



THE UNIVERSITY OF
CHICAGO

DEPARTMENT OF STATISTICS

Master's Thesis Presentation

Hang Won Jang

Department of Statistics
The University of Chicago

“News sentiment-drive portfolios using large language models”

Tuesday, February 13, 2024, at 1:30 PM
Jones 111, 5747 S. Ellis Avenue

Abstract

I develop investment portfolios based on news sentiment analysis, combining large language models with logistic regression. Instead of binary sentiment analysis, the analysis is done on 28 detailed emotions defined in the GoEmotions dataset. I train 28 separate logistic regression models, each focusing on one of the 28 emotions, on top of the embeddings of GoEmotions dataset from state-of-the-art large language models. Then, I conduct transfer learning to adapt the trained model to news data for sentiment analysis. I construct long-short investment portfolios based on these emotion scores. Using the correlation between emotion scores and the three-day stock return as weights will significantly improve the Sharpe ratio of portfolios.