Mini Project 2

Dataset – The data used in this experiment is the Sentiment140 dataset, a publicly available data set created by three graduate students at Stanford University: Alec Go, Richa Bhayani, and Lei Huang. The data comprises approximately 1,600,000 automatically annotated tweets.

Tasks

Task 1 – Build **two different machine learning models** for sentiment analysis on the given dataset. Remember that one criterion of a machine learning expert is to search for suitable methods and try them to get the best output. You could use any model presented during the course. You have also the freedom to choose the approaches that are not presented during the course, but that you think could be utilized after your research. These include the traditional approaches, DNNs with fine-tuning. Some examples are CNNs, RNNs, Transformers architectures, Random Forests, Naïve Bayes, or SVM.

Task 2 – Compare and visualize the performance of the two models.

Task 3 – Write a scientific report which includes

- Introduction (what is the problem you are solving?)
- Data processing (what are the choices you made in data processing and how you performed it?)
- Modelling (What are the modeling approaches? How have you performed them? Why do you think one model performed better than the other one? What can be done better?)
- Conclusions (what were the “scientific” bottlenecks? How did you overcome them?)

You need to hand in your Python code (preferably Jupyter notebook or Google Colab notebook) alongside a written report.