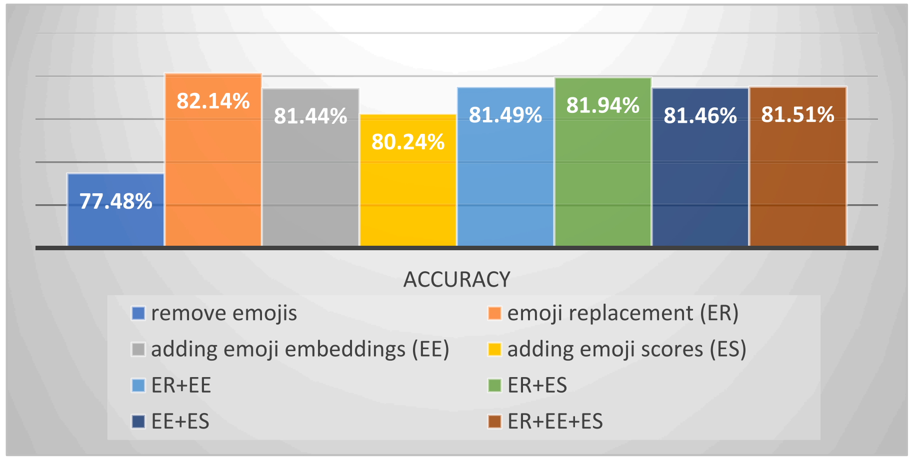
According to the various discussions above, the findings of this study first contribute to the theoretical understanding of how emojis and text interact in sentiment analysis. The proposed E-BiLSTM-CNN model, which incorporates both features in a balanced manner, addresses the limitations of previous models that ignore the sentiment information contained by emojis features or require intensive preprocessing. From an empirical standpoint, the model has demonstrated superior performance when compared to other models. With a competitive F1-score and accuracy, even when trained on a larger, more representative dataset, the E-BiLSTM-CNN model proves to be an effective tool for sentiment analysis. This success points to a significant advancement in the practical application of sentiment analysis models in social media contexts. In terms of marginal economic effect, the results of this study could significantly impact sectors that rely heavily on social media data. By applying our more accurate and efficient model, industries and governments can gain more precise insights into consumer sentiment. With the ability owned by the model to handle large datasets and maintain performance, they can analyze larger amounts of data in less time, leading to cost savings. Moreover, by not requiring additional preprocessing steps, resources can be allocated more efficiently, increasing the marginal returns of sentiment analysis.

In addition, understanding the sentiment of public opinion is crucial in managing market disasters caused by unforeseen circumstances, like unexpected regulations (U-R conflicts) or the COVID-19 pandemic. The *E*-BiLSTM-CNN model in this paper can also assist in such situations. First of all, the proposed model's superior performance in sentiment analysis can aid in the early detection of shifts in public sentiment. For instance, escalating public discontent due to sudden regulatory changes or public fears during the COVID-19 pandemic can be detected early by analyzing social media data. This allows policymakers, businesses, and other stakeholders to respond proactively and avert potential crises. Second, by understanding the prevalent sentiments in real-time, businesses and governments can tailor their communication strategies to address better public concerns, fears, or expectations to mitigate miscommunications or misunderstandings. Third, the proposed model can provide valuable feedback on the effectiveness of recovery efforts and allow adjustments to be made quickly.

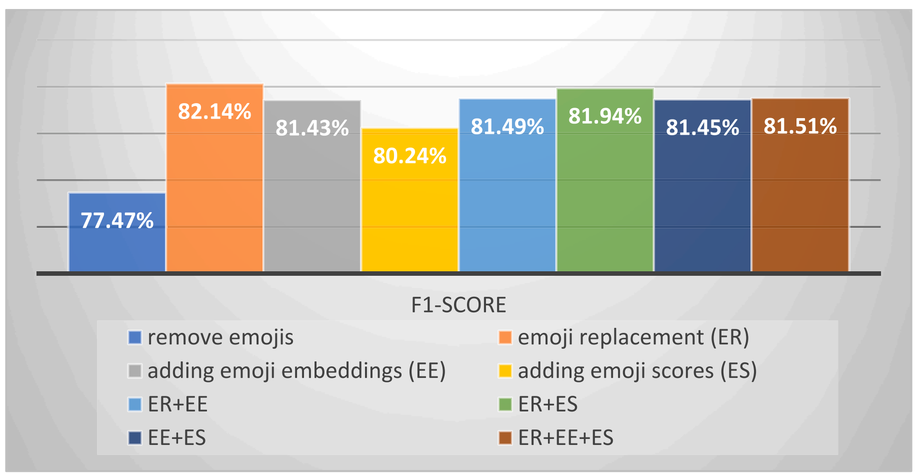
6. Conclusion

6.1. Main findings and contributions

From a multi-view learning perspective, this paper investigates the impact of emojis on identifying sentiments of posts users expressed on social media platforms. This study proposed three emoji handling methods, namely, Emoji Replacement, Adding Emoji Scores, and Creating Emoji Embeddings, and tested how well each sentiment classifier performs when incorporating emoji features processed by these methods individually or in combination. Three classical ML algorithms were employed to construct the baseline classifiers. Moreover, a novel multi-view deep learning model, *E*-BiLSTM-CNN, was also proposed and compared to the other classifiers. The main finding is that each sentiment classifier improves the performance of the classifiers when dealing with emoji features processed by the three methods, either individually or in combination. These results validate that text and emoji features can be used as different views to provide different sentiment information to the sentiment classification model. The performance of the Word_Emoji



(a)



(b)

Fig. 14. Accuracy (a) and F1-score (b) of different classifiers using BiLSTM-CNN.

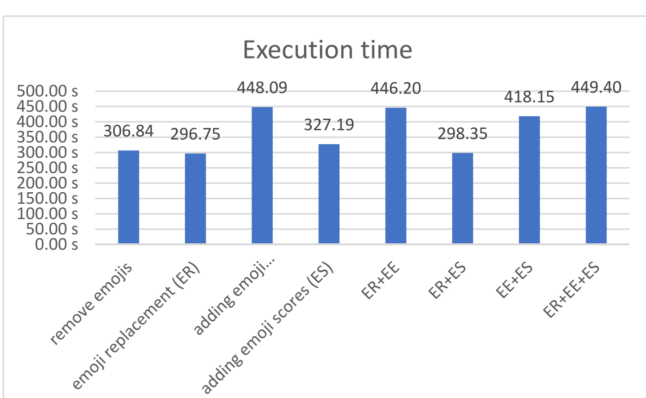


Fig. 15. Execution time of different classifiers using BiLSTM-CNN.

embedding matrix, which was implemented in the proposed E-BiLSTM-CNN model, was also evaluated, demonstrating notable effectiveness with a high F1 score of 81.4 %.

This research extends the understanding of sentiment analysis by proposing a multi-view learning approach that regards text and emojis as distinct, valuable sources of sentiment information. A significant contribution is the introduction of explainable sentiment analysis to this multi-view model. By utilizing explainable sentiment analysis, decision-makers can comprehend how the model develops its decisions and which features are deemed significant by the model. This enables them to evaluate the prediction themselves, combining their own experience to make the final decision, which can mitigate the influence of misleading decision forecasting on high-stakes businesses.

In addition to the effectiveness of considering text and emojis features in deep learning sentiment classification and providing explainable sentiment analysis, the current research has made several other contributions. The proposed multi-view sentiment analysis method is constructed by simulating the real distribution of emojis on the social media platform, which considers the issue of consistency between the dataset used and reality. Moreover, this study considered the efficiency of classifiers essential when applied in the real business world. The proposed application framework (Fig. 1) requires minimal preprocessing of social media posts, which ensures the system's efficiency and allows it to process large volumes of data in a timely and accurate manner. This streamlined approach to preprocessing significantly reduces the risk of errors and inaccuracies, allowing high-stakes businesses to make well-informed decisions based on reliable and accurate sentiment analysis.

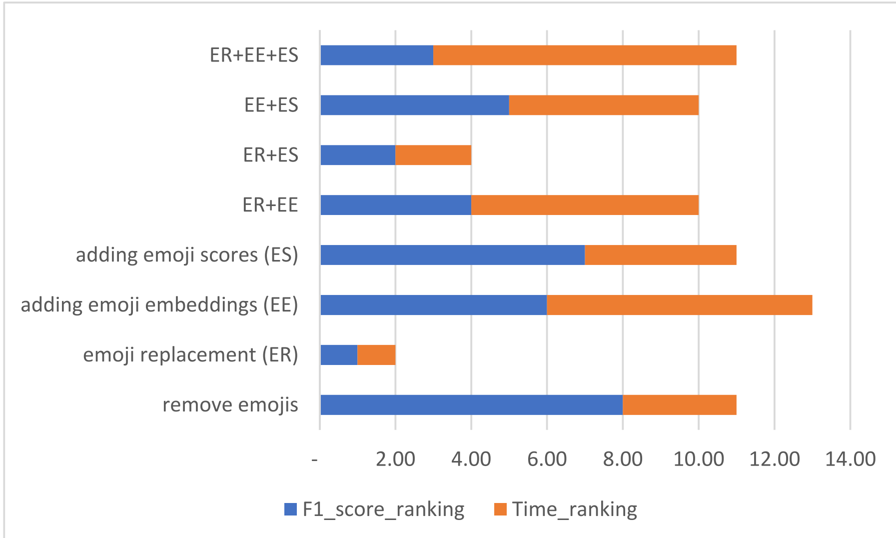


Fig. 16. The comprehensive ranking of different classifiers using BiLSTM-CNN

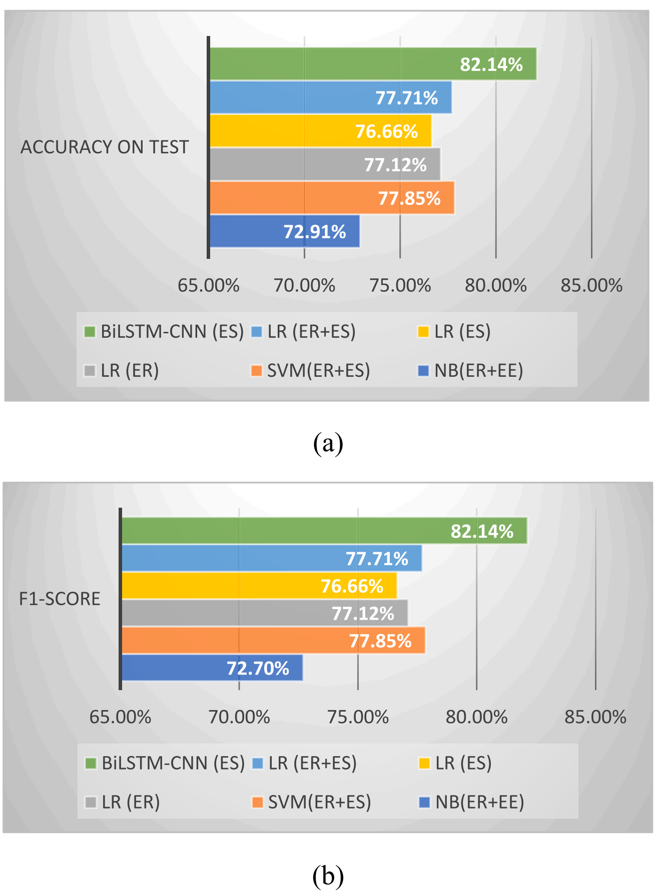


Fig. 17. Accuracy (a) and F1-score (b) of best classifiers for each algorithm.

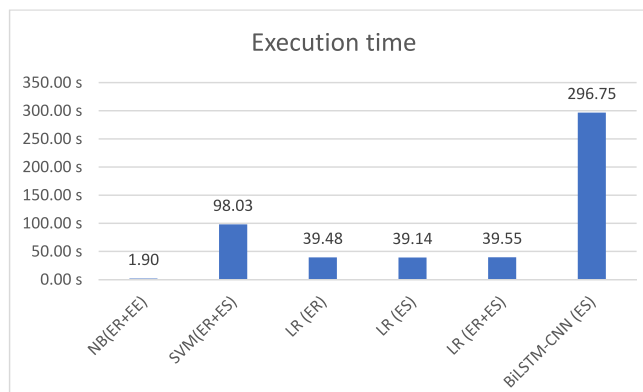


Fig. 18. Execution time of best classifiers for each algorithm.

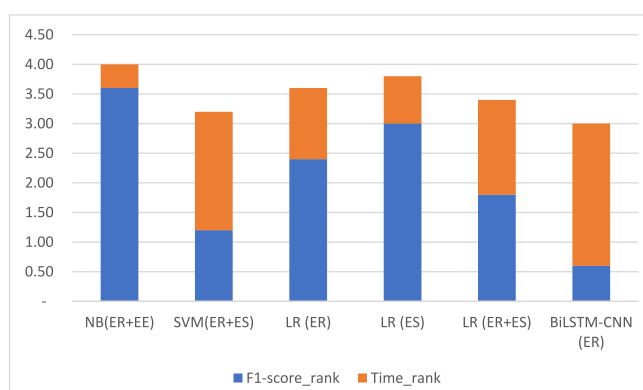


Fig. 19. The comprehensive ranking of best classifiers for each algorithm

6.2. Implications and stakeholder benefits

By illuminating the role of emojis in sentiment expression and demonstrating their impact on sentiment analysis, this study encourages stakeholders to give more attention to non-verbal cues in online communications when crafting policies.

Businesses in sectors such as retail, hospitality, and technology can utilize the study's findings to shape their social media monitoring policies. For instance, recognizing the importance of emojis in sentiment expression can help companies in these sectors refine their online customer service. This improved sentiment analysis capability can, for instance, enable a retail company to assess the reception of a new product more accurately based on online reviews and social media posts, thereby guiding marketing and production decisions. For machine learning practitioners and researchers, the proposed emoji handling

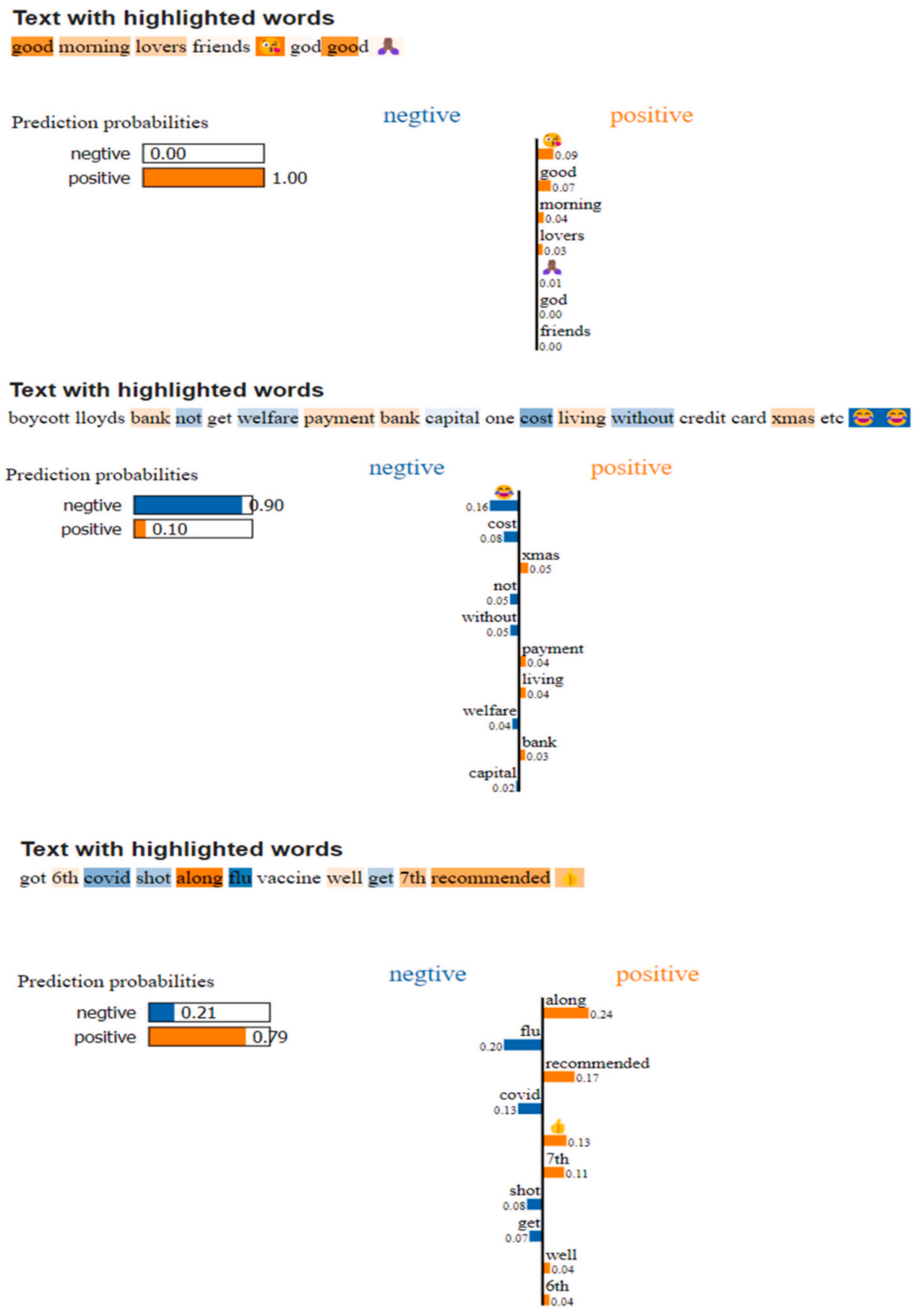


Fig. 20. Local explanation for an online review by the multi-view sentiment analysis model.

methods and the multi-view learning approach can be valuable additions to their toolkits. These novel methodologies can be used or further developed to improve the accuracy and interpretability of sentiment analysis models in future studies.

The impact of our study extends beyond just business applications. For instance, government agencies could use the proposed sentiment analysis model to gauge public sentiment towards new policies or public initiatives, such as the cost of living crisis in the UK and the government's response to public health events based on social media posts, thereby obtaining valuable feedback for policy adjustments.

By acknowledging the role of emojis in sentiment expression and proposing new ways to incorporate emojis into sentiment analysis, this study can potentially transform the way sentiment analysis is performed, leading to a more accurate and comprehensive understanding of online sentiments in various fields.

6.3. *Limitations and future work*

The present work has several limitations. While the dataset, Sentiment 140, is the most popular dataset used for sentiment analysis, it was not perfectly categorized as it was labeled by directly using the emoticons in the tweet. Therefore, the accuracy and F1-score may be lower than expected. Since the dataset of tweets containing emojis this study found is multi-domain, for consistency, Sentiment140 is the best choice among the available datasets. In the future, a primary dataset can be collected. In addition, one potential reason why adding emoji score methods does not perform as well as creating emoji embedding methods is that the emoji size (1662 emoji) used to train Emoji2Vec (Eisner et al., 2016) is larger than the emoji size in the emoji sentiment lexicon (751 emoji) provided by Kralj Novak et al. (2015). For future work, this study plans to train emoji embeddings and compute emoji scores based on the same emoji lexicon for a fairer comparison.

Table 5
Performance comparison of the *E*-BiLSTM-CNN model with other classifiers and building blocks.

Authors	Classifier	F1-score	Accuracy
Lou et al. (2020)	EA-Bi-LSTM	72.18 %	87.85 %
A. Singh et al. (2019)	EMJ-DESC	70.30 %	70.40 %
Liu et al. (2021)	CEmo-LSTM(text+E)	–	81.10 %
de Barros et al. (2021)	pre-trained BERT model-TweetSentBR	73.95 %	75.77 %
	pre-trained BERT model-2000-tweets-BR	81.51 %	83.16 %
Model 1	*E*-LSTM (building blocks of the proposed model)	81.33 %	81.33 %
Model 2	E-BiLSTM (building blocks of the proposed model)	81.31 %	81.31 %
Model 3	E-CNN (building blocks of the proposed model)	80.92 %	80.92 %
Model 4	E-BiLSTM-CNN model	81.43 %	81.44 %

CRediT authorship contribution statement

Qianwen Ariel Xu: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Chrisina Jayne:** Formal analysis, Supervision, Writing – review & editing. **Victor Chang:** Formal analysis, Project administration, Resources, Validation, Writing – review & editing.

Data availability

The datasets generated and analyzed during this study are available in the Kaggle repository:

Sentiment 140 Dataset: https://www.kaggle.com/datasets/kazanova/sentiment140;

Emoji Tweet Dataset: https://www.kaggle.com/datasets/nayan082/sentimentnewdataset

Acknowledgment

Prof Chang's work is partly supported by VC Research (VCR 0000207).

Appendix A

Table A1
Performance metrics for Naïve Bayes classifier.

Method	Accuracy	F1-score	Precision	Recall
Remove emojis	0.6841	0.6839	0.6844	0.6841
Emoji replacement (ER)	0.712	0.7119	0.7123	0.712
Adding emoji embeddings (EE)	0.7188	0.7164	0.7263	0.7187
Adding emoji scores (ES)	0.6887	0.6885	0.6891	0.6887
ER + EE	0.7291	0.727	0.7365	0.7291
ER + ES	0.7146	0.7144	0.7149	0.7146
EE + ES	0.7197	0.717	0.7283	0.7197
ER + EE + ES	0.7284	0.7259	0.7368	0.7283

Table A2
Performance metrics for support vector machine classifier.

Method	Accuracy	F1-score	Precision	Recall
Remove emojis	0.7379	0.7379	0.7379	0.7379
Emoji replacement (ER)	0.7709	0.7709	0.7709	0.7709
Adding emoji embeddings (EE)	0.7384	0.7384	0.7385	0.7384
Adding emoji scores (ES)	0.7661	0.7661	0.7663	0.7661
ER + EE	0.771	0.771	0.7711	0.771
ER + ES	0.7785	0.7785	0.7786	0.7785
EE + ES	0.7631	0.7631	0.7634	0.7631
ER + EE + ES	0.7782	0.7782	0.7782	0.7782

Table A3
Performance metrics for logistic regression classifier.

Method	Accuracy	F1-score	Precision	Recall
Remove emojis	0.7391	0.7391	0.7391	0.7391
Emoji replacement (ER)	0.7712	0.7712	0.7712	0.7712
Adding emoji embeddings (EE)	0.7387	0.7387	0.7387	0.7387
Adding emoji scores (ES)	0.7666	0.7666	0.7667	0.7666
ER + EE	0.7712	0.7712	0.7713	0.7712
ER + ES	0.7771	0.7771	0.7771	0.7771
EE + ES	0.7647	0.7647	0.7647	0.7647
ER + EE + ES	0.7773	0.7772	0.7773	0.7772

Table A4
Performance metrics for BiLSTM-CNN classifier.

Method	Accuracy	F1-score	Precision	Recall
Remove emojis	0.7748	0.7747	0.7748	0.7748
Emoji replacement (ER)	0.8214	0.8214	0.8215	0.8214
Adding emoji embeddings (EE)	0.8144	0.8143	0.8147	0.8144
Adding emoji scores (ES)	0.8024	0.8024	0.8025	0.8024
ER + EE	0.8149	0.8149	0.8151	0.8149
ER + ES	0.8194	0.8194	0.8194	0.8194
EE + ES	0.8146	0.8145	0.8147	0.8146
ER + EE + ES	0.8151	0.8151	0.8151	0.8151

References

Abedin, M.Z., Moon, M.H., Hassan, M.K., Hajek, P., 2021. Deep learning-based exchange rate prediction during the COVID-19 pandemic. Ann. Oper. Res. 2021, 1–52.

Agüero-Torales, M.M., Cobo, M.J., Herrera-Viedma, E., López-Herrera, A.G., 2019. A cloud-based tool for sentiment analysis in reviews about restaurants on TripAdvisor. Procedia Comput. Sci. 162, 392–399. https://doi.org/10.1016/j.procs.2019.12.002.

Bansal, B., Srivastava, S., 2019. Lexicon-based Twitter Sentiment Analysis for Vote Share Prediction Using Emoji and N-gram Features, p. 15. https://doi.org/10.1504/IJWBC.2019.098693.

Biswas, B., Sengupta, P., Kumar, A., Delen, D., Gupta, S., 2022. A critical assessment of consumer reviews: a hybrid NLP-based methodology. Decis. Support. Syst. 159, 113799 https://doi.org/10.1016/j.dss.2022.113799.

Bussmann, N., Giudici, P., Marinelli, D., Papenbrock, J., 2021. Explainable machine learning in credit risk management. *Comput. Econ. 57*, 203–216.

Chen, J., Chen, Y., He, Y., Xu, Y., Zhao, S., Zhang, Y., 2021. A classified feature representation three-way decision model for sentiment analysis. In: Applied INTELligence. SPRINGER. https://doi.org/10.1007/s10489-021-02809-1.

Chicco, D., Jurman, G., 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 21 (1), 6. https://doi.org/10.1186/s12864-019-6413-7.

Chowdhury, K.R., Sil, A., Shukla, S.R., 2021. Explaining a black-box sentiment analysis model with local interpretable model diagnostics explanation (LIME). In: Advances in Computing and Data Sciences: 5th International Conference, ICACDS 2021, Nashik, India, April 23–24, 2021, Revised Selected Papers, Part I 5. Springer International Publishing, pp. 90–101.

Das, S., Behera, R.K., kumar, M., & Rath, S. K., 2018. Real-time sentiment analysis of twitter streaming data for stock prediction. Procedia Comput. Sci. 132, 956–964. https://doi.org/10.1016/j.procs.2018.05.111.

de Barros, T. M., Pedrini, H., & Dias, Z. (2021). Leveraging emoji to improve sentiment classification of tweets. Proceedings of the 36th Annual ACM Symposium on Applied Computing, 845–852. doi:https://doi.org/10.1145/3412841.3441960.

Dehler-Holland, J., Okoh, M., Keles, D., 2022. Assessing technology legitimacy with topic models and sentiment analysis–the case of wind power in Germany. Technol. Forecast. Soc. Chang. 175, 121354.

Dewi, C., Tsai, B.J., Chen, R.C., 2022, November. Shapley additive explanations for text classification and sentiment analysis of internet movie database. In: Asian Conference on Intelligent Information and Database Systems. Springer Nature Singapore, Singapore, pp. 69–80.

Efat, M.I.A., Hajek, P., Abedin, M.Z., Azad, R.U., Jaber, M.A., Aditya, S., Hassan, M.K., 2022. Deep-learning model using hybrid adaptive trend estimated series for modelling and forecasting sales. Ann. Oper. Res. 2022, 1–32.

Eisner, B., Rocktäschel, T., Augenstein, I., Bošnjak, M., Riedel, S., 2016. emoji2vec: Learning Emoji Representations from their Description. https://doi.org/10.48550/ARXIV.1609.08359.

Emojipedia, 2022. Top Emoji Trends of 2021. https://blog.emojipedia.org/top-emoji-trends-of-2021/.

Ghosh, I., Chaudhuri, T.D., Alfaro-Cortés, E., Gámez, M., García, N., 2022. A hybrid approach to forecasting futures prices with simultaneous consideration of optimality in ensemble feature selection and advanced artificial intelligence. Technol. Forecast. Soc. Chang. 181, 121757.

Ghosh, I., Jana, R.K., Abedin, M.Z., 2023. An ensemble machine learning framework for Airbnb rental price modeling without using amenity-driven features. Int. J. Contemp. Hosp. Manag. 35 (10), 3592–3611. https://doi.org/10.1108/IJCHM-05-2022-0562.

Go, A., Bhayani, R., Huang, L., 2009. Twitter sentiment classification using distant. Supervision 1 (12), 6.

Graves, A., Schmidhuber, J., 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Netw. 18 (5–6), 602–610. https://doi.org/10.1016/j.neunet.2005.06.042.

HaCohen-Kerner, Y., Miller, D., Yigal, Y., 2020. The influence of preprocessing on text classification using a bag-of-words representation. PLoS One 15 (5), e0232525. https://doi.org/10.1371/journal.pone.0232525.

Hankamer, D., Liedtka, D., 2016. Twitter Sentiment Analysis with Emojis, 11.

Haque, A.B., Islam, A.N., Mikalef, P., 2023. Explainable Artificial Intelligence (XAI) from a user perspective: a synthesis of prior literature and problematizing avenues for future research. Technol. Forecast. Soc. Chang. 186, 122120.

He, X., Chen, Y., Ghamisi, P., 2020. Heterogeneous transfer learning for hyperspectral image classification based on convolutional neural network. IEEE Trans. Geosci. Remote Sens. 58 (5), 3246–3263. https://doi.org/10.1109/TGRS.2019.2951445.

Hirata, E., Matsuda, T., 2023. Examining logistics developments in post-pandemic Japan through sentiment analysis of twitter data. Asian Transport Stud. 9, 100110.

Hutto, C., Gilbert, E., 2014. VADER: a parsimonious rule-based model for sentiment analysis of social media text. Proc. Int. AAAI Conf. Web Soc. Media 8 (1), 1. https://ojs.aaai.org/index.php/ICWSM/article/view/14550.

Janssens, B., Bogaert, M., Maton, M., 2022a. Predicting the next pogačar: a data analytical approach to detect young professional cycling talents. Ann. Oper. Res. 1–32.

Janssens, B., Bogaert, M., Maton, M., 2022b. Predicting the next pogačar: a data analytical approach to detect young professional cycling talents. Ann. Oper. Res. 1–32.

Kamyab, M., Liu, G., Adjeisah, M., 2021. Attention-based CNN and bi-LSTM model based on TF-IDF and GloVe word embedding for sentiment analysis. Appl. Sci. 11 (23), 11255. https://doi.org/10.3390/app112311255.

Kastrati, Z., Dalipi, F., Imran, A.S., Pireva Nuci, K., Wani, M.A., 2021. Sentiment analysis of students' feedback with NLP and deep learning: a systematic mapping study. Appl. Sci. 11 (9), 3986.

Khan, Z.Y., Niu, Z., 2021. CNN with depthwise separable convolutions and combined kernels for rating prediction. Expert Syst. Appl. 170, 114528 https://doi.org/10.1016/j.eswa.2020.114528.

Kim, B., Park, J., Suh, J., 2020. Transparency and accountability in AI decision support: explaining and visualizing convolutional neural networks for text information. Decis. Support. Syst. 134, 113302.

Kim, D., Song, Y., Kim, S., Lee, S., Wu, Y., Shin, J., Lee, D., 2023. How should the results of artificial intelligence be explained to users?-research on consumer preferences in user-centered explainable artificial intelligence. Technol. Forecast. Soc. Chang. 188, 122343.

Kralj Novak, P., Smailović, J., Sluban, B., Mozetič, I., 2015. Sentiment of emojis. PLoS One 10 (12), e0144296. https://doi.org/10.1371/journal.pone.0144296.

Książek, W., Gandor, M., Pławiak, P., 2021. Comparison of various approaches to combine logistic regression with genetic algorithms in survival prediction of hepatocellular carcinoma. Comput. Biol. Med. 134, 104431 https://doi.org/10.1016/j.compbiomed.2021.104431.

Lampridis, O., Guidotti, R., Ruggieri, S., 2020, October. Explaining sentiment classification with synthetic exemplars and counter-exemplars. In: International Conference on Discovery Science. Springer International Publishing, Cham, pp. 357–373.

Leung, C.K., Pazdor, A.G., Souza, J., 2021, October. Explainable artificial intelligence for data science on customer churn. In: In 2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA). IEEE, pp. 1–10.

Liu, C., Fang, F., Lin, X., Cai, T., Tan, X., Liu, J., Lu, X., 2021. Improving sentiment analysis accuracy with emoji embedding. J. Saf. Sci. Resil. 2 (4), 246–252.

Lou, Y., Zhang, Y., Li, F., Qian, T., Ji, D., 2020. Emoji-based sentiment analysis using attention networks. ACM Trans. Asian Low-Resour. Language Information Process. 19 (5), 1–13. https://doi.org/10.1145/3389035.

Martín, C.A., Torres, J.M., Aguilar, R.M., Diaz, S., 2018. Using deep learning to predict sentiments: case study in tourism. Complexity.

Miron, V., Frasincar, F., Truşcă, M.M., 2023, June. Explaining a deep learning model for aspect-based sentiment classification using post-hoc local classifiers. In: International Conference on Applications of Natural Language to Information Systems. Springer Nature Switzerland, Cham, pp. 79–93.

Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L.T., Trajanov, D., 2020. Evaluation of sentiment analysis in finance: from lexicons to transformers. IEEE Access 8, 131662–131682.

Moreira, C., Chou, Y.L., Velmurugan, M., Ouyang, C., Sindhgatta, R., Bruza, P., 2021. LINDA-BN: An interpretable probabilistic approach for demystifying black-box predictive models. Decis. Support Syst. 150, 113561.

Nguyen, A., Pellerin, R., Lamouri, S., Lekens, B., 2023. Managing demand volatility of pharmaceutical products in times of disruption through news sentiment analysis. Int. J. Prod. Res. 61 (9), 2829–2840.

Pennington, J., Socher, R., Manning, C., 2014. Glove: global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in

Natural Language Processing (EMNLP), pp. 1532–1543. https://doi.org/10.3115/v1/D14-1162.
Priyadarshini, I., Cotton, C., 2021. A novel LSTM-CNN-grid search-based deep neural network for sentiment analysis. In: JOURNAL OF SUPERCOMPUTING, 77. SPRINGER, pp. 13911–13932. https://doi.org/10.1007/s11227-021-03838-w. Issue 12.
Rao, Y., Yang, F., 2022. A method for classifying information in education policy texts based on an improved attention mechanism model. Wirel. Commun. Mob. Comput. 2022, 1–7. https://doi.org/10.1155/2022/5467572.
Sahut, J.M., Hajek, P., 2022. Mining behavioural and sentiment-dependent linguistic patterns from restaurant reviews for fake review detection. Technol. Forecast. Soc. Chang. 177 (2022), 121532. https://doi.org/10.1016/j.techfore.2022.121532.
Salur, M.U., Aydin, I., 2020. A novel hybrid deep learning model for sentiment classification. In: IEEE Access, 8. IEEE-INST ELECTRICAL ELECTRONICS ENGINEERS INC, pp. 58080–58093. https://doi.org/10.1109/ACCESS.2020.2982538.
Samala, R.K., Chan, H.-P., Hadjiiski, L.M., Helvie, M.A., Cha, K.H., Richter, C.D., 2017. Multi-task transfer learning deep convolutional neural network: application to computer-aided diagnosis of breast cancer on mammograms. Phys. Med. Biol. 62 (23), 8894–8908. https://doi.org/10.1088/1361-6560/aa93d4.
Shin, D., 2021. Why does explainability matter in news analytic systems? Proposing explainable analytic journalism. Journal. Stud. 22 (8), 1047–1065.
Singh, A., Blanco, E., & Jin, W. (2019). Incorporating emoji descriptions improves tweet classification. Proceedings of the 2019 Conference of the North, 2096–2101. doi:10.18653/v1/N19-1214.
Singla, C., Al-Wesabi, F.N., Singh Pathania, Y., Sulaiman Alfurhood, B., Mustafa Hilal, A., Rizwanullah, M., Ahmed Hamza, M., Mahzari, M., 2022. An optimized deep learning model for emotion classification in tweets. Comput. Mater. Contin. 70 (3), 6365–6380. https://doi.org/10.32604/cmc.2022.020480.
Stöckli, D.R., Khobzi, H., 2021. Recommendation systems and convergence of online reviews: the type of product network matters! Decis. Support. Syst. 142, 113475 https://doi.org/10.1016/j.dss.2020.113475.
Suzuki, S., Zhang, X., Homma, N., Ichiji, K., Sugita, N., Kawasumi, Y., Ishibashi, T., Yoshizawa, M., 2016. Mass detection using deep convolutional neural network for mammographic computer-aided diagnosis. In: 2016 55th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), pp. 1382–1386. https://doi.org/10.1109/SICE.2016.7749265.
Wang, S., Tang, C., Sun, J., Zhang, Y., 2019. Cerebral micro-bleeding detection based on densely connected neural network. Front. Neurosci. 13, 422. https://doi.org/10.3389/fnins.2019.00422.
Wołk, K., 2020. Advanced social media sentiment analysis for short-term cryptocurrency price prediction. Expert. Syst. 37 (2), e12493.
Xiao, R., Cui, X., Qiao, H., Zheng, X., Zhang, Y., Zhang, C., Liu, X., 2021. Early diagnosis model of Alzheimer's disease based on sparse logistic regression with the generalized elastic net. Biomedical Signal Processing and Control 66, 102362. https://doi.org/10.1016/j.bspc.2020.102362.
Yan, Na, 2020. Sentiment-new-dataset—Text With Emoticon Tweets. https://www.kaggle.com/datasets/nayan082/sentimentnewdataset.
Yang, C., Abedin, M.Z., Zhang, H., Weng, F., Hajek, P., 2023. An interpretable system for predicting the impact of COVID-19 government interventions on stock market sectors. Ann. Oper. Res. 2023, 1–28.
Zhang, Z., Wei, X., Zheng, X., Li, Q., Zeng, D.D., 2022. Detecting product adoption intentions via multiview deep learning. INFORMS J. Comput. 34 (1), 541–556. https://doi.org/10.1287/ijoc.2021.1083.
Zheng, J., Zheng, L., 2019. A hybrid bidirectional recurrent convolutional neural network attention-based model for text classification. In: IEEE Access, 7. IEEE-INST ELECTRICAL ELECTRONICS ENGINEERS INC, pp. 106673–106685. https://doi.org/10.1109/ACCESS.2019.2932619.
Zytek, A., Liu, D., Vaithianathan, R., Veeramachaneni, K., 2021. Sibyl: understanding and addressing the usability challenges of machine learning in high-stakes decision making. IEEE Trans. Vis. Comput. Graph. 28 (1), 1161–1171.

Qianwen Ariel Xu is a PhD student at Operations and Information Management, Aston Business School, Aston University UK. She was previously a PhD student in Computer Science at the School of Computing, Engineering and Digital Technologies at Teesside University. Her dissertation research focuses on box-office prediction based on techniques of Artificial Intelligence and Data Science. She completed her master's degree in Business Analytics with Distinctions from the University of Liverpool, UK. She is also a member of Prof. Chang's research team. She is a hardworking, dedicated and resourceful student who can make things happen. She has published several publications in refereed academic journals, such as Technological Forecasting and Social change, Information Systems Frontiers, Expert Systems, Journal of Global Information Systems, etc.

Chrisina Jayne is Dean of the School of Computing & Digital Technologies. She received her PhD degree in Applied Mathematics and an MSc in Mathematics and Informatics from Sofia University, Bulgaria. She also holds an MSc in Computing Science and a Postgraduate Diploma in Management from Birkbeck College, University of London. Chrisina joined Teesside University in August 2019, from Oxford Brookes University, where she was the Head of the School of Engineering, Computing and Mathematics. As well as Oxford Brookes University, Chrisina has previously worked in a number of universities in the UK, including Head of the School of Computing and Digital media at Robert Gordon University and Head of the Department of Computing at Coventry University. Chrisina is a Senior Fellow of the Higher Education Academy and Chartered IT Professional Fellow of the British Computer Society (BCS). She was awarded a UK National Teaching Fellowship award in 2009 in recognition of excellence in learning and teaching.

Victor Chang is a Professor of Business Analytics at Operations and Information Management, Aston Business School, Aston University UK, since mid-May 2022. He was previously a Professor of Data Science and Information Systems at the School of Computing, Engineering and Digital Technologies, Teesside University, UK, between September 2019 and mid-May 2022. He has deep knowledge and extensive experience in AI-oriented Data Science and has significant contributions in multiple disciplines. Within 4 years, Prof Chang completed Ph.D. (CS, Southampton) and PGCert (Higher Education, Fellow, Greenwich) while working for several projects simultaneously. Before becoming an academic, he has achieved 97 % on average in 27 IT certifications. He won 2001 full Scholarship, a European Award on Cloud Migration in 2011, IEEE Outstanding Service Award in 2015, best papers in 2012, 2015 and 2018, the 2016 European award: Best Project in Research, 2016–2018 SEID Excellent Scholar, Suzhou, China, Outstanding Young Scientist award in 2017, 2017 special award on Data Science, 2017–2022 INSTICC Service Awards, Talent Award Suzhou 2019, Top 2 % Scientist 2018/2019, 2019/2020, 2020/2021, 2021/2022 and 2022/2023, the most productive AI-based Data Analytics Scientist between 2010 and 2019, Highly Cited Researcher 2021 and numerous awards mainly since 2011. Prof Chang was involved in different projects worth more than £14 million in Europe and Asia. He has published 3 books as sole authors and the editor of 2 books on Cloud Computing and related technologies. He published 1 book on web development, 1 book on mobile app and 1 book on Neo4j. He gave 40 keynotes at international conferences. He is widely regarded as one of the most active and influential young scientist and expert in IoT/Data Science/Cloud/security/AI/IS, as he has the experience to develop 10 different services for multiple disciplines. He is the founding conference chair for IoTBDS, COMPLEXIS and FEMIB to build up and foster active research communities globally with positive impacts.