Fundamentals of Data Science
Final project

Stock market predictions from Reddit WorldNews Channel titlelines

Andrea Messina
Leonardo Di Nino
Adil Atahan Atik
Borbála Sára Tóth Gy
Maria Alejandra Peña

The task

We want to deal with a binary classification problem about the relation between news titles and stock market opening.

The motivation

This is a good opportunity to discover the mechanics of the influence network that links macro socio-political dynamics with economics trends and sometimes unexpected derangement.

The dataset

- Reddit NewsChannel titlelines in 8 years
- Corresponding Dow Jones Industrial Average

https://www.kaggle.com/datasets/aaron7sun/stocknews

The method

We are going to train a DNN network in PyTorch framework.

We'll need a NLP processor to extract information.

From a *model design* perspective will focus on optimality of the learning rate, feature engineering and overfitting detecting.

From a *performance evaluation* perspective we are eventually benchmarking over state-of-art works.

References

Model design and optimization

- Cyclical Learning Rate to train neural networks, Leslie N. Smith, 2015
- Cross-validation for detecting and preventing overfitting, Andrew W. Moore, 2001
- Text mining at Feature level: A Review, T. Shruti, M. Choudhary, 2016

Previous work and state of art

- Text mining approaches for stock market prediction, A. Nikfarjam, E. Emadzadeh, 2010
- On the Importance of Text Analysis for Stock Price Prediction, H Lee, M Surdeanu, 2014
- Text mining of news articles for stock price predictions, Kim-George Aase, 2011