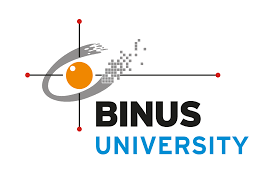
****

**TEXT EMOTION DETECTOR REPORT**

**COMP6576 – Natural Language Processing**

**LC01**

**2301859335 - Adrian Rianto**

**2301920941 - Kelly Cornelya Kesuma**

**2301853136 -  Reinaldy Sukamto**

**2301960440 - William Gazali**

Bina Nusantara University

Jakarta, Indonesia

1. **Background**

Texting has gone through a lot of changes through the years and it is heavily popular between young adults and adults alike. It is the most popular form of communication [2] that we use nowadays. Research has shown that even some people do not rely on just one medium of texting but instead use a whole different application to chat with their friends [1]. There are a lot of applications out there, some of which are iMessage, Line, WhatsApp, Telegram, etc.

In search for more expressive messaging, emoticons and emojis are used as different formats of communication have evolved. Emoticons and emojis are used to express emotions, avoid misunderstanding and for social purposes. Some people even perceive people who use emojis and emoticons to be more enjoyable and outgoing! [3]

Although some people like to use emojis, sometimes people can misunderstand what the sender is trying to convey and thus have a bit of a misunderstanding. So we aim to try to fix that by analysing what the sender sends and conveying certain emotions that are favourable for that text.

1. **Objective and benefits**

This study aims to be able to classify certain text messages into emotions based on the context of what the sender is trying to say and hopefully get a really good representation of the emotion that the sender is experiencing based on their text.

We hope that this study can evolve into a usable feature in a text messaging app in the future because this feature can help in determining what the sender is trying to convey, and it will be a usable feature inside a text messaging app.

1. **Data**

We got our dataset from kaggle <https://www.kaggle.com/praveengovi/emotions-dataset-for-nlp>

which is a bunch of text along with the emotion behind the context.

1. **Research Methodology**

We use a recurrent neural network model and multinomial naive bayes in the hopes that we can compare how the two models perform on the given dataset.

Recurrent neural network is a neural network architecture that takes into account the previous iterations, unlike conventional neural networks, this neural network passes the output along into the next iteration with gate usage. There are three gates in a recurrent neural network, the forget gate, the remember gate and the input gate. This type of neural network has long term dependencies having the ability to remember previous inputs.

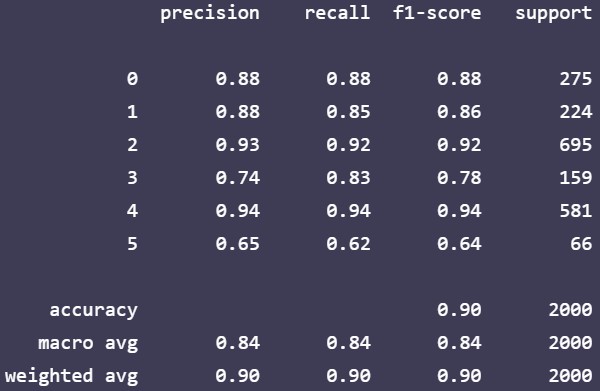
The input gate is where the input comes in each cell which is later processed, The remember gate is where the previous out passed through and the forget gate is where each cell loses some information, the output from the gates is passed through the same cell for n number of EPOCHS. Since we are working with a text based dataset, it makes sense to use a model that takes in the previous input as consideration.

Multinomial Naive Bayes take in the input and calculate the probability of that input having a given output, and tune the weights from the output. It is one of the popular models to work on an NLP problem especially sentiment analysis [4] so it is one of the models that we have to try.

We preprocess the input lemmatizing the words. The way it works is that the text was brought back into its original form for example given -> give, but it takes into account the context of the given text, this is the same as stemming except stemming did not take into consideration the context of the given sentence. Then we tokenize the words before it is fed to the model.

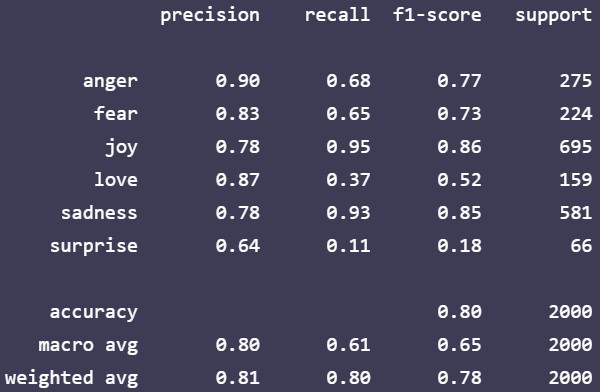
To make things easier and readable we use the tensorflow library and the sklearn library for our model.

1. **Hasil Eksperimen dan Analisis**
2. Recurrent Neural Network (LSTM)



We got a fairly general model with 90% accuracy on the testing dataset, although we do lack on classifying the “surprise” emotion because we lack the data on that label but for the rest, our model can classify just fine.

1. Multinomial Naive Bayes (MNB)

****

With the multinomial naive bayes we also got a fairly general model with 80% accuracy, and the same problem with the “surprised” label.

In conclusion, based on our finding and the dataset, we conclude that LSTM is better suited for this particular dataset with a good accuracy of 90% over 80% but that does not mean that MNB is bad, because it also performs well with 80% accuracy. Both models are still applicable in practice. We also can say that emotion detection is also applicable to fit as a feature in chatting apps and we hope that this study will be beneficial.

1. **Conclusion**

To conclude, first we define the use of messaging apps and emoticons, and we find that messaging is the most prefered way of communication amongst young adults [1] moreover as messaging transforms and extends into a better format, emoticons and emojis are used. [2]

We get a sample dataset from kaggle that includes a lot of sentences with certain emotions that labels them. We used 2 models which are the recurrent neural network and multinomial naive bayes. We got a fairly general model with 90% with the RNN model and 80% on the MNB.

Although the accuracy is high we still can get an unsatisfying result but we can fix that by adding more datasets so that the model can generalise better. Since this study is on a smaller scale there are ways to improve but we can say that this feature can be beneficial for chatting apps, to avoid misunderstanding between recipients and the sender.

1. **Referensi**
2. Perceptions of Mobile Instant Messaging Apps Are Comparable to Texting for Young Adults in the United States Shannon K.T. Bailey, Bradford L. Schroeder, Daphne E. Whitmer, & Valerie K. Sims University of Central Florida
3. A Large Scale Study of Text Messaging Use Agathe Battestini, VidyaSetlur, Timothy Sohn Nokia Research Center955 Page Mill RoadPalo Alto, CA 94304 USA{agathe.battestini,vidya.setlur,tim.sohn}@nokia.com
4. Emoticon, Emoji, and Sticker Use in Computer-Mediated Communication: A Review of Theories and Research Findings. Ying Tang, Indiana University Bloomington, Khe Foon Hew The University of Hong Kong
5. https://towardsdatascience.com/sentiment-analysis-of-tweets-using-multinomial-naive-bayes-1009ed24276b