T-820-DEEP Introduction to Deep Learning
Assignment: Beat the market! (10%)

Objective

The objective of the assignment is to get hands-on experience in developing deep-learning neural-network models in *PyTorch*, going through the process of using external (partially processed) data, preprocess it, and test, train, experiment with, and employ ANN models.

Logistic

- You are expected to work on the assignment in a <u>group of two</u> (in exceptional cases, given instructor approval, you can work individually or in a group of three).
- Hand in:
    - Shared link to the CoLab notebook with the code you used for training your network (submit only the link).
    - A report in a PDF format called *A1Report.pdf*.
    - Two Python files: *model.py,* defining your *PyTorch* model (i.e., the ANN), and *predictor.py*, for employing your pre-trained model. Also, hand in your pre-trained model in a file called *pmodel.pt* (using "*torch**.**save(model**.**state_dict(), PATH)"*, see: https://pytorch.org/tutorials/beginner/saving_loading_models.html).

Description

The financial consulting company *UpUpUp Inc.* has been operating for over two years, providing various services to its clients, most notably recommendations for opportunistic high-frequency traders taking advantage of (minor) day-to-day variations in stock prices.

Generally speaking, it is believed that the stock market is efficient and transparent enough for such trading not to be profitable, even impossible. Also, *UpUpUp's* track-record in its recommendations so far has been far from stellar, especially with respect to predicting the stock-price movements of high-tech companies. It is thus starting to lose some of its clients.

The company is thus considering stopping offering this service, however, before doing so it first wants to see if its stock-movement prediction can be improved. All the hype around the potentials of deep-learning has caught their attention, so the company has contacted you --- as well as other leading experts in deep learning--- with a business proposal.

The proposal is for building an improved stock-movement prediction model for the stock-prices of three high-tech companies, the ones it has had the most difficulty with predicting. Over a period of 90 days, your model will be used to predict the day-to-day stock movements for the three companies. For each correct prediction you will receive $1,000, but nothing for wrong predictions. However, the real carrot is, that after this evaluation period *UpUpUp Inc.* will pay the expert coming up with the best performing model $1,000,000 for an exclusive right to its use, as well as financing further model development.

The requirements more prediction model must meet are listed below (because this is a proof-of-concept development there are some simplifying assumptions made, to be addressed later once the best performing model has been chosen):

- Your model is supposed to predict, for the three companies, whether their stock price will go up or not (note this is a binary decision, as you are not distinguishing between the price dropping or staying the same).

- You will get a historical day-to-day record of the stock-price development of the three companies, along with various information (features) that *UpUpUp Inc.* is currently using for its predictions. For details, see appendix. Note that you will have to preprocess the data to be suitable as an input to your model. Feel free to omit/use any of the provided information as you deem necessary.

- You should build a single model that predicts all three stock-movements.

- Your models will be evaluated by *UpUpUp Inc.* using a unified Python interface, provided in a file *predictor.py*. You need to implement the predict method in there to suit your model. *Make sure to adhere to the specification, otherwise your submission cannot be evaluated*.

- All models should be built in Google CoLab using *PyTorch*. Please place the names of the group members as a text field at the top of your notebook.

- Hand in a short report (3-4 pages) where you describe the main features of your model, state what approaches you tried (also what didn't work) and summarize your experimental findings, for example, compare the performance (both training and testing wise) of your model.

Appendix: Data

The *UpUpUp Inc.* has to clear various issues with its legal department before it can deliver you the data, a process that will get resolved in the very next days. However, so that you can start working right away (there is one million at stake!) you are here provided with a description of the data.

- In a file called *stock_prices.txt*, you get a historical record of the development of the stock-prices of the three companies. The number listed is the stock-price at the close of the market that trading day. All the values have to scaled to be relative to the stock value of each company at the opening of the first day listed (100.0).

NAME	TYPE	RANGE
company	integer	0 – 2
year	integer	2017 – 2020
day	integer	1 – 366
quarter	integer	0 – 3
stock_price	float	0.0 – …

- In a file, *market_analysis.txt*, you have a market analysis trend for each quarter (in the data). The data is provided just-in-time, that is, the analysis for a given quarter is available at the close of market the day before the start of the quarter. A trend value of -1 means negative prospects, 0 neutral prospects, and +1 positive prospects.

NAME	TYPE	RANGE
segment	categorical	{ BIO, IT }
year	integer	2017 – 2020
quarter	integer	0 – 3
trend	integer	{-1, 0, 1}

- In a file, *market_segments.txt*, you see in which market segment each of the three companies operates.

NAME	TYPE	RANGE
company	integer	0 – 2
segment	categorical	{ BIO, IT }

- In a file, *info.txt*, you get information about various features that *UpUpUp Inc.* is using in their current prediction model. It uses two independent expert predictions, (one indicating a prediction of the stock going up, 0 otherwise). The sentiment analysis score is the results of an extensive analysis of the discussion of the company and its products in the media over the past 24 hours, where a score of 0 means bad and 10 an excellent publicity. Finally, you are given four mysterious values that *UpUpUp Inc.* is producing themselves, however, they are not willing to explain in detail what they are. All they say is that some of the features are based on accumulated historical values of the stock's performance (you suspect that some of them might also be based on some insider information, but that's a different story).

NAME	TYPE	RANGE
company	integer	0 – 2
year	integer	2017 – 2020
day	integer	1 – 365
quarter	integer	0 – 3
expert1_prediction	binary	{ 0, 1 }
expert2_prediction	binary	{ 0, 1 }
sentiment_analysis	integer	0 – 10
m1	float	-10.0 – 10.0
m2	float	0.0 – 10000
m3	float	-1.0 – 1.0
m4	binary	{ 0, 1 }