APP FOR ANALYZING SENTIMENT IN GERMAN TEXTS

We apply sentiment analysis for a self-help application supporting users with psychological issues such as stress, anxiety, or depression.

We aim to build a mobile application (called ROLI) that is suitable for both Android and IOS systems, and that acts as a diary between the patient and the psychologist. The vision of the application is to be used as a self-help therapy tool by tracking the patient's sentimental state for a period of time, those sentimental states could then be discussed with the psychologist, thus, being able to track the effectiveness of the treatment, and help the patient regulate his mood, and having valuable data in fewer sessions, and spending less time in the psychologic clinic.

In particular, we are going to analyze the user sentiments through the use of ROLI applications, and with help of machine learning techniques. For this purpose, we implemented different ML-Models to be evaluated against each other to find the optimal model. We summarized in the process different machine learning approaches, tools, and the current state-of-the-art in the domain of sentiment analysis and discussed their advantages and limitations. We have tried different machine learning classifiers and a pre-trained model namely Multinomial NB, Vector Space, Maximum Entropy, Random Forest, and BERT the pre-trained model. To find the optimal model, we conducted a couple of experiments each with two datasets (corpora). The first dataset was data samples that were extracted from Twitter, and news articles and the other was from an external source (SB1OK). The first corpora were manually annotated with three classes namely positive, negative, and neutral. For the annotation process, three annotators were chosen to annotate the corpus separately. Each of the datasets was pre-processed before training/testing with a model, the pre-processing included stopwords, punctuations, symbol removal, and tokenization. The SB1OK required further processing such as removing usernames, hashtags, and URLs.

After training the different models, Multinomial NB was the best model with an accuracy of 66.6% and an f1-score of 66%. Using our internal dataset, the BERT model archived an accuracy of ≈ 61%.

We created a RESTFUL API using a python library named fast API, this library can interact effectively with a machine learning model to get a particular prediction. Visualization will be displayed in the application, so the user does not have to show his notes to the psychologist. The user can still modify the sentiment of his notes, in case the machine has made
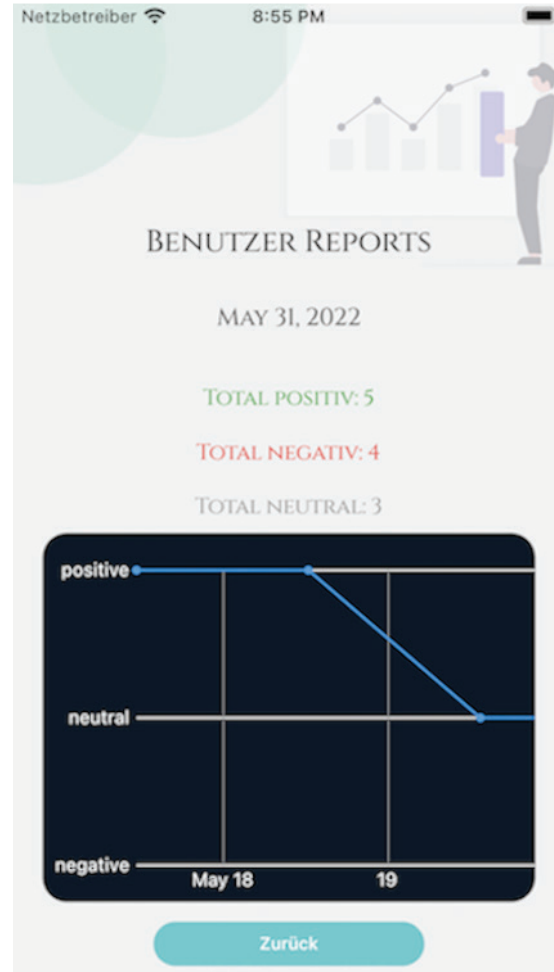
Sentimental state diagram in a specific period of time

Rodi Abdullah
078 659 06 41

Ali Hinnawi
079 906 63 91