**Title of the project:** Twitter sentiment prediction

**Supervisor(s), faculty, email**

Tom Viering, EEMCS, [t.j.viering@tudelft.nl](mailto:t.j.viering@tudelft.nl)

**Which faculties and bachelor programs may enrol for this project?**

EEMCS, applied mathematics

**Background**

Every day, over 500 million Tweets are posted on the social media platform Twitter. These are short text messages of no more than 280 characters. Many use Twitter to share opinions and emotions.

**Project aim**

The objective is to create an application which estimates the mood or sentiment of a specified person or area based on Twitter data. The user provides a Twitter username or geographical area. The application gathers the most recent Tweets based on that input, and uses a machine learning model to estimate the general mood of those tweets and presents this to the user.

**Moscow analysis**

* **Must** take a Twitter handle or geographical area as input
* **Must** estimate the mood of Tweets (positive/negative/neutral) using a NLP model
* **Must** come up with a suitable evaluation metric for the task
* **Should** have multiple ML models trained on data
* **Should** perform a model comparison to find the best model for the task
* **Should** classify the mood on a scale
* **Should** have an interface (possibly command line interface)
* **Could** have a web interface
* **Could** interact with the Twitter API to collect tweets (time consuming!)
* **Could** use a dataset annotated by students themselves (time consuming!)
* **Could** use a state-of-the-art neural network fine-tuned on their own data
* **Won’t** interact with Twitter in any other way (send tweets, etc.)
* **Won’t** implement a machine learning algorithm from scratch

**Ethical aspects**

One issue regarding the Twitter sentiment classification example is that this application cannot be released in the EU. There are two problems: (1) according to the EU AI act users need to be informed when their mood is analysed through automated means, and (2) that Twitter users their data is used for a different purpose than originally intended. These may pose limitations for this application.

**AI approaches and methods**

Supervised learning, fine tuning existing NLP model, model selection.

**Prerequisites**

None

**Provided data and resources**

Data is collected using the Twitter API or an existing sentiment twitter dataset used ([example code](https://www.kaggle.com/kazanova/sentiment140))

**Recommended libraries, tools and code**

NLTK, Huggingface Transformers, Flask

**References**

Alshammari, Norah Fahad, and Amal Abdullah AlMansour. "State-of-the-art review on Twitter Sentiment Analysis." 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS). IEEE, 2019.