

On the Effect of Prior Knowledge in Text-Based Emotion Recognition

Mohammed A. Almulla

Computer Science Department – Kuwait University

DOI: <https://doi.org/10.51244/IJRSI.2024.1108005>

Received: 23 June 2024; Revised: 03 July 2024; Accepted: 08 July 2024; Published: 27 August 2024

ABSTRACT

Human emotion recognition has significant importance across various applications, including chatbots, customer support, and airport security. Detecting emotions from text still poses big challenges for both humans and machine learning algorithms. Recognizing emotions from facial expressions or audio recordings tends to yield more accurate results than from textual data. In this paper we suggest a novel approach to text-based emotion recognition by supporting the machine learning algorithm with prior knowledge capabilities. Simply put, we propose combining natural language processing and sentiment analysis to build a database of keywords that are frequently associated with human-specific emotions. This database constitutes the prior knowledge that we use to make predictions about the emotion corresponding a given text. The aim is to enhance prediction accuracy by leveraging a database of keywords with fine granularity. The experimental results of testing this approach confirmed that our algorithm achieved a higher accuracy rate when the prior knowledge was introduced. Initially, the machine learning model achieved recognition accuracy of 99.79% on the training subset and an accuracy rate of 79.02% on the testing subset. With the help of the knowledge-driven database, the accuracy rate of the testing subset became 97.85%, which confirms that prior of keywords associated with emotion classes has a great impact on the performance of the text-based emotion recognition algorithm.

Keywords: Text-based emotion recognition, machine learning, prior knowledge, human–computer interaction.

INTRODUCTION

The development and prosperity of information and communications technology (ICT) increased the reliance on human-computer interaction (HCI). Emotions are being used as a communication tool between humans and computers. Usually, humans express their emotions through body movements, facial expressions, speech, text, hand gestures, body posture, and physiological signals. Emotion recognition provides computers with interaction capabilities that consider the user's condition, which in turn enables computers to offer better services [1]. Lately, emotion recognition has garnered considerable attention in the machine learning research community. With the ongoing evolution of applications and online services in the current revolution of artificial intelligence, laptops and mobile devices have become an integral part of our life style. Their efficacy and ability to provide more nuanced responses improve significantly when comprehending human emotions. Various scenarios illustrate this, for example emotion recognition aids clinicians in diagnosing mental illnesses and enhances the provision of welfare services for the elderly. Chatbots equipped with emotion recognition can seamlessly assist individuals in their daily routines. Tailoring conversations based on understanding patients' emotional states can greatly benefit their well-being. Also, recent trends in software engineering encourage software developers to design and implement applications that can quickly adapt to the users' emotional states, especially in the world of game-playing. Additionally, there is a growing interest among researchers in recognizing and classifying human emotions derived facial expressions, voice, texting, physiological signals, and body language [2].

Human emotions are often categorized into joy, sadness, anger, surprise, love, and fear. The ability to discern these emotions from textual content written by individuals holds significant importance across various applications, including chatbots, automated customer support services, and sentiment analysis of customer feedback and reviews. However, identifying emotions from text poses a pivotal challenge, even for humans. For instance, expressions such as "Why don't you ever text me!" can be interpreted as conveying either anger or

sadness, showcasing the ambiguity that machines also face. In [3], the authors argued that it might be easier for humans to emotionally communicate face-2-face (F2F) rather than by using computer-mediated communication (CMC). Consequently, machine learning algorithms still demonstrate varying levels of performance in this area. Presently, many artificial intelligence (AI) researchers argue that recognizing emotions from facial expressions or audio recordings tends to yield more accurate results than recognizing emotions from textual data. Indeed, achieving precise text-based emotion recognition requires careful consideration of several nuances, with context-dependency being particularly critical.

Researchers typically use a unimodal approach for emotion recognition, relying on a single input source like text, video, or audio. While this method can achieve high accuracy in certain situations, it is often prone to errors due to noise and variations in the input data. For example, physiological signals can be disrupted by physical activity and stress, and facial expression recognition can be affected by camera angles and lighting conditions. [4]. Thus, researchers are striving to enhance recognition accuracy by employing machine learning and deep learning algorithms or shifting their attention toward a multimodal emotion recognition approach. As noted by [5], different modalities possess distinct features, and understanding their interrelationships is crucial for accurately classifying complex human emotions.

In this paper, we propose employing natural language processing (NLP) to construct a database containing keywords that are frequently associated with specific human emotions (i.e., context-dependent keywords). This database which we call prior knowledge is then utilized to improve predictions made by the machine learning algorithm. The primary aim in this study is to harness statistical heuristics to build prior knowledge which can be used in a tool capable of discerning emotions from user-generated text. The objective of this research is to empirically examine the effect of prior knowledge on the performance of machine learning (ML) models by comparing the performance of these models with and without prior knowledge.

The paper is organized as follows: Section 1 introduces the topic and sets the context for the research, explaining the significance and objectives of the study. Section 2 delves into the relevant literature, providing a comprehensive overview of previous research in the field. It discusses key findings, methodologies, and gaps that the current study aims to address. Section 3 details the methodology adopted for the research. It includes a description of the dataset used, explaining its sources, characteristics, and any pre-processing steps undertaken. An explanation of the machine learning model employed, including the choice of algorithm, architecture, and training process. Also, it covers the validation process of the predictions, highlighting how prior knowledge was utilized to enhance the accuracy and reliability of the results. Section 4 presents the experimental findings, showcasing the performance of the model through various metrics and visualizations. It includes a thorough analysis of the results, comparing them with existing benchmarks and discussing any anomalies or interesting observations. The last section summarizes the key contributions and outcomes of the research. It reflects on the implications of the findings, acknowledges limitations, and suggests potential directions for future research to build upon the current work.

RELATED WORK

Human emotions can be recognized through different modalities. For instance, unimodal behaviour involves using a single input source such as facial expressions, voice, texting, eye gaze, hand gestures, or body movements [6-8]. Meanwhile, bimodal behaviour integrates any two input sources with the promise to give better performance, such as combining facial expressions and voice, or voice and texting [9-11]. The third modality type is the multimodal behaviour in which the model utilizes multiple input sources, like forming a combination of facial expressions, voice, and text messages [12-14].

Remarkable advancements have been made in the realms of machine learning (ML) and artificial intelligence (AI) in the recent years, particularly after the integration of deep learning with computer vision and natural language processing. This era of data-driven innovation has significantly boosted productivity, ranging from enhancing the predictive capabilities of ML algorithms to optimizing processes in advanced manufacturing. Nevertheless, substantial concerns persist regarding certain inherent limitations of conventional ML methodologies: (a) the necessity for large and often costly datasets and (b) challenges related to generalizability and interpretability.

To address these limitations, researchers have experimented with incorporating knowledge into these tasks. This knowledge may encompass both scientific insights and human expertise. For instance, Li et al. [15] introduced knowledge-driven approaches to machine learning-based channel estimation within massive multiple-input multiple-output (MIMO) systems. One year later, the same group applied knowledge-driven machine learning to wireless communications [16]. Similarly, Hussain et al. [17] devised a knowledge-driven machine learning framework for predicting early-stage disease risks in critical environments. In 2023, Qian et al. [18] proposed a knowledge-driven approach to learning, optimization, and experimental design under uncertainty, particularly aimed at the discovery of materials.

As for emotion recognition from text, we found a few studies in the literature that applied domain knowledge to the emotion recognition problem. For example, detecting emotions in casual, everyday conversations presents a big challenge, as the emotions expressed by speakers are often less overt than those in scripted or acted speech. It is noted that human annotators exhibit considerable disagreement when annotating spontaneous speech. However, this disagreement decreases significantly when they are given supplementary knowledge pertaining to the nature of the conversation. This observation has inspired an emotion recognition framework that was suggested by Chakraborty et al. [19]. The proposed framework took as input existing domain knowledge. Using this knowledge, the framework demonstrated a significant improvement in recognizing the emotion of the speaker.

Qi et al. [20] explored video highlight detection by identifying key frames based on users' specific interests, recognizing that these interests are significantly influenced by subjective emotions. They developed an emotion-driven video detection framework to model human emotions and assess highlight significance. First, they extracted concept representations from video clips and built an emotion-based knowledge graph. They modeled relationships within this graph using external public knowledge graphs (EPKGs). Then, they utilized graph convolutional networks (GCNs) to capture dependencies between nodes and facilitate information flow. Finally, they created an emotion-aware video representation from the GCN layers to predict highlight emotions. Their method outperformed similar state-of-the-art approaches in experiments on two benchmark datasets.

Teng et al. [21] proposed an emotion recognition conversation system based on an automatic knowledge database architecture. Using a knowledge base that the system constructs automatically, their system recognizes the emotions within conversation contents. Experiments results showed that their method achieved better results in practice. In the current work, we incorporate deep domain knowledge into text-based emotion recognition with a novel ML model to forward this research area.

METHODOLOGY

One can utilize prior knowledge in emotion recognition in several ways. For example, one can integrate knowledge graphs or ontologies [22]. These structured representations of knowledge capture relationships between concepts, which can be leveraged to enhance emotion recognition systems. Another method is to explore the use of ontologies to model emotions, their causes, and their expressions [23]. By incorporating this prior knowledge into deep learning models, such as machine learning architectures, the systems can better understand the context of emotions, leading to more accurate recognition [24]. Yet another approach is the incorporation of common-sense knowledge [25]. Emotions are often influenced by everyday situations and social interactions, which can be captured using common-sense reasoning techniques. By integrating common-sense knowledge bases into emotion recognition systems, researchers aim to improve the systems' ability to understand and interpret emotions in context.

Focusing on transferring knowledge from related domains is another possibility. For instance, knowledge learned from natural language processing tasks, such as sentiment analysis or opinion mining, can be transferred to emotion recognition tasks. This transfer learning approach enables the model to leverage existing knowledge and adapt it to the specific nuances of emotion recognition. To summarize, these strategies highlight the potential of incorporating prior knowledge from various sources, such as knowledge graphs, ontologies, commonsense reasoning, and related domains, to enhance the performance of emotion recognition systems and provide deeper insights into human emotions.

In this study, we employ the knowledge transfer path from natural language processing and sentiment analysis to emotion recognition from text. Hence, our study scrutinizes and enhances the resilience of text-based emotion recognition, introducing a novel methodology that augments conventional machine learning predictions with prior emotional knowledge. The aim of this research is to enhance the intelligence and user-friendliness of the machines we interact with every day. By equipping these machines with the ability to understand human emotional states and respond appropriately, the research seeks to create more intuitive and empathetic interactions. This involves developing advanced algorithm that can accurately detect and interpret human emotions based on text messages. By integrating these capabilities, machines will be able to adapt their responses in real-time, offering a more personalized and engaging user experience. Ultimately, this research aspires to make our daily interactions with technology more seamless, efficient, and emotionally attuned, significantly improving the overall user experience. This can be achieved through the following four phases:

The Emotion Dataset

Text-based emotion recognition is a core challenge of content-based classification, drawing upon principles from natural language processing (NLP) and machine learning. Consequently, this study proposes a machine learning approach—supported by semantic text analysis—for text-based emotion recognition in large datasets. Various prominent datasets have been developed and deployed to facilitate research and make experimental advancements in this domain. For example, the 2019 SemEval dataset consists of the “emotion_data.csv” and “emotion_data_prep.csv” datasets encompassing, respectively, 55,774 tweets and 62,015 tweets sourced from Twitter. The tweets in both files are subjected to various preprocessing steps, such as lowercasing, tokenization, and lemmatization, as well as the removal of mentions, URLs, punctuation marks, and stop words. Both datasets are annotated with five emotion classes: neutral, happy, sad, love, and anger [26].

The Essay dataset [27] is another standard benchmark dataset used for solving the text-based emotion recognition problem. Essay is based on the stream of consciousness collected by Pennebaker and Laura King from 2,468 daily writing submissions from 34 psychology students, spanning a wide age range. Additionally, the ISEAR.csv dataset (International Survey on Emotion Antecedents and Reactions) features records covering seven major emotions: joy, fear, anger, sadness, disgust, shame, and guilt, totaling 7,516 entries. Lastly, the EmoBank repository is a comprehensive text corpus meticulously annotated with emotions based on the psychological valence–arousal–dominance (VAD) model. Developed at the JULIE Lab at Jena University, this resource is detailed in [28-29]. The VAD model categorizes emotions along three dimensions: valence (pleasantness), arousal (intensity of emotion), and dominance (control). EmoBank provides valuable data for researchers exploring the intersection of language and emotion. The repository is organized into two main folders Corpus and Pilot. The latter includes the pilot data, serving as an initial dataset for preliminary studies and methodological validation. By offering a robust, annotated dataset, EmoBank facilitates advancements in natural language processing, sentiment analysis, and emotional AI, enabling machines to better understand and respond to human emotions in text.

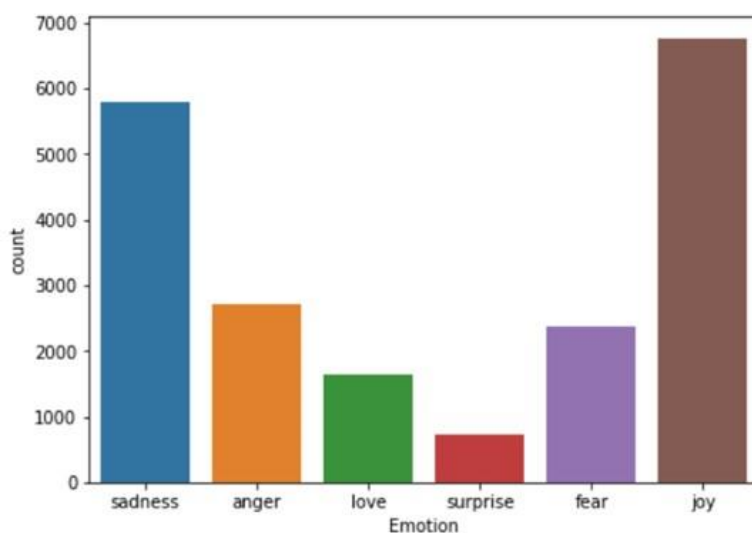


Fig. 1: Emotion class distribution in the training dataset.

In this work, we used a publicly available dataset from Kaggle’s Hugging Face website [30]. This Emotion dataset contains three files containing tweets with their corresponding emotion labels. The first file is “train.txt,” which contains 16,000 tweets embedding emotions for training. The second file is “test.txt,” which contains 2,000 tweets for testing our machine learning algorithm. The third file is “val.txt,” which contains 2,000 tweets for validating the performance of our machine learning algorithm. In all files, each tweet is labeled with one of six emotions: sadness (0), joy (1), love (2), anger (3), fear (4), and surprise (5). The dataset can be accessed online following the datalink <https://huggingface.co/datasets/dair-ai/emotion>. Figure 1 shows the number of tweets labeled with each of the six emotion classes.

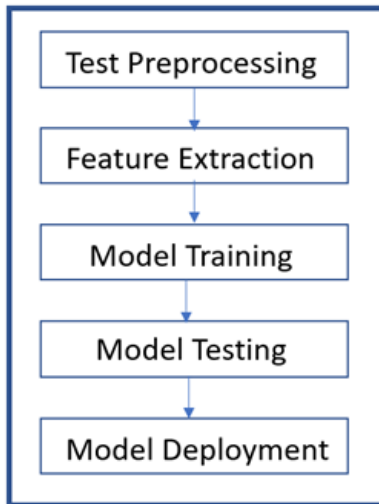


Fig. 2. The data acquisition and preparation procedure.

The data acquisition and preparation undergoes a procedure shown in Figure 2. It involves the following steps: (a) Text preprocessing: stop word removal, punctuation removal, and extraction of the most common words for each emotion. (b) Feature extraction: using CountVectorizer to convert the cleaned text into numerical features that can be used for machine learning. (c) Model training: training the ML model on the training dataset. (d) Model testing: testing the ML model using the accuracy index. Interpretation of the model was performed using eli5. (e) Model deployment: saving the final trained model for future invocation.

The Prior Knowledge Database

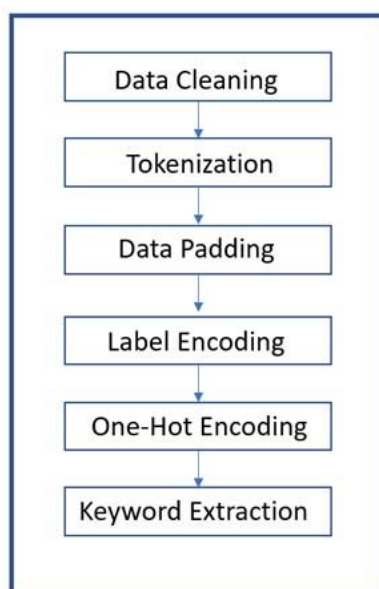


Fig. 3. Construction of the prior knowledge database.

Building the prior knowledge database is illustrated in Figure 3. It involves the following six key steps: (a) Data cleaning: We clean the lines of text by performing the following operations: (1) convert the text to lowercase,

(2) split the text into tokens (3) reduce the tokens to their stemmed version (4) eliminate mentions and hashtags, (5) eliminate URLs (6) eliminate punctuation marks (7) eliminate stop words, and (8) eliminate duplicate words. Figure 4 shows the nine steps of data cleaning. (b) Tokenization: We tokenize the cleaned textual data, breaking the sentences down into individual words or tokens. (c) Padding: To ensure compatibility with the neural network, pad the sequences to a uniform length, allowing them to be processed simultaneously. (d) Label Encoding: Convert the emotion classes from strings to integers using the label encoder. This conversion is essential because machine learning algorithms require numerical input. (e) One-Hot Encoding: Transform categorical labels into a binary format, where each label is represented as a vector containing all zeros except for a single 1 at the index corresponding to that label. (f) Keyword extraction: Emotional keywords are extracted from the tokens of each emotion class and stored in a separate list (i.e., databases). These lists of keywords corresponding to the emotion classes constitute the databases, which play a crucial role in classifying the emotions in the validation data. To build the prior knowledge database, we used the 2000 tweets with their labelled emotion classes in the validation file. The database consists of six lists, one for each recognizable emotion class by the system. Each list holds the keywords that are frequently associated with the labelled list index.

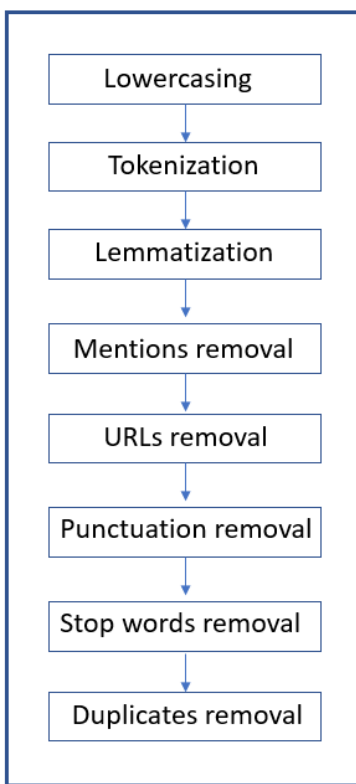


Fig. 4. Input data cleaning steps.

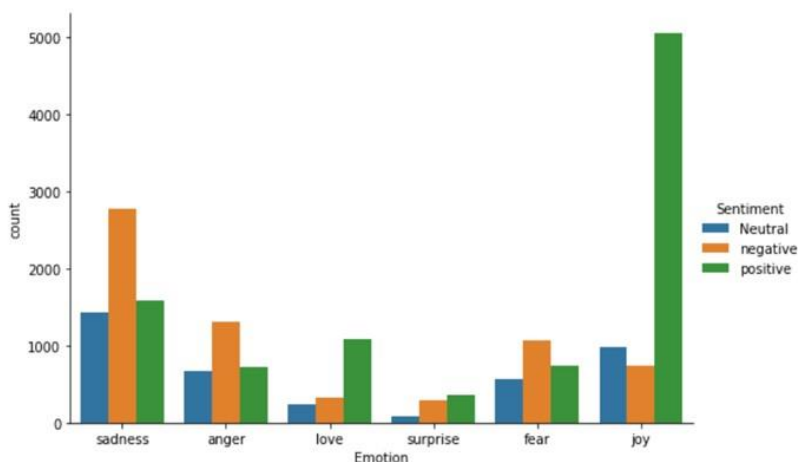


Fig.5 Sentiment analysis results of the training and testing data.

The TextBlob library in Python allows the user to do sentiment analysis on the input data. The outcome of the analysis can be one of three folds (1) positive if the sentiment is greater than zero, (2) negative if the sentiment is less than zero, and (3) neutral if the sentiment is equal zero. We ran the sentiment analysis on the training and testing data. Figure 5 reflects the results of this analysis. For each emotion class, we count the number of positive, negative and neutral sentiments. One can see that the negative tweets for the sadness, anger and fear emotion classes outnumber the positive and neutral tweets. Similarly, the positive tweets for the joy, love and surprise emotion classes outnumber the negative and neutral tweets.

The Machine Learning Model

The neural network architecture of our model is defined using the Python Keras library which is part of the Tensorflow API. Our network uses the sequential model, which comprises an embedding layer, a flatten layer, and two dense layers with two activation functions namely “relu” and “softmax”. The model is compiled with the Adam optimizer and the “categorical_crossentropy” loss function. We split the 20000 lines of input in the Emotion dataset into a training subset (80% = 16000 tweets) and the test and validation subsets (20% = 4000 tweets, 2000 tweets each). The validation step is used to track performance improvement after the training and testing of the model.

Prediction Improvement

Once the machine learning model is trained and tested, we construct the prior knowledge database from the lines of text in the validation data subset. Upon completion of the construction procedure, we train and test a multinomial naive Bayes (NB) model using the keywords stored in the prior knowledge database. Next, the multinomial NB model predicts the emotion classes of the validated lines that were misclassified by the original machine learning model. Unless the accuracy rate of the ML model is 100%, there will be some misclassified emotions from the validation file. Assume that μ is the set of misclassified statements by the original model. For every statement $s \in \mu$, its labeled emotion class is listed in the validation file (let us call this emotion class E_s). We now apply the NLP steps discussed above to the statement s to extract the most important keywords in s (call this set of keywords K_s).

After training and testing the multinomial model on the statements in the dataset, we ask the model to predict the classes of all the statements in μ . We calculate the accuracy percentage of the correctly classified statements and add this rate to the accuracy rate obtained in testing the ML model. This gives the accuracy rate of our overall approach. Figure 6 shows the accuracy and loss curves obtained by training and testing the multinomial NB model. Since this model is trained and tested on optimized prior knowledge (i.e., databases of emotional classes), this will surely increase the prediction accuracy of the overall proposed approach.

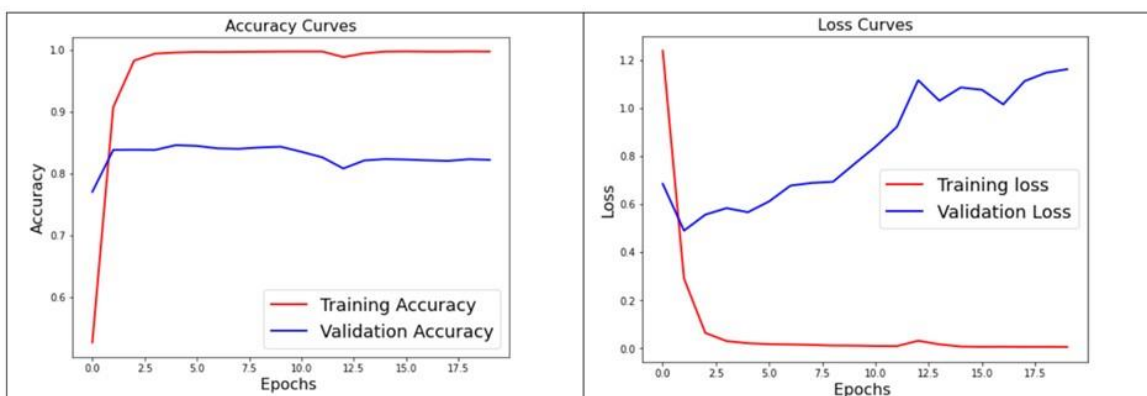


Fig. 6. Training the multinomial NB model using the Adam optimizer.

EXPERIMENTAL RESULTS

We used the accuracy evaluation metric to assess the performance of our approach. The machine learning model was trained multiple times on different number of epochs whereas the learning rate is fixed at 0.001. The experiments were conducted on a laptop running Windows 11 operating system, and equipped with a 13th

Generation Intel Core i7–1355U Processor (1700 MHz, 10 cores, 12 logical processors), 16 GB of RAM, and an NVIDIA GeForce MX-550 graphics card. Our neural network architecture consists of a sequential model with three layers. We employed categorical cross-entropy as the loss function, which is well-suited for multi-class classification problems. To optimize the training process, we used the Adam optimizer, which is known for its efficiency and ability to handle sparse gradients on noisy problems. This combination of hardware and parametric choices ensured robust performance and accurate evaluation of our machine learning approach.

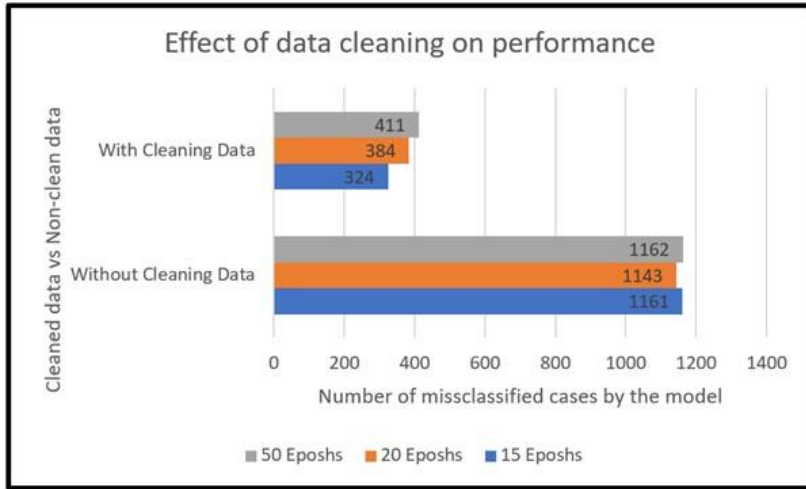


Fig. 7. Effect of data cleaning on the emotion recognition model.

The Emotion dataset that we used as a benchmark to test our approach contains 20,000 sample text sentences (i.e. tweets) divided into training data (train_size = 16000), test data (test_size = 2000), and validation data (validation_size = 2000). We tested our approach twice, with and without cleaning the data. Figure 7 shows the effect of cleaning the data on the performance of the sequential machine learning model. Moreover, to determine the accuracy measure for the proposed approach, we tested the neural network on the 2,000 lines of text in the validation data file. Without cleaning the data (epochs = 15, 20, 50), the model misclassified, respectively, 1161, 1143, and 1162 cases out of a total of 2000 test cases. On the other hand, with cleaning the data (epochs = 15, 20, 50), the model misclassified, respectively, 324, 384, and 411 cases out of the 2000 test cases.

With the help of the prior knowledge database, the old machine learning model misclassified 66 cases out of the 2,000 validation cases, which means that the model’s validation accuracy rate is 96.7% of the cases. For our prior knowledge dependent approach, we examined two predetermined classification models (logistic regression and multinomial NB) and adopted the one with the highest correct classification score. Out of the 66 cases that were misclassified by the ML model, the logistic regression classifier correctly classified 22 cases, whereas the multinomial NB model correctly classified 23 cases, leaving only 43 misclassified cases among the 2000 validation tweets. Thus, out of the 2000 tweets in the validation file, our approach misclassified only 43 tweets (i.e. correctly classified 1957 tweets), which means the accuracy rate of our approach $(1957/2000) \times 100\% = 97.85\%$. Figure 8 shows the confusion matrix and the classification report of the multinomial NB model.

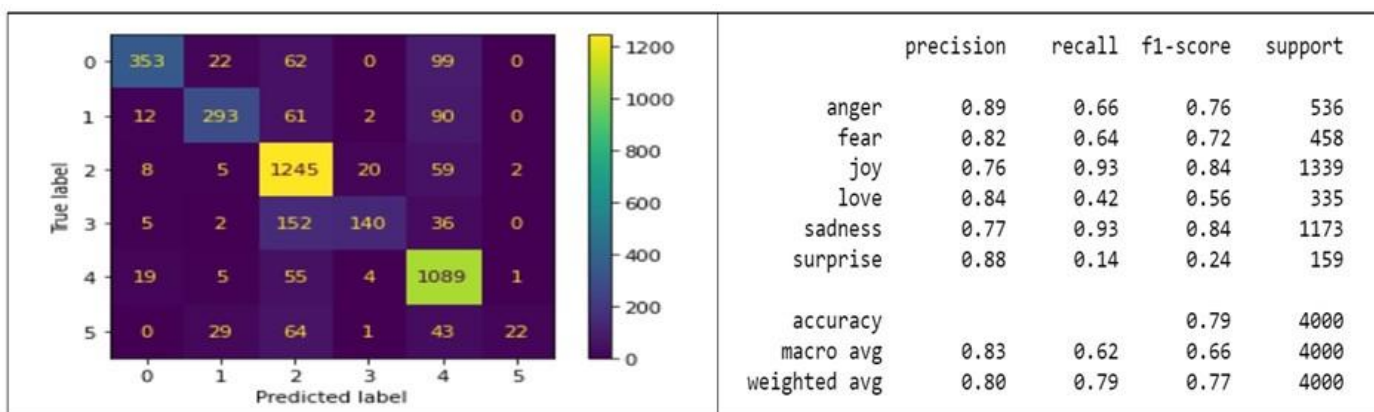


Fig. 8. Confusion matrix and classification report of the multinomial NB model.

CONCLUSION AND FUTURE WORK

In this work we demonstrate how prior knowledge may positively influence the performance of machine learning classification algorithm. First, we introduced a machine learning model designed to recognize and classify human emotional states expressed in textual data. The model achieved a recognition accuracy of 99.79% on the training subset and an accuracy rate of 79.02% on the test subset. To enhance the model's predictive capabilities, we augmented a multinomial Naïve-Bayes model with a prior knowledge database containing keywords associated with distinct emotional categories. This step boosted the recognition accuracy of our approach to reach 97.85%, which confirms that prior knowledge of keywords associated with emotion classes has a great impact on the performance of the machine learning algorithm. Currently, our approach disregards the analysis of user personality traits, which significantly influence emotional interpretations, encompassing traits such as openness, conscientiousness, extraversion, agreeableness, and neuroticism. In the future, to enhance our understanding of users' emotional fluctuations, we intend to incorporate the analysis of personality traits into the machine learning algorithm. This initiative may enrich the user experience and improve the accuracy of emotion recognition capabilities.

ACKNOWLEDGES

This work was conducted in the Artificial Intelligence Laboratory, one of the computing facilities that belong to the Computer Science Department in the Faculty of Science at Kuwait University. The author expresses his gratitude to all individuals affiliated with the Artificial Intelligence Laboratory at the Computer Science Department of Kuwait University for their valuable assistance and support throughout this endeavor. Additionally, he extends his appreciation to the open-source community and the developers behind diverse Python libraries and machine-learning algorithms, whose contributions were instrumental in facilitating this research.

REFERENCES

1. Ren, F. (2009). Affective Information Processing and Recognizing Human Emotion, *Electronic Notes in Theoretical Computer Science* 225 (2009) 39–50.
2. Abdulsalam, W.H., Alhamdani, R.S., and Abdullah, M.N., (2019). Emotion Recognition System Based on Hybrid Techniques, *International Journal of Machine Learning and Computing*, 9(4): 490-495.
3. Derks, D., Fischer, A., and Bos, A. (2008). The role of emotion in computer-mediated communication: A review, *Computers in Human Behaviour*, 24(3):766-785. <https://doi:10.1016/j.chb.2007.04.004>.
4. Sun, Q., Liang, L., Dang, X., and Chen, Y. (2022). Deep learning-based dimensional emotion recognition combining the attention mechanism and global second-order feature representations, *Computers and Electrical Engineering*, 104(B), 108469, <https://doi.org/10.1016/j.compeleceng.2022.108469>.
5. Mehta P. (2018). Multimodal deep learning fusion of multiple modalities using deep learning. Source: (<https://towardsdatascience.com/multimodal-deep-learning-ce7d1d994f4>).
6. Bagwan, R., Chintawar, S., Dhapudkar, K., Balamwar, A., and Gore, S. (2021). Facial Emotion Recognition using Convolution Neural Network, *International Journal of Trend in Scientific Research and Development*, 5(3): 800-801.
7. Parlak, C. and Diri, B. (2013). Emotion recognition from the human voice, In the Proceeding of the IEEE Signal Processing and Communications Applications Conference (SIU), <https://doi:10.1109/SIU.2013.6531196>.
8. Rashid, M., Abu-Bakar, S.A.R., and Mokji, M. (2013). Human emotion recognition from videos using spatio-temporal and audio features. *The Visual Computer*, 29: 1269–1275.
9. Kumar, P., Malik, S. and Raman, B. (2023). Interpretable multimodal emotion recognition using hybrid fusion of speech and image data. *Multimed Tools Appl.* <https://doi.org/10.1007/s11042-023-16443-1>.
10. Xu, Y., Su, H., Ma, G., and Liu, X. (2023). A novel dual-modal emotion recognition algorithm with fusing hybrid features of audio signal and speech context. *Complex & Intelligent Systems*. 9: 951–963, <https://doi.org/10.1007/s40747-022-00841-3>.
11. Ma, Y., Hao, Y., Chen, M., Chen, J., and Lu, P. (2019). Košir, A. Audio-visual emotion fusion (AVEF):

- A deep efficient weighted approach. *Inf. Fusion*, 46, 184–192. 13.
12. Hayes, Th., Zhang, S., Yin, X., Pang, G., Sheng, S., Yang, H., Ge, S., Hu, Q., and Parikh, D. (2022). MUGEN:A Playground for Video-Audio-Text Multimodal Understanding and GENERation, *European Conference on Computer Vision*, pp. 431–449.
 13. Poria, S., Cambria, E., Howard, N., Huang, G.B., and Hussain, A., (2016). Fusing audio, visual and textual clues for sentiment analysis from multimodal content, *Neurocomputing*, 174(A): 50-59.
 14. Almulla, M. A. (2024). A multimodal emotion recognition system using deep convolution neural networks, *Journal of Engineering Research*, Available online 27 March 2024. <https://doi.org/10.1016/j.jer.2024.03.021>.
 15. Li, D., Xu, Y., Zhao, M., Zhang, S., and Zhu, J. (2021) "Knowledge-Driven Machine Learning-based Channel Estimation in Massive MIMO System," *IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, Nanjing, China, pp. 1-6, <https://doi.org/10.1109/WCNCW49093.2021.9420022>.
 16. Li, D., Xu, Y., Zhao, M., Zhu, J., and Zhang, S. (2022) "Knowledge-Driven Machine Learning and Applications in Wireless Communications," in *IEEE Transactions on Cognitive Communications and Networking*, 8(2): 454-467, <https://doi.org/10.1109/TCCN.2021.3128597>.
 17. Hossain, M.A., Ferdousi, R., and Alhamid, M.F. (2020). Knowledge-driven machine learning based framework for early-stage disease risk prediction in edge environment, *Journal of Parallel and Distributed Computing*, 146: 25-34, <https://doi.org/10.1016/j.jpdc.2020.07.003>.
 18. Qian, X., Yoon, B-J., Arróyave, R., Qian, X., and Dougherty, E.R. (2023). Knowledge-driven learning, optimization, and experimental design under uncertainty for materials discovery, *Patterns*, 4(11): 100863, <https://doi.org/10.1016/j.patter.2023.100863>.
 19. Chakraborty, R., Pandharipande, M., and Kopparapu, S.K. (2016). Knowledge-based Framework for Intelligent Emotion Recognition in Spontaneous Speech, *Procedia Computer Science*, 96: 587-596, <https://doi.org/10.1016/j.procs.2016.08.239>.
 20. Qi, F., Yang, X., Yang, X., and Xu, Ch. (2020). Emotion Knowledge Driven Video Highlight Detection, *IEEE Transactions on Multimedia*, (99):1-1, <https://doi.org/10.1109/TMM.2020.3035285>.
 21. Teng, Z., Ren, F., and Kuroiwa, S. (2007). An Emotion Recognition Conversation System Based on Knowledge Database Automatic Architecture, *International Conference on Intelligent Computing (ICIC-2007): Advanced Intelligent Computing Theories and Applications with Aspects of Contemporary Intelligent Computing Techniques*, pp 722–731. *Communications in Computer and Information Science*, vol 2. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-540-74282-1>.
 22. Zhang, D. Chen, X., Xu, Sh. and Xu, B., (2020). Knowledge Aware Emotion Recognition in Textual Conversations via Multi-Task Incremental Transformer, *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 4429–4440, Barcelona, Spain, December 8-13, 2020.
 23. Zhang, X., Hu, B., Chen, J. and Moore, Ph. (2013). Ontology-based context modeling for emotion recognition in an intelligent web. *World Wide Web* 16:497-513. <https://doi.org/10.1007/s11280-012-0181-5>.
 24. Mumuni, F. and Mumuni, A. (2024). Improving deep learning with prior knowledge and cognitive models: A survey on enhancing explainability, adversarial robustness and zero-shot learning, *Cognitive Systems Research*, 84: 101188, <https://doi.org/10.1016/j.cogsys.2023.101188>.
 25. Chen, J., Yang, T., Huang, Z. and Wang, K. (2023). Incorporating structured emotion commonsense knowledge and interpersonal relation into context-aware emotion recognition. *Applied Intelligence* 53(6):1-17. <https://doi.org/10.1007/s10489-022-03729-4>.
 26. Alfred, M. (2019). The semeval-data-set-2019, *Proceedings of the 13th International Workshop on Semantic Evaluation*, <https://kaggle.com/datasets/maxjohnalfred/semeval-data-set-2019>.
 27. Pennebaker, J.W. and King, L. A. (1999). Linguistic styles: language use as an individual difference. *Journal of personality and social psychology*, 77(6): 1296-1312. <https://doi.org/10.1037/0022-3514.77.6.1296>.
 28. Buechel, S. and Hahn, U. (2017). EmoBank: Studying the Impact of Annotation Perspective and Representation Format on Dimensional Emotion Analysis. In *EACL 2017 - Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. Valencia, Spain, April 3-7, 2017. Short Papers, pages 578-585. Available: <http://aclweb.org/anthology/E17-2092>.

29. Buechel, S. and Hahn, U. (2017). Readers vs. writers vs. texts: Coping with different perspectives of text understanding in emotion annotation. In LAW 2017 - Proceedings of the 11th Linguistic Annotation Workshop @ EACL 2017. Valencia, Spain, April 3, 2017, pages 1-12. Available: <https://sigann.github.io/LAW-XI-2017/papers/LAW01.pdf>.
30. Saravia, E., Liu, H. C. T., Huang, Y. H., Wu, J., and Chen, Y. S. (2018). Carer: Contextualized affect representations for emotion recognition. In Proceedings of the 2018 conference on empirical methods in natural language processing, pp. 3687–3697.