Summary of tasks Sentimental Analysis Bachelor-Thesis

Author: Giorgio Bakhiet Derias
Tutor: Prof. Dr. Mascha Kurpicz-Briki
Expert: Andreas Dürsteler

List of all the task

1 Project research

This task covers the initial part of the project:
- initial information search
- meeting
- time planning

Next, there will be a constant search for information, as the project is not only programming, but also research.

2 Research dataset

The dataset is one of the key parts of the project, so there will be a search for one or more datasets, appropriate for the job.
The datasets will need to meet these requirements:
- large size
- binary labelled

3 Set up virtual machine

This task covers all the work that needs to be done to set up the virtual machine:
- order a vm
- initialize the machine
- install the necessary software
- check that there are no errors

4 Jupyter Notebook / Repository Setup

For my machine learning project I have chosen a Jupyter notebook as my workspace, for this reason in this task I will be engaged in:
- managing the notebook
- import libraries
- manage the libraries
- correct any errors in the imports

Later will be created a git to have a versioning of my work.

5 BERT model research

State of the art regarding sentiment analysis is the BERT model, in order to use this model, I will need:
- initial research on technology
- search of the adapted model
- various evaluations

6 Work on dataset

Once I have the information I need I can work on the dataset:
- Import dataset
- Extract dataset
- Transforming dataset
- Categorize reviews from dataset
- Present findings from data set in diagrams
- Resample dataset
- Preprocessing Dataset

7 Work on model

Once I am done with the dataset I can deal with the model:
- Create model
- Training model
- Test model
- Tuning model

8 Tests

This task is to do more testing on the goodness of the newly trained model.

9 Documentation

Documentation will be created at the same time so that at the end of the project we will have:
- a documentation of the work with deadline 17.6.2021, 18h00
- file in which will be marked the planned and worked hours with deadline 17.6.2021, 18h00
- a book with deadline 17.06.2021, 18h00
- a short film with deadline 17.06.2021, 18h00
- a poster with deadline 07.06.2021, 12h00