Project Lead : **Sanjit**
Sentiment Trading

LSESU Data Science Society

Outline and objectives	The thoughts of market participants often move prices and there's a wealth of information contained in the millions of tweets that are sent everyday. We will use Natural Language Processing techniques to generate trading signals from tweets, then backtest a strategy which uses these signals. The final deliverables are a Jupyter notebook and a report which analyses the performance of our trading strategy. The ultimate result is a consistently profitable strategy.
Approach	We're analysing sentiment, which involves labelling tweets as e.g. bullish / neutral / bearish. To obtain a vector representation of each tweet, we'll use a pre-trained word embedding model. Scraped tweets are unlabelled, so we'll initially use a clustering algorithm to generate labels for them. We could also try transfer learning with a pre-trained classifier. Finally, we'll use the now-labelled tweets to generate trading signals.
Data sources	We'll scrape twitter for tweets using Mottl/GetOldTweets3: A Python 3 library and a corresponding command line utility for accessing old tweets. We could use this Stock-Market Sentiment Dataset or this Financial PhraseBank for transfer learning.
Difficulty	Intermediate / Advanced
Machine Learning / Data Science techniques	• Text preprocessing • Text feature extraction (N-grams / Bag of words / BERT / FinBERT / word2vec / GloVe / ELMo) • Clustering algorithms (k-means, Gaussian, KNN) • Classification algorithms (Support vector machines, Logistic regression)
Timeline	Estimated effort: 5 hours a week • Text preprocessing (1 week) • Feature extraction (3 weeks) • Model training and evaluation (3 weeks) • Backtest and report (1 week)

Prerequisites	Applicants should have intermediate knowledge of Python as well as an ability to use documentation well. They should also have a solid understanding of the basics of machine learning. Most importantly, they should expect and be willing to research extensively.
Useful resources	1. Everything There Is to Know about Sentiment Analysis 2. Ultimate Guide To Deal With Text Data 3. What Are Word Embeddings for Text?