CF969-7-SU-CO: Big Data for Computational Finance

This assignment is for students enrolled under the CF969-7-SU-CO module code and consists of two parts.

Part I: Report (worth 66%)

You are asked to write a report on **one (1)** recent research paper in applications of machine learning for computational finance. Examples of possible papers to review appear in the following list:

- “Using deep Q-learning to understand the tax evasion behavior of risk-averse firms” by N. Goumagias, D. Hristu-Varsakelis, and Y. Assael.
  Available at: https://arxiv.org/abs/1801.09466
- “Improving financial trading decisions using deep Q-learning: Predicting the number of shares, action strategies, and transfer learning” by G. Jeong and H.Y. Kim.
  Available at: https://www.sciencedirect.com/science/article/pii/S0957417418306134
- "Ascertaining price formation in cryptocurrency markets with Deep Learning" by F. Fang, W. Chung, C. Ventre, Michail Basios, Leslie Kanthan, L. Lid, and F. Wu.
  Available at https://arxiv.org/abs/2003.00803
- “Deep Reinforcement Learning for Asset Allocation in US Equities” by M. Noquer i Alonso and S. Srivastava.
  Available at: https://arxiv.org/abs/2010.04404
- “An intelligent financial portfolio trading strategy using deep Q-learning” by H. Park, M. K. Sim, and D. G. Choi.
  Available at: https://arxiv.org/abs/1907.03665
- “Deep Modeling Complex Couplings Within Financial Markets”, W. Cao, L. Hu and L. Cao.
  Available at:
  https://pdfs.semanticscholar.org/6466/650d45a97941e9b583a84db1a734746016dc.pdf
- “Deep Learning for Mortgage Risk” by J. Sirignano, A. Sadhwani and K. Giesecke.
  Available at: https://ssrn.com/abstract=2799443
- “An application of deep reinforcement learning to algorithmic trading” by T. Theate and D. Ernst.
  Available at: https://arxiv.org/abs/2004.06627
- “Deep learning with long short-term memory networks for financial market predictions” by T. Fischer and C. Krauss.
  Available at: https://www.econstor.eu/bitstream/10419/157808/1/886576210.pdf
- "FinBERT: A Pre-trained Financial Language Representation Model for Financial Text Mining" by Z. Liu, D. Huang, K. Huang, Z. Li, and J. Zhao.
  Available at: https://www.ijcai.org/proceedings/2020/622
- “A deep learning framework for financial time series using stacked autoencoders and long-short term memory” by W. Bao, J. Yue , and Y. Rao.
  Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0180944
- “Deep Learning in Finance” by J. B. Heaton, N. G. Polson and J. H. Witte.
  Available at: https://arxiv.org/pdf/1602.06561v2.pdf

- “Machine Learning Algorithms for Financial Asset Price Forecasting” by P. Ndikum.
  Available at: https://arxiv.org/abs/2004.01504
- “Classification-based Financial Markets Prediction using Deep Neural Networks” by M. Dixon, D. Klabjan and J. H. Bang.
  Available at: https://arxiv.org/abs/1603.08604
- “CLVSA: A Convolutional LSTM Based Variational Sequence-to-Sequence Model with Attention for Predicting Trends of Financial Markets” by J. Wang, T. Sun, B. Liu, Y. Cao and H. Zhu.
  Available at: https://www.ijcai.org/Proceedings/2019/0514.pdf
- “Deep Learning for Event-Driven Stock Prediction” by X. Ding, Y. Zhang, T. Liu, J. Duan.
  Available at: https://www.ijcai.org/Proceedings/15/Papers/329.pdf
- "Deep Learning for Stock Market Prediction" by M. Nabipour, P. Nayyeri, H. Jabani, A. Mosavi, E. Salwana, and Shahab S.
  Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7517440/pdf/entropy-22-00840.pdf
- "Learning about risk: Machine learning for risk assessment" by Nicola Paltrinieri, Louise Comfort, and Genserik Reniers.
  Available at: https://www.sciencedirect.com/science/article/pii/S0925753518311184
- "Forecasting and trading cryptocurrencies with machine learning under changing market conditions" by Helder Sebastião and Pedro Godinho.
  Available at: https://jfin-swufe.springeropen.com/articles/10.1186/s40854-020-00217-x
- "Short-term bitcoin market prediction via machine learning" by Patrick Jaquart, David Dann, and Christof Weinhardt.
  Available at: https://www.sciencedirect.com/science/article/pii/S2405918821000027

You are encouraged to search the literature and choose your own academic paper (i.e., they have to contain serious scientific contributions along the lines of the aforementioned papers). If in doubt, you are welcome to contact me for feedback on the paper choice.

This part of your assignment assesses the following learning outcomes:

1. Understand the principles of (data-driven) algorithms such as modern machine learning and data mining algorithms;
2. Understand the application of (data-driven) algorithms on financial industry.

A review report of at most 1500 words must be written; you can include more information in an appendix if needed.

The report should both **summarize** and **evaluate** the article. You should aim to avoid summarizing just the knowledge from the lectures in your review; you can take this knowledge as a given. Your review should demonstrate that you are able to understand the contribution of the article and that you are able to critically reflect on it. As a guideline, a good report should address satisfactorily the following questions, namely:

- What is the paper about (i.e., what is the topic)?
- How do the authors approach the problem? I.e., what is the method they use?
- What are the results?
- What are strong points in the paper, in your view?
- What are weak points in the paper, in your view?

The report will be assessed attending factors such as its contents, presentation, organisation, clarity, explanations, references, grounding in the theory of machine/deep learning, etc.

Part II: Software for financial industry (worth 34%)

You are asked to come up with a case study that demonstrates your knowledge of software tools for the financial industry. Examples of areas include: Algorithmic Trading, Risk Assessment, Fraud Detection, Portfolio Management, Asset Pricing and Derivatives Market, Cryptocurrency/Blockchain, Financial Sentiment Analysis and Behavioral Finance, etc. You can use as a possible example our study of the correlation between closing values of European markets and American ones (the focus of labs #6 and #7). You can look at other sources of financial data (e.g., forex exchange rates) and provide a piece of software that does either some optimization or learning using the data.

Any language/software we covered during the modules is acceptable; you are also welcome to use a different one that you might have learnt independently or in some other module.

This part of your assignment assesses the following learning outcome:

3. Use software tools to build up data-driven algorithms and analyse the huge amount of historical data.

You have to produce a Jupyter-like notebook (if you do not use Python you do not need to include "In/Out code" in the notebook) discussing your case study covering the aims of your software and your solution. You should also comment on whether (and possibly why) your software has been successful or not in fulfilling the project aims.

Your submission will be assessed attending factors such as contents, clarity, explanations, etc. of the Jupyter-like notebook and correctness, techniques and style of the piece of software. Any part of the code that has been inspired by another source, should be explicitly identified as such in the comments.

- In a Jupyter Notebook it is possible to interchange blocks of codes with blocks of text, which can **and should** be used to explain what you are doing. For examples: Take a look at all the Jupyter Notebook files that contain the lab exercises for this module.
- Various essential elements should be present in your Jupyter Notebook: It should contain and discuss your code as stated above, but should also discuss (i.) the data (and where you got it from), (ii.) what experiments you are doing with the data, (iii.) the results of the experiments (iv.) a discussion of to what extent the software/experiments were successful, and why.
- Complex pieces of code should also include comments **within** the blocks of code (for Python, you use the '#'-character for this) which may discuss the technical details of the code, potentially even line-by-line.
- If you are using a language other than Python, then you can either create a document that replicates the style of a Jupyter-like notebook, or you can opt to include a file containing just the source code together with an **auxiliary document** that discusses your software, the data, explains how to run the software, and describes and discusses the results of your experiments.
- In the latter case, the code should still be explained, which can either be done by inserting extensive comments throughout the code, which explain what you are doing, **or** including such a break-down of the code in your auxiliary document.

Deliverables

The submission of the assignment must be done through FASER. The following parts should be uploaded independently (and not bundled together in a zip file):

- the review report (in PDF format) for Part I;
- the sources (including the data files) and executables (if any, compiled, either on a Mac or PC) of your software in Part II, to allow me to verify your results if needed;
- a Jupyter-like notebook discussing your piece of software in Part II.

Please refer to the Student's handbook on the School's Policy on Plagiarism and Late Submission.