Task 2:

For Task 3 applied a basic LSTM to the time period of 2014-2015 to the stocknet data attached in the Jupyter Notebook as well as a conda environment. The results are pretty good at almost 50%, but since only based on the historical prices the benchmark a couple of % points below StockNet from 2018 when also taking into account the processed tweets. The adversarial ALSTM previous Sota or DTML from 2021 also taking into account more macroeconomic market factors as well as via an autoencoder and the leading transformer architectures to take into account economic correlations between stocks and industries gives another few percentage points in both accuracy as well as Matthew coeffs as seen in the table below.

Table 1[1]: Classification accuracy (ACC) and the Matthews correlation coefficients (MCC) of our DTML and the baselines. DTML gives the state-of-the-art accuracy in all six datasets with up to 3.6 points higher ACC and 10.8 points higher MCC over the best competitors, which are significant amounts considering the difficulty of the problem.

Model	**ACL18 (US)**	
	ACC	MCC
LSTM [24]	0.4987 ± 0.0127	0.0337 ± 0.0398
ALSTM [31]	0.4919 ± 0.0142	0.0142 ± 0.0275
StockNet [31]	0.5285 ± 0.0020	0.0187 ± 0.0011
Adv-ALSTM [9]	0.5380 ± 0.0177	0.0830 ± 0.0353
DTML (proposed)	**0.5744 ± 0.0194**	**0.1910 ± 0.0315**

[1] https://datalab.snu.ac.kr/dtml/resources/YooSPK21.pdf