Emotion Recognition on Social Media Using Natural Language Processing (NLP) Techniques

Romero Gomez Luis
School of Computing, Engineering & the Build Environment, Edinburgh Napier University, Edinburgh EH10 5DT, United Kingdom
40408203@live.napier.ac.uk

Tess Watt
School of Computing, Engineering & the Built Environment, Edinburgh Napier University, Edinburgh EH10 5DT, United Kingdom
t.watt@napier.ac.uk

Kehinde O. Babaagba
School of Computing, Engineering & the Built Environment, Edinburgh Napier University, Edinburgh EH10 5DT, United Kingdom
K.Babaagba@napier.ac.uk

Christos Chrysoulas*
School of Computing, Engineering & the Built Environment, Edinburgh Napier University, Edinburgh EH10 5DT, United Kingdom
C.Chrysoulas@napier.ac.uk

Aydin Homay
Computer Science, TU Dresden
aydin.homay@mailbox.tu-dresden.de

Raghuraman Rangarajan
Nomadcrow Informatica, Porto, Portugal
raghu@pronomadic.com

Xiaodong Liu
School of Computing, Engineering & the Built Environment, Edinburgh Napier University, Edinburgh EH10 5DT, United Kingdom
x.liu@napier.ac.uk

ABSTRACT

In recent years, text has been the main form of communication on social media platforms such as Twitter, Reddit, Facebook, Instagram and YouTube. Emotion Recognition from these platforms can be exploited for all sorts of applications. Through the means of a review of the current literature, it was found that Transformer-based deep learning models show very promising results when trained and fine-tuned for emotion recognition tasks. This paper provides an overview of the architecture for three of the most popular Transformer-based models, BERT Base, DistilBERT, and RoBERTa. These models are also fine-tuned using the "Emotions" dataset; a data corpus composed of English tweets annotated in six (6) different emotions, and the performance of the models is evaluated. The results of this experiment showed that while all of the models demonstrated excellent emotion recognition capabilities by obtaining over 92% F1-score, DistilBERT could be trained in nearly half of the time compared to the other models. Thus, the use of DistilBERT for emotion recognition tasks is encouraged.

*Corresponding author

ICISS 2023, August 11–13, 2023, Edinburgh, United Kingdom

ACM ISBN 979-8-4007-0820-6/23/08.
https://doi.org/10.1145/3625156.3625173

CCS CONCEPTS

• **Information systems**; • **Computing methodologies** → Artificial Intelligence; Natural language processing;

KEYWORDS

Emotion Recognition, Transformer Based Models, BERT Base, DistilBert, RoBERTa, NLP

ACM Reference Format:
Romero Gomez Luis, Tess Watt, Kehinde O. Babaagba, Christos Chrysoulas, Aydin Homay, Raghuraman Rangarajan, and Xiaodong Liu. 2023. Emotion Recognition on Social Media Using Natural Language Processing (NLP) Techniques. In *2023 The 6th International Conference on Information Science and Systems (ICISS 2023), August 11–13, 2023, Edinburgh, United Kingdom.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3625156.3625173

1 INTRODUCTION

Nowadays, communication on social media platforms such as Twitter, Reddit, and Facebook have predominantly been text-based. This includes those on secondary communication platforms such as Instagram and YouTube. Emotion Recognition from these platforms can be applied to fields such as marketing, recommendation systems, education, health and mental wellbeing, law enforcement among others. To process and analyze such enormous amounts of data and ultimately extract emotions, there exist several types of techniques such as Neural Networks, Natural Language Processing (NLP), Deep Learning (DL), Parallel computing, and so much more. However, emotion recognition also faces many challenges in both language abstraction and computation, such as accuracy of emotion detection, fuzzy boundaries, grammatical mistakes, multiple and variety of emotions, sentence structure, efficiency, casual writing, to

mention a few [2]. One of the major problems in emotion classification seems to be the lack of a general consensus on what categories or basic emotions to use for these tasks.

One of the most prominent theories of the 20th century is Robert Plutchik's wheel of emotions [10]. In his paper, Plutchik proposes a dimensional model comprising eight basic emotions: joy, sadness, trust, disgust, fear, anger, surprise, and anticipation. He further explains these basic emotions can overlap and combine with each other to produce secondary and even tertiary emotions, like the different hues and shades in a color wheel. For example, fear and surprise could result in "awe", while the combination of joy and anticipation might produce "optimism". On the other hand, [6] proposes a classification model which measures and evaluates the movements of facial muscles as well as those of the eyes and head. Based on his theory, [6] proposed a discrete and categorical approach to the basic emotions problem, suggesting that there only exists six of them and they are shared across cultures and countries: joy, sadness, surprise, fear, anger, and disgust. He reached this conclusion after experimenting and analyzing human facial expressions.

This paper aims to explore the current emotion recognition methods within the NLP techniques and compare state-of-the-art NLP deep learning models to reach a final recommendation. Therefore, the main contribution of this work is twofold and includes: 1) the implementation of the proposed emotion recognition models using an appropriate dataset for this task and 2) a justified and factual comparison between the results of testing the models implemented.

The rest of the paper is structured as follows: Section 2 provides a background of the research and reviews related work. Section 3 provides a detailed overview of the models chosen to be implemented. In Section 4 the experiments and results are discussed. Finally, Section 5 concludes the paper and puts forward areas for future research.

2 RELATED WORK

Emotion recognition has gained increased attention over the past years and has even been described as the key component in human nature. Textual emotion recognition is in high demand and has been increasingly promoted in the NLP and Artificial Intelligence (AI) fields because of its wide range of applications and interests [3].

Emotion recognition on social media carried out during a given situation can help to understand and predict people's feelings towards an ongoing event or incident; [1] conducted a study on two different case scenarios where they used emotion analysis on tweets to successfully predict the outcome of a general election result. Emotion analysis can also find application in areas of new interest such as online learning, where considering emotional aspects of the learners and combining these with more traditional analytical data can help to offer a complete view of the learning experience and fine tune it, depending on the learner's profile [14].

[9] describes how the advancements made on Convolutional Neural Networks (CNN) for computer vision inspired the future of NLP. In CNNs, pre-trained models were used to initialize more complex and deeper models, this translated into using word vectors to transfer information from large amounts of unlabeled data for NLP models. To demonstrate this, they developed CoVe, an encoder pre-trained on supervised machine translation, which performed better in general NLP tasks than previous baseline models that were initialized by using random word vectors.

As a result of these developments and the use of transformer-based models, OpenAI released their own language model known as Generative Pre-trained Transformer (GPT) [12] \cite{radford2018improving} which would surpass the performance of state-of-the-art discriminatively trained models in a wide range of benchmarks. Further improved versions of the GPT model were then released known as GPT-2(2019) and GPT-3(2020). Another pre-trained model that is highly regarded in the NLP field and that surged after the advancements made following GPT is Bidirectional Encoder Representations from Transformers (BERT), presented in the study by [4]. Unlike the previously cited models, BERT is designed to pre-train deep bidirectional representations from text by jointly conditioning on both left and right context in all layers. BERT also offers the possibility of fine-tuning, meaning that the base model can be "re-trained" for multiple other text classification, text generation, and question-answering tasks. This essentially enabled BERT and its derived versions to be the current state-of-the-art language understanding model [4], [11], and to set a new standard for natural language processing [3].

Furthermore, emotional word embedding was introduced for emotion recognition in text. Emotional word embedding was initially inspired by typical word embedding models as a way of representing emotional knowledge for emotion recognition and sentiment analysis tasks. In their paper [16], they proposed a model called Emo2Vec to encode emotional semantics into vectors. This model was pre-trained in various emotion-related tasks that included emotion and sentiment analysis among others. Overall, Emo2Vec is proven to offer very competitive performance when combined with other word embedding models. While many NLP models rely on pre-trained word embeddings, [5] highlight the importance of the emoji in social media and electronic messaging and the lack of studies and models capable of processing emoji representations. In their paper, Emoji2Vec was proposed to offer a solution to this problem, an NLP model pre-trained on emoji embeddings that can be used alongside other models such as Word2Vec for sentiment analysis tasks. DeepMoji was then proposed in [7] as a tool for emotion analysis models to learn richer representations through emoji prediction. It is pre-trained on a dataset of 1246 million tweets and utilizes a Bidirectional Long Short-Term Memory (BiLSTM) network to estimate the emotional content of text through emoji labelling. This study shows how emojis can be used to classify emotional content accurately and clarify seemingly ambiguous and sarcastic sentences.

3 MODEL ARCHITECTURE

The Transformer architecture [15] has brought great advancements to NLP, language modelling, and language understanding. BERT [4], short for Bidirectional Encoder Representations from Transformers, is considered to be the best-known model based on this architecture. A general overview of the Transformer architecture is provided below to gain a better understanding of how BERT operates. The Transformer approach uses the attention mechanism, essentially a

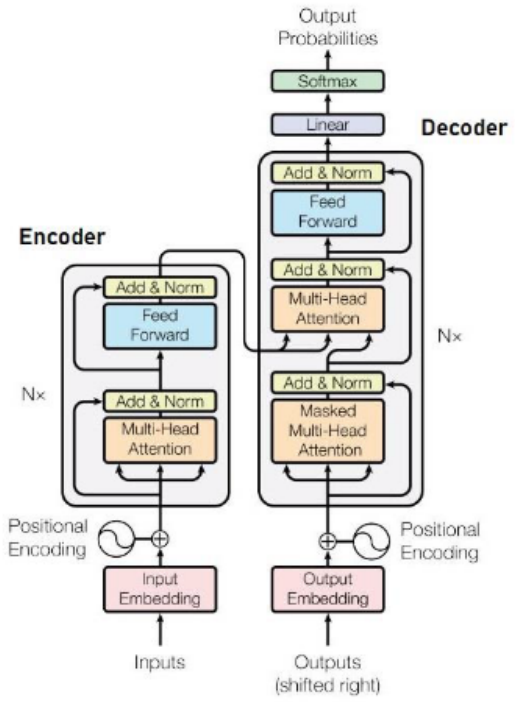

Figure 1: The Transformer Model Architecture [15].

deep-learning algorithm, to understand the existing relationships between words. It uses this mechanism to create different weights associated with each input word, indicating which ones in a given sentence provide the most critical information.

A Transformer model operates by stacking multiple layers of Encoders and Decoders, both using a self-attention algorithm as seen in Fig. 1. Encoder layers aim to provide an understanding of the language and context, by utilizing input word embedding. These embeddings encapsulate the meaning of the words by giving similar vector values to those words with synonymous meanings. Decoder layers take the resulting embeddings from the encoders and combine these with the desired outputs to produce the output probabilities and predictions, in other words, the decoder is responsible for establishing the relationships between inputs and outputs.

Fundamentally, BERT consists of a stack of Encoder layers extracted from the Transformer architecture and it is also worth noting that there are two main versions of BERT available: BERT Base, using 12 Transformer Encoder layers, and BERT Large, using 24 of these layers. For the experimental work done in this paper, the base model is used due to its reduced computational requirements and training times. The implementation of the BERT model (and its variants) usually consists of two phases: pre-training and fine-tuning as seen in Fig. 2. This is what allows a base BERT model to be applied to very diverse NLP tasks as opposed to Transformers, which are most generally used for language translation.

DistilBERT is a general-purpose language representation model proposed in [13], that can be fine-tuned on a wide range of language processing tasks while offering good performances. DistilBERT is trained by distilling the BERT base model, reducing its size by 40% compared to the original model, while claiming to retain 97% of its language understanding capabilities and being 60% faster [13]. It is essentially described as being the smaller, faster, cheaper to pre-train, and lighter version of BERT.

To achieve this, DistilBERT is trained by applying a technique known as knowledge distillation. Knowledge distillation is a compression method in which a student model (in this case, DistilBERT) is trained to reproduce the behaviour of a larger model or teacher (BERT). For the scope of this paper, knowledge distillation will not be explored further, but more information can be found in the original paper [13]. DistilBERT has generally the same architecture as BERT but has the initial segment embedding layer and a final pooling layer removed. In addition, the number of encoder layers is reduced in half, leaving 6 Transformer encoding layers as opposed to the 12 of the BERT base model. The student model is also initialized by using the teacher model's weights in a one layer out of two fashion, taking advantage of the common dimensionality between the two as seen in Fig. 3

RoBERTa (Robustly Optimized BERT Pretraining Approach) is a BERT-based model proposed in the study conducted in [8]. In their paper, they claim to have found that the original BERT model was significantly undertrained, leaving a reasonable margin for improvement. They developed RoBERTa to capitalize on this, with the intention of matching or exceeding the performance of the BERT methods. To achieve such a goal, the RoBERTa model follows the same architecture as BERT and introduced 4 different improvements over the original model: a) Training the model longer, using bigger batches, and over more data; b) Changing the masking pattern applied to the training data; c) Removing the next sentence prediction objective; and d) Training on larger batches.

4 EXPERIMENTS AND DISCUSSION

This section presents and discusses the results obtained from training and testing the three individual models. Information about the accuracy, F1-score, testing loss, test samples processed per second, and the time taken for each model to train forms the basis of performance comparison of the models. The parameters employed in this research were chosen after empirical analysis and include a batch size of 64, number of training epochs of 8, learning rate of $2e^{-5}$, weight decay of 0.01. The evaluation metric for the best model was set to F1 score as it combines both precision and recall and thus takes into account how the data is distributed through the dataset. The experiments then were conducted, and results are analyzed in the subsections below to answer our research questions.

4.1 Are Transformer-based Models Capable of Emotion Recognition?

To answer the first research question, we tested the Transformer-based models - BERT, DistilBERT, and RoBERTa on the Emotion dataset. As shown in Table 1, the results of testing the models

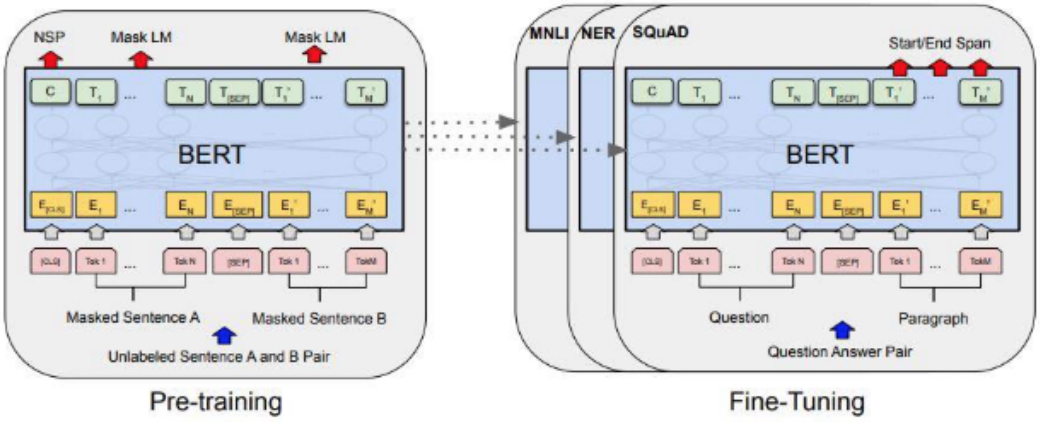

Figure 2: Overview of Pre-training and Fine-tuning phases of BERT [4].

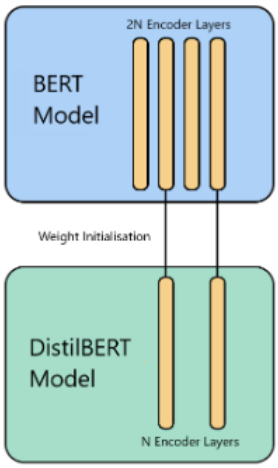

Figure 3: DistilBERT layer weight initializations from BERT.

demonstrate that the three models were more than capable of recognizing the emotions from the dataset as they all obtained accuracy scores and F1 scores over 92%.

4.2 Extensive Comparison of the Transformer-based Models

In answering the second research question, we compare the three models using their accuracy, F1-score, testing loss, test samples processed per second, and the time taken to train. It can be seen from Table 1 that DistilBERT outperformed BERT and RoBERTa in terms of accuracy, F1-score and samples tested per second. Despite the margin of improvement of DistilBERT over the other two models being very small in every metric, it is still considered relevant as the three models have very similar architectures.

The RoBERTa model also offered slightly better accuracy results over BERT Base, but their overall F1-score was the same. It is also worth noting that the F1-score is the most relevant performance metric for this experiment and provides the best overview of model performance. This is because F1-score combines both precision and recall and thus takes into account how the data is distributed through the dataset. This is a key observation since the dataset used for the model's training presents a very uneven distribution of labels.

In addition, the most significant difference in results across the three models is the time taken for training as seen in Table 2. For this metric, DistilBERT clearly yielded the best time as it only took 16:03 minutes to train against the 31:44 and 30:43 minutes of BERT Base and RoBERTa (respectively), nearly halving their time. The loss score represents the sum of the errors made during validation and testing. In this case, the loss score obtained was the same across all three models, rendering it irrelevant for the purposes of this comparison. Owing to its lighter and faster architecture, DistilBERT

Table 1: Comparison of Accuracy, F1 Scores, Loss and Samples/s Obtained during Testing.

Model	Accuracy	F1 Score	Loss	Samples/s
BERT Base	0.925	0.925	0.2	260.6
DistilBERT	0.928	0.93	0.2	427.7
RoBERTa	0.926	0.925	0.2	297.8

Table 2: Training (Fine-Tuning) Times Comparison.

Model	Training Time (mm:ss)
BERT Base	31:44
DistilBERT	16:03
RoBERTa	30:43

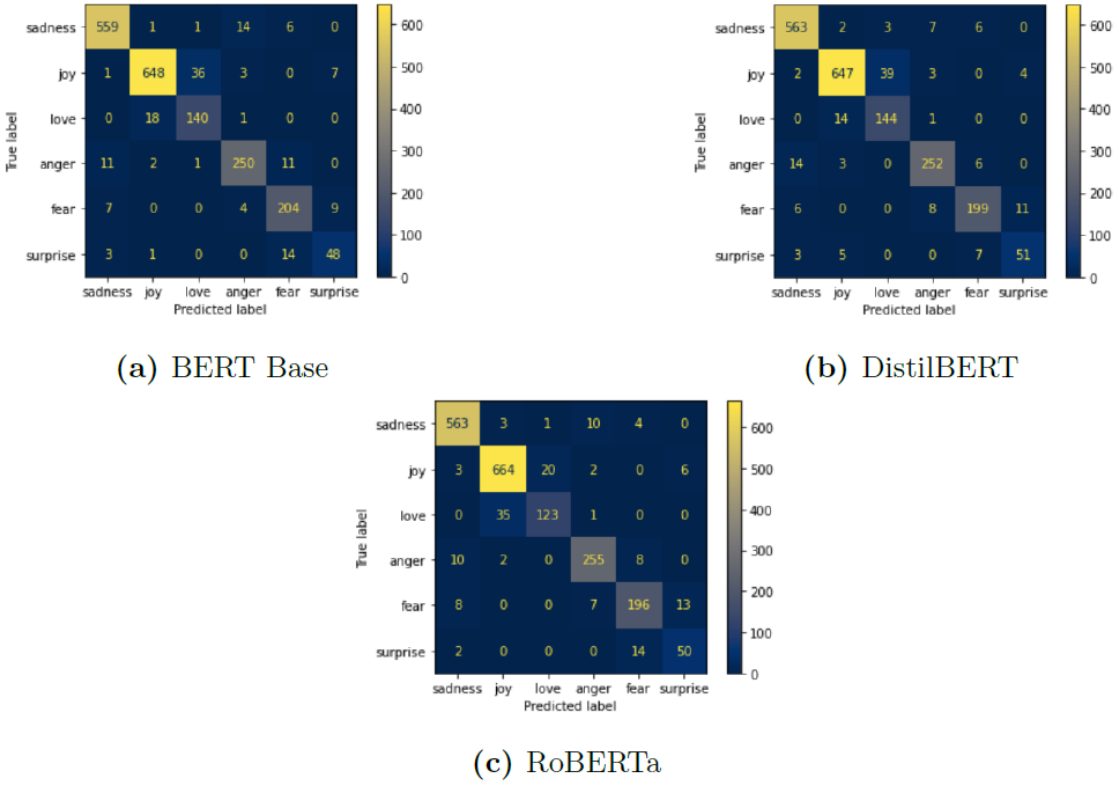

Figure 4: Confusion matrices for BERT Base, DistilBERT, and RoBERTa.

also managed to test significantly more samples per second than the other two models as seen in Table 1 which translated into taking less time during the testing task.

Furthermore, Fig. 4, shows the confusion matrices obtained after testing the BERT Base, DistilBERT, and RoBERTa models on the Emotions dataset respectively. All three matrices show extremely similar results, which further confirms the observations about the differences in the performance of the three models being very marginal. The higher number of "joy" and "sadness" predicted labels also highlights the uneven nature of the label distribution. The biggest take away from observing the confusion matrices is that while all three models were highly effective in recognizing all 6 emotions, they seem to have trouble identifying "joy" and "love" labels. This could be due to these emotions being semantically more like each other, given that Transformer-based models do present an understanding of language context and semantics.

Additionally, when analyzing the F1-scores obtained after each epoch of the training process across the three models as seen in Fig 5; BERT Base and DistilBERT reached the best performance in epoch 7, and RoBERTa in epoch 6. Therefore, the results obtained show that there are little to no diminishing returns present when training at 8 epochs for this scenario. This could potentially mean that the models do obtain better results when training above 5 epochs, as opposed to what is often suggested for fine-tuning tasks.

When comparing the results obtained in this experiment against the aforementioned literature in Section 2, it is found that the values obtained in [13] when comparing the accuracy scores of BERT against DistilBERT do share some similarities with those obtained in this paper. Their values only differed by 0.6%, which is a relatively small difference. On the other hand, the authors in [8] claim to have obtained a higher 3.7% difference in F1-scores when comparing BERT and RoBERTa on text classification tasks. This differs from the much smaller 0.5% determined from the testing done in this

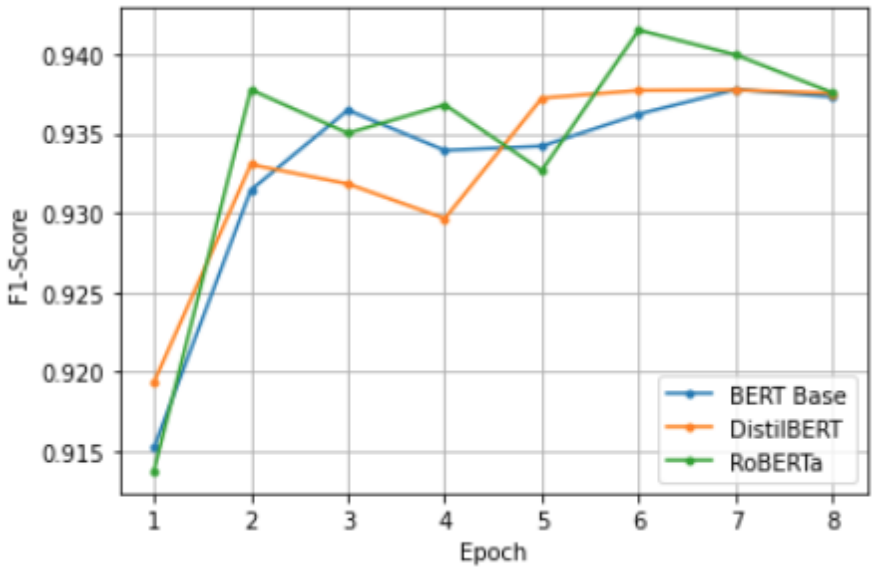

Figure 5: DistilBERT layer weight initializations from BERT.

paper. However, it must be kept in mind that for a comparison to stand its ground, it needs to have been made using the same or similar variables, such as dataset, epoch number, learning weights, machine, among others.

5 CONCLUSIONS

In this paper, the performance of BERT, DistilBERT, and RoBERTa are assessed when fine-tuned for emotion recognition using the Emotion dataset. The results showed that all the models proved to be more than capable of successfully carrying out the classification task, achieving F1-scores over 92%. DistilBERT is recommended over BERT and RoBERTa, as it achieved slightly superior results while halving training times and reducing resource consumption. When comparing the results with the literature, small differences were found that could be attributed to the dataset and machine differences.

In the future, further testing utilizing various training configurations will be done. The comparison of other Transformer-based models against different architectures could also potentially produce more varied and interesting results. In addition, the adoption of a bigger, more evenly distributed dataset could benefit the performance of the models.

REFERENCES

[1] Anjaria, M., Guddeti, R.M.R.: Influence factor based opinion mining of twitter data using supervised learning. In: 2014 Sixth International Conference on Communication Systems and Networks (COMSNETS). pp. 1–8 (2014)
[2] Ashraf, M.U., Rehman, M., Zahid, Q., Naqvi, M.H., Ilyas, I.: A survey on emotion detection from text in social media platforms. Lahore Garrison University Research Journal of Computer Science and Information Technology 5(2), 48–61 (2021)
[3] Deng, J., Ren, F.: A survey of textual emotion recognition and its challenges. IEEE Transactions on Affective Computing pp. 1–1 (2021)
[4] Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding (2018), https://arxiv.org/abs/1810.04805
[5] Eisner, B., Rockt äschel, T., Augenstein, I., Bo ˇsnjak, M., Riedel, S.: emoji2vec: Learning emoji representations from their description. In: Proceedings of the Fourth International Workshop on Natural Language Processing for Social Media. pp. 48–54. Association for Computational Linguistics, Austin, TX, USA (Nov 2016)
[6] Ekman, P.: Are there basic emotions? In: Psychological Review. vol. 99, pp. 550–553. American Psychological Association, US (1992)
[7] Felbo, B., Mislove, A., Søgaard, A., Rahwan, I., Lehmann, S.: Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. pp. 1615–1625. Association for Computational Linguistics, Copenhagen, Denmark (Sep 2017)
[8] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized bert pretraining approach (2019), https://arxiv.org/abs/1907.11692
[9] McCann, B., Bradbury, J., Xiong, C., Socher, R.: Learned in translation: Contextualized word vectors. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. p. 6297–6308. NIPS'17, Curran Associates Inc., Red Hook, NY, USA (2017)
[10] Plutchik, R.: The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. American Scientist 89(4), 344–350 (2001)
[11] Polignano, M., Basile, P., De Gemmis, M., Semeraro, G., Basile, V.: Alberto: Italian bert language understanding model for nlp challenging tasks based on tweets. In: 6th Italian Conference on Computational Linguistics, CLiC-it 2019. vol. 2481, pp. 1–6. CEUR (2019)
[12] Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., *et al.*: Improving language understanding by generative pre-training (2018)
[13] Sanh, V., Debut, L., Chaumond, J., Wolf, T.: Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108 (2019)
[14] Suero Montero, C., Suhonen, J.: Emotion analysis meets learning analytics: Online learner profiling beyond numerical data. In: Proceedings of the 14th Koli Calling International Conference on Computing Education Research. p. 165–169. Koli Calling '14, Association for Computing Machinery, New York, NY, USA (2014)
[15] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L.u., Polosu hin, I.: Attention is all you need. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 30. Curran Associates, Inc. (2017)
[16] Xu, P., Madotto, A., Wu, C.S., Park, J.H., Fung, P.: Emo2Vec: Learning generalized emotion representation by multi-task training. In: Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. pp. 292–298. Association for Computational Linguistics, Brussels, Belgium (Oct 2018)