A Novel Machine Learning Model for Predicting the Meaning of an Emojis String in Social Media Platforms

Mossaad BEN AYED[1,2]*, Ali ALSAAWI[3]

[1] Department of Electronic Industrial, ENISo, Sousse University, Tunisia
mossaad.benayed@eniso.u-sousse.tn, mossaad_benayed@yahoo.fr (**Corresponding author*)

[2] Computer and Embedded Systems Laboratory, ENIS, Sfax University, Tunisia

[3] Department of English, College of Humanities & Sciences, Majmaah University, Majmaah, Saudi Arabia
a.alsaawi@mu.edu.sa

Abstract: This paper delves into how social media, an increasingly pervasive global phenomenon, shapes the language nowadays. It zeroes in on emojis, a rapidly evolving means of communication that deviates from traditional verbal and nonverbal expressions, examining their impact on linguistic development. Emojis, initially conveying emotions, have shifted towards representing feelings rather than explicit meanings. Users now create entire messages using strings of emojis instead of sentences or phrases. The present study examines how people understand emoji-based messages, particularly through a survey conducted in Saudi Arabia. Additionally, an intelligent system leveraging machine learning is introduced to decode the meanings within emoji messages. The creation of EmojiString, a novel dataset, aids in better understanding these messages by utilizing advanced models like long short-term memory (LSTM) and MultiLayer Perceptron (MLP). The proposed model boasts an average accuracy of 82.22%, surpassing the existing methods. These results strongly support the idea that emojis serve as vital contextual cues in everyday communication. They are not just whimsical symbols, but meaningful elements that shape the interactions between persons. This research underscores the need to recognize emojis' nuanced roles in the evolution of modern language, marking a significant step forward in understanding their impact on how people communicate with one another.

Keywords: Emojis, WhatsApp, Natural language, Computer-mediated discourse, Machine learning, LSTM, MLP.

1. Introduction

The use of emojis in communication has proliferated in recent years, most frequently on social media platforms. The rise of communication technologies has triggered the adoption of emojis as a distinct way of communicating. The popularity of emojis has attracted the attention of researchers from different backgrounds and contexts (Albawardi, 2018). Some of these studies have explored the effect of emojis and how they are used, with fewer examining emojis from a linguistic perspective, particularly related to the Arabic language (Yafooz, 2021).

Among various social media platforms, WhatsApp stands out as a popular place for the use of emojis, especially in Saudi Arabia. It is considered one of the most prevalent social platforms in the region. The popularity of WhatsApp highlights the importance of emojis as a unique tool employed by users for exchanging images, audio, and videos individually or in groups on their smartphones. Figure 1 presents a categorization of the common uses of emojis. As shown in Figure 1, emojis are often used to express feelings and emotions, a subject of many studies.

Mahmud et al. (2020) demonstrated that emojis effectively convey emotional meanings, improving the accuracy of emotion recognition compared to using text alone. Most emojis are employed to express positive emotions, such as happiness and excitement. Additionally, emojis play a role in reducing misunderstandings in digital communication, clarifying the intended tone of a message. For instance, inserting a smiling face emoji can signal that a message is meant to be humorous rather than serious (Kutsuzawa et al., 2022).

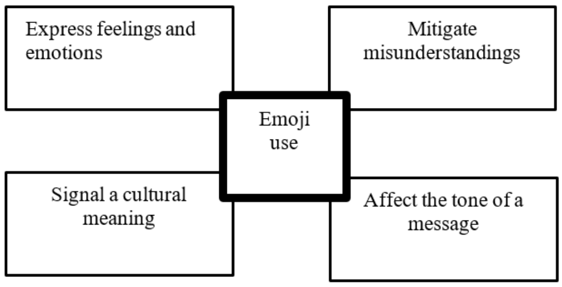

Figure 1. The widespread use of emojis

Emojis can have different meanings depending on the cultural context in which they are used, conveying specific messages. For example, the interpretation of the "thumbs up" emoji can vary between Western cultures and Middle Eastern countries (Chandra Guntuku et al., 2019). Moreover, the use of emojis influences the tone of a message, helping to identify whether a message is casual or formal, and can be used to soften the impact of bad news (Robertson et al., 2020).

https://doi.org/10.24846/v33i1y202408

This paper introduces a novel use of emojis, specifically as a short phrase. Limited research has been conducted on the replacement of natural language with emojis. Cohn et al. (2019) explored emoji-only communication and replacing words with emoji revealing significant challenges and high creativity. Siever et al. (2020) surveyed the relationship between emojis and text on Instagram, assessing the capability of emojis to represent syntactic and semantic elements of the text.

This paper is based on a survey that explores the functions for which Saudi users involve emojis in WhatsApp messages as part of natural language. The study then developed a system to transform a string of emojis into the predicted meaning, creating a new dataset to ensure prediction accuracy based on machine learning methods.

The paper is organized as follows. Section 2 presents a comprehensive review of the existing specialised literature regarding the usage of emoji. A detailed exposition of the proposed system architecture and of the utilised method is illustrated in section 3. Then, section 4 delineates the implementation process and details the experimentation carried out. Section 5 provides the discussions regarding the result of the experiment proposed in this paper. Section 6, the final one, offers the concluding remarks and displays the key findings and insights for a possible subsequent development of the proposed research.

2. Literature Review

Social media is increasingly used worldwide for various purposes, both professional and private. Emojis, such as faces with different expressions, are a core element in messaging used to express feelings and to add emotion to an utterance (Al-Azani & El-Alfy, 2018; Velioğlu et al., 2018). Moreover, emojis are seen as important contextualization cues that contribute to the mechanics of virtual communication (Gumperz, 1992). The understanding of these cues is partly shaped by the individual's cultural and social background. Thus, both verbal and non-verbal communicative tools are utilized in interactions.

Early studies have examined the use of emojis focused on emotional icons, such as facial expressions (Al Rashdi, 2018). However, few studies have addressed the functions performed by the use of emojis. Identifying these would aid in understanding the intended meaning of the text (Dresner & Herring, 2010). Thus, it is crucial to understand the reasons for employing certain emojis in text. Studies have found a range of uses of emojis (Neel et al., 2023; Tang & Hew, 2019). Kelly & Watts (2015) conducted a study exploring the reasons for the use of emojis in computer-mediated communication. They found that it was a way of expressing feelings, maintaining conversations, and providing a joyful interaction environment. Skovholt et al. (2014) examined the functions of emoticons in workplace e-mails and found that they were used to show positive attitudes, act as markers of jokes, and hedge.

Some recent studies have been conducted in the Gulf region concerning the use of emojis and their functions. For instance, Al Rashdi (2018) found that Omanis employ emojis to indirectly and playfully reduce face attacks. Additionally, examining why Omanis (males and females) use emojis in WhatsApp, they found it was due to several reasons, such as indicating emotions and the tone of the message, and to create alignment between participants. Another study, conducted by Albawardi (2018), found that female Saudi university students use emojis to manage interpersonal relationships in WhatsApp messages.

Yeole et al. (2015) presented a novel approach to determine a user's emotion through the input text and emojis. They supposed that emojis are used for indirect emotions, to express positive, negative, or neutral sentiments. Their proposed system classifies emotion based on the text and emojis the user enters. Rodrigues et al. (2018) developed a new dataset called the Lisbon Emoji and Emoticon Database (LEED) to interpret the use of emojis in online communication. The dataset is built to highlight the relationship between emojis and their meanings. The authors focused on the aspects of familiarity, concreteness, aesthetic appeal, meaningfulness, visual complexity, and valence.

Chandra & Prasad (2021) introduced a generalized system capable of integrating with various machine learning and deep learning classification algorithms. The system's primary objective is to interpret emojis within sentences, specifically focusing on detecting and translating offensive or sexual connotations represented by emojis into English words. Chandra & Prasad's (2021) approach involves the utilization of a hierarchical lookup data structure for storing and retrieving emojis along with their meanings. They employed the cartesian approach which is a noteworthy method for establishing correlations between words and emojis. This approach involves a systematic comparison of the Cartesian product of words and emojis, aiming to identify meaningful associations. By examining the cross-combinations of words and emojis, the system seeks to discern patterns that indicate the intended interpretations or semantic relationships within a given context. Consequently, the system successfully identifies the intended interpretation of emojis, considering all possible meanings, and this interpretation is further validated by the high accuracy of the employed deep learning classifier for classifying the translated sentences.

Seyednezhad et al. (2018) studied the pattern of use of emojis, demonstrating that they are often used in short messages. The authors built a co-occurrence bipartite emoji-word network to examine the reference to sentiment and meaning in emoji use, finding that emojis tend to be used to express positive feelings rather than addressing the semantics of the subject of the conversation. As identified by Gomes & Casais (2018), the meaning of an emoji directly relates to the significance of the word in the message.

Ebel & Dutra (2022) studied the language of emojis and explored young people's ability to communicate in emoji-only messaging scenarios. The results demonstrated that only 20% composed messages entirely of emojis. The findings also indicated that the participants gained an increasing understanding of the meaning of emojis used.

The studies reviewed above indicate that the use of emojis on social media platforms is of interest to researchers; however, more studies are needed to achieve greater clarity and understanding, especially regarding the functions that emojis perform. The main challenge concerns the multiple interpretations of an emoji string and the risk of misunderstanding. Hence, this study focuses on the meaning of emoji strings. The meanings attributed to the use of emojis by Saudis in WhatsApp was established as a natural language. Then, a model based on artificial intelligence was proposed to predict the meaning of an emoji string entered into the system.

3. Method

The main problem addressed in this study was to propose an accurate system allowing reliable translation of a string of emojis into the correct meaning based on Long Short-Term Memory (LSTM) and MultiLayer Perceptron (MLP). The proposed system, illustrated in Figure 2, comprises two phases: (1) the translation phase, and (2) the classification phase.

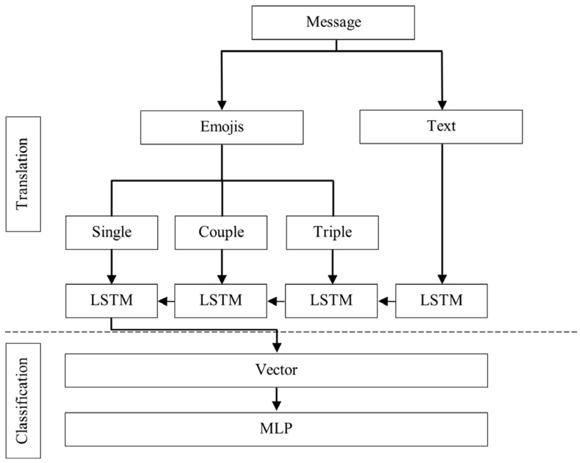

Figure 2. Proposed model based on LSTM and MLP

LSTM is a type of recurrent neural network (RNN) architecture designed to overcome the limitations of traditional RNNs in capturing long-term dependencies in sequential data. It is particularly effective in tasks involving sequences, such as natural language processing and time-series prediction. LSTM units have memory cells and gating mechanisms that enable the model to selectively store, read, and erase information over extended sequences, making them well-suited for handling context over time. MLP is a type of artificial neural network characterized by multiple layers of interconnected nodes (neurons). It consists of an input layer, one or more hidden layers, and an output layer. Each connection between nodes has an associated weight, and each node applies an activation function to the weighted sum of its inputs. MLPs are widely used for various machine learning tasks, including classification and regression. They are known for their ability to learn complex relationships in data through training on labelled examples.

The translation phase aims to determine the meaning of emojis based on the created dataset. It begins by splitting the message into textual and emoji components. The system then extracts potential meanings from the textual component of the message. The emoji components are divided into various sizes, ranging from a single emoji to a string of emojis. A combination of a set of emojis and a set of emoji strings is performed to explore suggested meanings. A similarity matrix is created to establish meaningful associations between the identified textual components and the emoji components. It facilitates the matching of textual and emoji meanings, aiding in the interpretation of the overall message. The proposed translation method can also be applied when the message consists entirely of emojis.

The study used the word2vec tool (Mikolov et al., 2013) to transform words into vectors, enabling computation. This tool defines the context based on words that are semantically similar. In the present proposal model, the skip-gram model in word2vec is preferred over Continuous Bag-of-Words (CBOW) because the prediction of context depends on the current words used.

The classification phase, employing LSTM and MLP, aims to find accurate meanings. The widely used LSTM model in natural language processing addresses issues such as gradient explosion and long-term dependence. The LSTM output generates a vector of sentences. The MLP functions as a sorter to determine the optimal (most probable) meaning of the message. The MLP comprises 32 neurons and one output layer with a single neuron. The sigmoid equation serves as the activation function. The MLP output layer indicates the probability of the sentence being correct. Thus, the selected sentence will be the synonym of the entered message, encompassing both the text and emojis.

In the training phase, the binary cross-entropy loss function is applied to samples, described by equation 1.

$$L = -\frac{1}{n}\sum s[R_l \log(p) + (1 - R_l)\log(1 - p) \quad (1)$$

where s is the number of sentences in the vector, R_l is the real label of a sentence, and p is the probability of the predicted model.

4. Experiment

Several datasets in the literature focus on emojis and their meanings, primarily extracting feelings and emotions from a message containing emojis. To assess the accurate recognition of the meaning of an emoji string, we designed a survey for Saudis using WhatsApp consisting of 20 common sentences. The results, presented in Table 1, indicate a mean recognition rate of 42.2%. This outcome underscores the importance of extending the dataset beyond single emoji. Therefore, our study aims to enhance these datasets to extract meanings, particularly from the usage of emoji strings.

Table 1. Survey results

English	*Arabic*	*Emojis*	*Recognition rate (%)*
"What's up?"	ايش الاخبار	🙋‍♂️🙋‍♀️	37.5
"How's it going?"	كيف الامور	👋🏼💬	50
"Hey there!"	هلا و الله	👋🏼	81.3
"Good morning"	صباح الخير	🌅	31.3
"How are you?"	كيف الحال	😊😃	37.5
"What's new?"	ايش الجديد	💬🆕	75
"Long time no see"	مبطين ماشفناك	👋🏼🕵️‍♂️	12.5
"How have you been?"	وين ايامك	👋🏼💬😃	12.5
"What's up with you?"	وش علومك	🤔💬	75
"How was your day?"	كيف الوضع اليوم	🤔😃	43.8
"What have you been up to?"	وش عندك جديد	🤔💬💼	25
"Catch up soon?"	خلنا على تواصل	💬🕰️	18.8
"Let's chat"	خل نسولف	💬💭	31.3
"What's the plan?"	ايش الخطة	💬🗺️	43.8
"Where are you?"	وينك	📍👀	75
"What are you doing?"	وش تسوي	🤔💼	31.3
"You around?"	انت قريب	📍👀💬	75
"How's work?"	كيف الشغل	💼👨‍💼👩‍💼	12.5
"How's everything?"	كيف الأوضاع	💬😃	37.5
"Just checking in"	تجربة	💬👀	37.5

The EmojiString dataset, developed in this study, encompasses the meaning of a set of emojis (String Emojis). Notably, the EmojiString dataset encompasses more senses than the EmojiNET (Chandra & Prasad, 2021) dataset and WordNet (Toet & van Erp, 2019). WordNet is considered a lexical database and semantic network of the English language. It groups English words into sets of synonyms called synsets, representing distinct lexical concepts. These synsets are linked by various semantic relations, such as hypernyms (generalization) and hyponyms (specific instances). WordNet provides a structured organization of words, capturing relationships between them and facilitating applications in natural language processing, linguistics, and artificial intelligence.

The EmojiString dataset offers a comprehensive collection of emoji combinations, each associated with a diverse range of meanings and contexts. Unlike the EmojiNET and WordNet datasets, which primarily focus on individual emoji meanings, EmojiString extends beyond individual emojis to capture the nuances of entire emoji sequences and their interpretations. This broader scope allows EmojiString to encapsulate multiple senses and connotations within a single entry, accommodating the varied ways in which emoji sequences are used to convey complex emotions, actions, or concepts. The EmojiString offers a more nuanced and contextually rich portrayal of emoji semantics, making it a valuable resource for understanding the intricate dynamics of emoji-based communication.

Unlike EmojiNET, which primarily focuses on single meanings for each emoji, EmojiString expands this perspective by encompassing a broader range of potential interpretations, capturing the multifaceted nature of emojis in communication.

To illustrate, while EmojiNET may attribute one specific meaning to a thumbs-up emoji, EmojiString could include multiple senses such as approval, encouragement, or positivity, reflecting the varied ways users may interpret and use the same emoji in different contexts. This approach allows EmojiString to account for the cultural, contextual, and subjective dimensions of emoji meanings, providing a more comprehensive representation of the semantic richness inherent in these symbols. Unfortunately, specific quantitative data on the exact number of senses may not be available, but the emphasis lies in the qualitative expansion of interpretive possibilities offered by EmojiString compared to EmojiNET.

The EmojiString dataset compiles 24902 emojis and a set of emoji meanings sourced from various channels. The utilized dataset encompasses a diverse array of emoji combinations, each linked to specific meanings or contexts. For instance, within the EmojiString dataset, one entry might feature a sequence of emojis like 🚗🌞🏖️, symbolizing a 'summer road trip.' These instances not only showcase the dataset's depth and intricacy but also emphasize its effectiveness in capturing nuanced emotions and feelings conveyed through emoji sequences.

The dataset was created based on the following criteria:

- Data source reliability: Collected data were compiled from various trustworthy sources, including the EmojiNET dataset, the WordNet dataset, chat logs, survey data, and searches of artificial intelligence engines, such as ChatGPT (ChatGPT, n.d.) and BlackBox AI (BlackBox AI, n.d.);
- Diversity of emojis: The dataset incorporates diverse emojis supporting various emotions, objects, and expressions. Certain emojis may have different meanings based on cultural, regional, and religious backgrounds;
- Labelling: Each emoji is assigned one or more labels.

The EmojiString consists of 98 strings of emojis with different and similar meanings. Each sample is associated with one or more sentences.

Additionally, EmojiString prioritizes the diversity of emojis, including those with different interpretations across cultural, regional, and religious backgrounds. Each emoji in the dataset is labelled with one or more descriptors, enhancing the granularity of the dataset and providing a more nuanced understanding of the possible meanings associated with each symbol.

In summary, EmojiString is a rich and nuanced dataset that goes beyond the conventional single-meaning association of emojis. It provides a valuable resource for exploring the multifaceted nature of emoji usage in natural language processing and communication analysis.

5. Results and Discussion

The LSTM classifier demonstrates excellent predictive accuracy among all suggested meanings (both textual and emojis) with high probability. During the splitting phase, senses are extracted from the textual and emoji components. The pre-trained model based on the EmojiString and EmojiNet datasets is applied to the emoji component. This model randomly assigns 80% of the dataset for training and 20% for testing. The output vector provided by the LSTM classifier is filled with potential textual senses, and the spaCy library, available for Python, is employed to derive textual meaning.

In the classification phase, the MLP manages keywords and sentences, utilizing the Twitter Developer's API (Ertam, 2018) as a dataset to identify senses. This dataset comprises 1,440,438 sentences with corresponding categories. Post-training, the model can categorize words based on their respective categories.

Principle Component Analysis (PCA) (Siddiqui, et al., 2018) is applied to reduce the number of features and determine their required number. Figure 3 illustrates the relationship between

PCA features and the variance graph. The aim of describing PCA features versus variance is to understand the distribution and importance of features in a dataset after applying PCA. PCA is a dimensionality reduction technique that transforms the original features of a dataset into a new set of uncorrelated variables called "principal components". It is observed that the curve flattens out at around 840 features. This implies that, according to the analysis, the inclusion of up to 840 features is effective in capturing most of the variability in the dataset. Beyond this point, adding more features does not significantly enhance the cumulative explained variance. Therefore, the dataset is divided into training and testing sets at an 8:2 ratio. PCA also ensures the normalization of all values to reduce computation complexity.

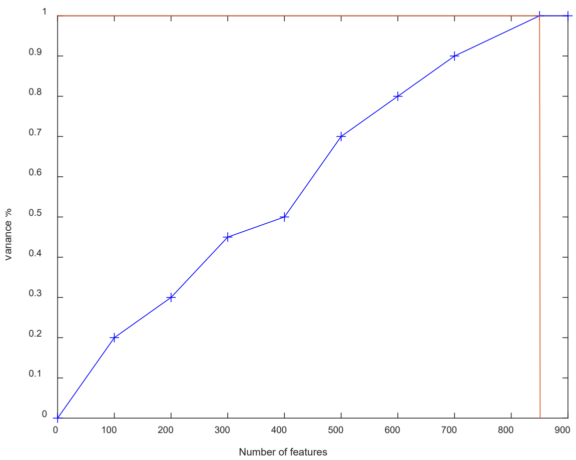

Figure 3. PCA features vs. variance

The MLP model was trained for 50, 100, 150, and 250 epochs, with batch sizes varying from 300 to 500. After conducting an extensive analysis, an epoch size of 150 and batch size of 500 were identified as the optimal parameters. Cross-validation testing was carried out tenfold during training. The model indicates a mean accuracy of 82.23%, demonstrating high efficiency. The accuracy is defined as the ratio of correctly predicted instances to the total number of instances in the dataset. The formula for accuracy is:

$$Accuracy = \frac{Number\ of\ Correct\ \mathrm{Pr}\,edictions}{Total\ Number\ of\ \mathrm{Pr}\,edictions} \times 100 \quad (2)$$

The model also demonstrates a low variance value of 1.83 x 10^-5. This finding confirms that the model's predictions are consistent, closely aligned with the expected outcomes, and exhibit stability. The model's validation was performed on the testing set of the Twitter Developer's API dataset, achieving an accuracy of around 82%, as depicted in Figure 4.

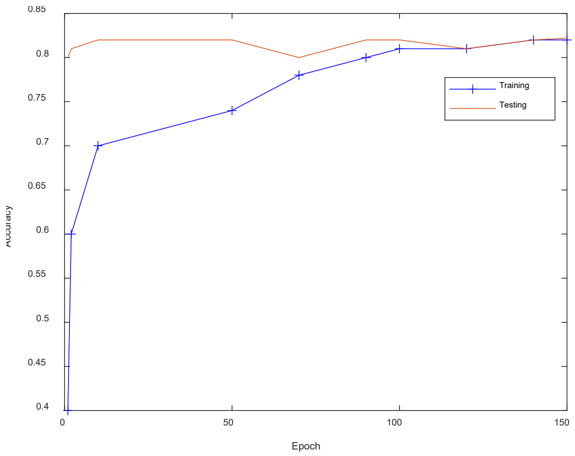

Figure 4. MLP trained-test accuracy

The comparison presented in Table 2 utilized the same model as previous studies, but it was applied to the EmojiString dataset proposed in the present paper. While this method may limit direct comparisons with studies using different datasets, it provides valuable insights into how the model performs in a new setting. The goal of this work was to assess how well the model adapts and accurately captures the subtle meanings conveyed by emoji sequences in the present specific dataset. Although the comparison may not directly translate to studies using other datasets, the approach proposed in this paper allows for a focused evaluation of the model's effectiveness within the context of emoji-based communication in the present dataset. This approach extends the findings of previous studies to a new dataset, enhancing the understanding of the model's capabilities across various data domains. Çöltekin & Rama (2018) aimed to identify the meaning of emojis using the Support Vector Machine (SVM) classifier, achieving an accuracy of 58.88%, with precision and recall of 0.36, and an F1 score of about 0.35. Himabindu et al. (2022) proposed a model based on BiLSTM Attention Random Forest (BARF) modeling, obtaining better results than Çöltekin & Rama (2018), with an accuracy of 61.14%, precision of 0.66, recall of 0.59, and F1 score of 0.59. The present proposed model, which examines the meaning of string emojis using an LSTM-MLP machine learning model, improves accuracy to 82.22%, with a precision of 0.87, and recall and F1 score of about 0.84. The proposed model achieved better performance, because the EmojiString dataset is an enhanced dataset, and the combination of LSTM and MLP is more efficient than other models.

To sum up, emojis are becoming increasingly important in people's communication. This

Table 2. Comparison between models

Model	*Study*	*Year*	*Accuracy*	*Precision*	*Recall*	*F1 Score*
SVM	Çöltekin & Rama (2018)	2018	58.88%	0.36	0.36	0.35
BARF	Himabindu et al. (2022)	2022	61.14%	0.66	0.59	0.59
LSTM-MLP	Study proposal	2023	**82.22%**	**0.87**	**0.84**	**0.84**

emergence is not a threat or a replacement for language, but is part of normal language development. The growing use of emojis is likely to continue with the rapidly rising influence of technology in daily life. Context is always a crucial language, and this becomes even more significant with the use of emojis, which Gumperz (1992) refers to as "contextualization cues". The model proposed in this paper focuses on ascertaining the meaning of string emojis via the use of LSTM-MLP machine-learning models. The results reveal that the LSTM classifier has a good predictive accuracy among all the suggested meanings (textual and emoji), with a high probability.

6. Conclusion

This paper highlights the widespread use of emojis in online social platforms among Saudis, introducing a machine learning-based model designed to translate messages, encompassing both text and emojis, to derive an accurate meaning. The LSTM model explores all potential meanings of message components, including text, and emoji strings. Subsequently, the MLP model identifies the best meaning based on higher probabilities. The proposed model achieves reasonable accuracy, attributed to the creation of the "EmojiString" dataset which allows for the identification of the meaning of a set of emojis, rather than the one of a single emoji.

Looking ahead, the EmojiString dataset could be improved by including samples composed of more than three emojis and with more varied meanings. Such improvements could contribute to a deeper understanding of emojis and their inclusion in messages and daily conversations. Future research is encouraged to delve into the functions of the proposed dataset. Additionally, exploring the cultural impact of this model among users, especially about the potential replacement of text, is essential. Emojis are here to stay, and acknowledging them as part of language development is crucial.

REFERENCES

Al-Azani, S. & El-Alfy, E.-S. M. (2018) Combining emojis with Arabic textual features for sentiment classification. In: *2018 9th International Conference on Information and Communication Systems (ICICS), 03-05 April 2018, Irbid, Jordan.* New York City, USA, Institute for Electrical and Electronics Engineers (IEEE). pp. 139-144. doi: 10.1109/IACS.2018.8355456.

Albawardi, A. (2018) The translingual digital practices of Saudi females on WhatsApp. *Discourse, Context & Media.* 25, 68–77. doi: 10.1016/j.dcm.2018.03.009.

Al Rashdi, F. (2018) Functions of emojis in WhatsApp interaction among Omanis. *Discourse, Context & Media*. 26, 117–126. doi: 10.1016/j.dcm.2018.07.001

BlackBox AI. (n.d.) *How can I help you today?*. https://www.blackbox.ai/ [Accessed 15th February 2023].

Chandra Guntuku, S., Li, M., Tay, L. & Ungar, L. H. (2019) Studying cultural differences in emoji usage across the east and the west. In: *Proceedings of the International AAAI Conference on Web and Social Media*, *11-14 June 2019, Münich, Germany*. Washington, DC, USA, Association for the Advancement of Artificial Intelligence. pp. 226-235. doi: 10.1609/icwsm.v13i01.3224.

Chandra, P. & Prasad, U. (2021) Classification of Emojis using Artificial Neural Network and Natural Language Processing. In: *2021 8th International Conference on Computing for Sustainable Global Development (INDIACom), 17-19 March 2021, New Delhi, India*. New York City, USA, Institute for Electrical and Electronics Engineers (IEEE). pp. 205–212.

ChatGPT (n.d.) *Try ChatGPT*. https://chat.openai.com/auth/login [Accessed 10th December 2022].

Cohn, N., Engelen, J. & Schilperoord, J. (2019) The grammar of emoji? Constraints on communicative pictorial sequencing. *Cognitive Research: Principles and Implications*. 4, 1–18. doi: 10.1186/s41235-019-0177-0.

Çöltekin, Ç. & Rama, T. (2018) Tübingen-Oslo at SemEval-2018 task 2: SVMs perform better than RNNs in emoji prediction. In: *Proceedings of the 12th International Workshop on Semantic Evaluation*

(SemEval-2018), 5-6 June 2018, New Orleans, Louisiana. Association for Computational Linguistics. pp. 34–38. doi: 10.18653/v1/S18-1004.

Dresner, E. & and Herring, S. C. (2010) Functions of the nonverbal in CMC: Emoticons and illocutionary force. *Communication Theory*. 20(3), 249–268. doi: 10.1111/j.1468-2885.2010.01362.x.

Ebel, I. & Dutra, J. P. (2022) "Do You Speak Emoji? The Language of Emojis. In: Tuten, T. L. & Hanlon, A. (eds.) *The SAGE Handbook of Social Media Marketing*. London, UK, SAGE Publications Inc. pp. 320-336. doi: 10.4135/9781529782493.

Ertam, F. (2018) Deep learning based text classification with Web Scraping methods. In: *2018 International Conference on Artificial Intelligence and Data Processing (IDAP), 28-30 September 2018, Malatya, Turkey*. New York City, USA, Institute for Electrical and Electronics Engineers (IEEE). pp. 1–4. doi: 10.1109/IDAP.2018.8620790.

Gomes, R. F. & Casais, B. (2018) Feelings generated by threat appeals in social marketing: Text and emoji analysis of user reactions to anorexia nervosa campaigns in social media. *International Review on Public and Nonprofit Marketing*. 15, 591–607.

Gumperz, J. J. (1992) Contextualization revisited. In: Peter Auer, O. & Di Luzio, A. (eds.) *The Contextualization of Language*. Pragmatics & Beyond New Series, 22. Amsterdam, Netherlands, John Benjamins Publishing Company, pp. 39–53. doi: 10.1075/pbns.22.04gum.

Himabindu, G. S. S. N., Rao, R. & Sethia, D. (2022) A self-attention hybrid emoji prediction model for code-mixed language: (Hinglish). *Social Network Analysis and Mining.* 12(1), 137. doi: 10.1007/S13278-022-00961-1.

Kelly, R. & Watts, L. (2015) Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships. In: *European Conference on Computer Supported Cooperative Work (ECSCW 2015). Experiences of Technology Appropriation: Unanticipated Users, Usage, Circumstances, and Design*, *20 September 2015, Oslo, Norway*. University of Bath. pp. 1-7.

Kutsuzawa, G., Umemura, H., Eto, K. & Kobayashi, Y. (2022) Age Differences in the Interpretation of Facial Emojis: Classification on the Arousal-Valence Space. *Frontiers in Psychology*. 13, 915550. doi: 10.3389/fpsyg.2022.915550.

Mahmud, M. S., Huang, J. Z., Salloum, S., Emara, T. Z & Sadatdiynov, K. (2020) A survey of data partitioning and sampling methods to support big data analysis. *Big Data Mining and Analytics*. 3(2), 85–101. doi: 10.26599/BDMA.2019.9020015.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. (2013) Distributed representations of words and phrases and their compositionality. To be published in *Advances in Neural Information Processing Systems*. 26 [Preprint]. https://doi.org/10.48550/arXiv.1310.4546. [Accessed 25th March 2023].

Neel, L. A. G., McKechnie, J. G., Robus, C. M. & Hand, C. J. (2023) Emoji alter the perception of emotion in affectively neutral text messages. *Journal of Nonverbal Behavior*. 42, 83–97. doi: 10.1007/s10919-022-00421-6.

Robertson, A., Magdy, W. & Goldwater, S. (2020) Emoji skin tone modifiers: Analyzing variation in usage on social media. *ACM Transactions on Social Computing*. 3(2), 1–25. doi: 10.1145/3377479.

Rodrigues, D., Prada, M., Gaspar, R., Garrido, M. V & Lopes, D. (2018) Lisbon Emoji and Emoticon Database (LEED): Norms for emoji and emoticons in seven evaluative dimensions. *Behavior Research Methods.* 50, 392–405. doi: 10.3758/s13428-017-0878-6.

Seyednezhad, S. M. M., Fede, H., Herrera, I. & Menezes, R. (2018) Emoji-word network analysis: Sentiments and semantics. In: *Proceedings of the Thirty-First International Florida Artificial Intelligence Research Society Conference (FLAIRS 2018)*, *21-23 May 2018, Melbourne, Florida, USA*. Washington, DC, USA, Association for the Advancement of Artificial Intelligence. pp. 271-274.

Siddiqui, M. A., Ali, S. A. & Haider, N. G. (2018) Reduced Feature Set for Emotion Based Spoken Utterances of Normal and Special Children Using Multivariate Analysis and Decision Trees. *Engineering, Technology & Applied Science Research*. 8(4), 3228–3233. doi: 10.48084/etasr.2177.

Siever, C. M., Siever, T. & Stöckl, H. (2020) Emoji-text relations on Instagram: Empirical corpus studies on multimodal uses of the iconographetic mode. In: Stöckl, H., Caple. H. & Pflaeging, J. (eds.) *Shifts toward Image-Centricity in Contemporary Multimodal Practices*. London, England, Routledge, pp. 177–203. doi: 10.4324/9780429487965.

Skovholt, K., Grønning A. & Kankaanranta, A. (2014) The communicative functions of emoticons in workplace e-mails: :-). *Journal of Computer-Mediated Communication*. 19(4), 780–797. doi: 10.1111/jcc4.12063.

Tang, Y. & Hew, K. F. (2019) Emoticon, emoji, and sticker use in computer-mediated communication: A review of theories and research findings. *International Journal of Communication.* 13, 2457-2483.

Toet, A. & van Erp, J. B. F. (2019) The EmojiGrid as a tool to assess experienced and perceived emotions. *Psych*. 1(1), 469–481. doi: 10.3390/psych1010036.

Velioğlu, R., Yıldız, T. & Yıldırım, S. (2018) Sentiment analysis using learning approaches over emojis for Turkish tweets. In: *2018 3rd International Conference on Computer Science and Engineering (UBMK), 20-23 September 2018, Sarajevo, Bosnia and Herzegovina*. New York City, USA, Institute for Electrical and Electronics Engineers (IEEE). pp. 303–307. doi: 10.1109/UBMK.2018.8566260.

Yafooz, W. M. S., Hizam, E. A. & Alromema, W. A. (2021) Arabic Sentiment Analysis on Chewing Khat Leaves using Machine Learning and Ensemble Methods. *Engineering, Technology & Applied Science Research*. 11(2), 6845-6848. doi: 10.48084/etasr.4026.

Yeole, V., Chavan, P. V & Nikose, M. C. (2015) Opinion mining for emotions determination. In: *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, *19-20 March 2015, Coimbatore, India*. New York City, USA, Institute for Electrical and Electronics Engineers (IEEE). pp. 1–5. doi: 10.1109/ICIIECS.2015.7192931.