W266 - Project Proposal - Summer 2023, Thursday @ 6:30pm - Mark Butler

**Team Members:**

[Vincent Goldberg](mailto:vincent.goldberg@berkeley.edu) and Samantha Williams

**Topic Proposed:**

Classifying Finance-Related Tweets

**Objective:**

The objective of this project is to develop an effective classification model for finance-related tweets by accurately categorizing these tweets based on their topic.

**Importance and Challenges:**

Finance-related tweets contain valuable information that can influence financial markets and investor behavior. However, extracting relevant insights has its challenges including: the noisy and abbreviated nature of tweets, sarcasm and irony, difficulty of accurately classifying tweets into specific finance-related topics.

**Dataset:**

We will utilize the [Twitter Financial News](https://www.kaggle.com/datasets/sulphatet/twitter-financial-news?select=valid_data.csv) dataset from kaggle, consisting of 21,107 annotated finance-related tweets. This dataset has been labeled with finance-related topics, such as earnings, currencies, macro, and company news. We believe it will be a suitable foundation for training and evaluating our classification model.

**Algorithms and Implementation:**

We intend to exploit the FinBERT pretrain language model (PLM) for classifying our financial tweets. We believe the domain-specific knowledge of FinBERT will outperform the BERT-base model. A series of experiments will be conducted and compared that both combine and individually implement the fine-tuning and inter-training strategies presented in Sun et al. (2019) and Shnarch et al. (2022), respectively. These experiments will be implemented on both the FinBERT and BERT-base models with the results being compared. Finally, to ensure our advanced models are producing advanced results, will be utilized the methodologies presented in Lin et al. (2023) to compare the PLM to linear classifiers which have been shown to produce competitive performances for text classification.

**References:**

1. Yang, Yi & UY, Mark & Huang, Allen. (2020). FinBERT: A Pretrained Language Model for Financial Communications. https://www.researchgate.net/publication/342198406\_FinBERT\_A\_Pretrained\_Language\_Model\_for\_Financial\_Communications
2. Lin, Y.-C., Chen, S.-A., Liu, J.-J., & Lin, C.-J. (2023, June 12). *Linear Classifier: An Often-Forgotten Baseline for Text Classification*. arXiv.org. https://arxiv.org/abs/2306.07111
3. Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019, May 14). *How to Fine-Tune BERT for Text Classification?* arXiv.org. https://arxiv.org/abs/1905.05583
4. Shnarch, E., Gera, A., Halfon, A., Dankin, L., Choshen, L., Aharonov, R., & Slonim, N. (2022, March 20). *Cluster & Tune: Boost Cold Start Performance in Text Classification*. arXiv.org. https://arxiv.org/abs/2203.10581