Stock Market Analysis Using Deep Learning and Natural Language Processing During the Covid-19 Pandemic

Yan-Ni Tai, Department of Statistics, National Cheng Kung University

Abstract

This project focuses on the **Taiwan 50 Index** as a comprehensive indicator of the stock market, selecting six industrial stocks significantly impacted during the Covid-19 pandemic. The number of confirmed Covid-19 cases is used as an indicator to evaluate the pandemic's impact, aiming to enhance model performance.

Moreover, over **60,000 financial news** articles from cnyes.com during the Covid-19 period were analyzed for **sentiment analysis**. The project employs the time series **ARIMA** model and the deep learning **LSTM** model to predict stock index trends. It also compares the accuracy of the prediction directions using a **rolling window** approach.

Methods

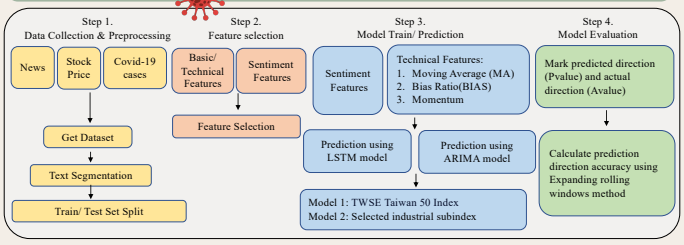

Figure 4. Step flow chart

- **Phase 1**: Implemented a **web crawler** to collect **60,000+** financial news articles from cnyes.com, integrating stock price data and Covid-19-related information while labeling stock price sentiment. Utilized **CKIPTagger** for word segmentation, incorporated a sentiment lexicon, and applied **feature selection** techniques to extract key text features.
- **Phase 2**: Performed sentiment analysis using **NLP lexicon methods** and extracted sentiment features by computing ratios and mean values of different sentiment scores.
- **Phase 3**: Developed **time series model** (ARIMA) and **deep learning model** (LSTM). Integrated sentiment features, COVID-19 data, and technical stock indicators for enhanced prediction accuracy.
- **Phase 4**: Labeled **predicted vs. actual trends** and applied an **expanding rolling window** to compute prediction accuracy, validating and evaluating model performance.

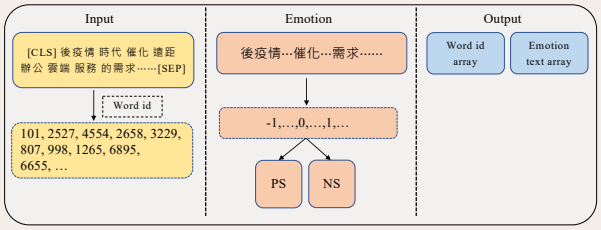

Figure 5. Fine-tuning Output Flowchart

Motivation

A comparison plot illustrates the relationship between the TWII and the number of confirmed Covid-19 cases from January 16, 2020, to December 31, 2021, and from January to July 2022. Key dates where the TWII experienced significant drops are highlighted, showing corresponding index values and confirmed case counts, suggesting a correlation between the two.

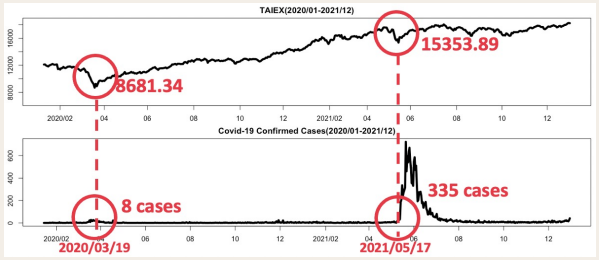

Figure 1. Comparison of TWII Performance and Confirmed COVID-19 Cases (2020-2021)

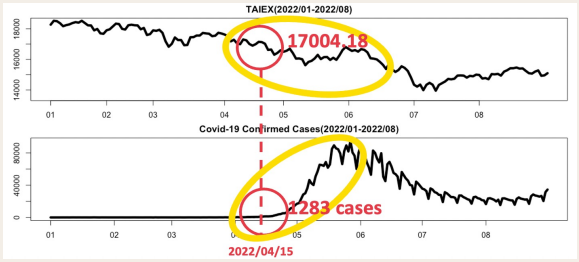

Figure 2. Comparison of TWII Performance and Confirmed COVID-19 Cases (2022)

This study examines the impact of the Covid-19 pandemic on various industries by standardizing stock indices and analyzing market trends. The findings indicate that sectors such as **chemicals, biotechnology, medical, food, shipping, tourism, and retail** experienced significant fluctuations.

This project aims to leverage government-reported Covid-19 case data as an indicator, apply sentiment analysis to financial news, and predict the Taiwan Weighted Index (TWII) along with selected stock indices.

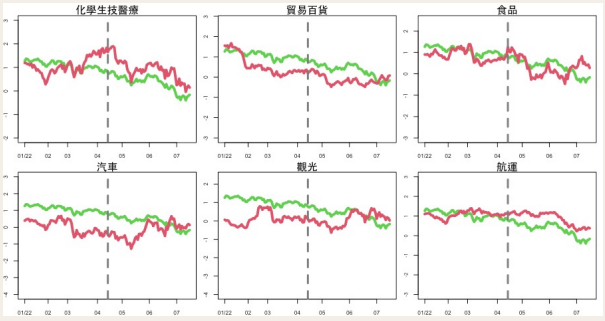

Figure 3. Selected stock index and TWII after standardization

Result

The goal of this project focuses on analyzing whether the sentiment in the news under Covid-19 affects the stock market and which of the two models can predict the stock market more accurately. The following are the results of this project:

1. The accuracy of LSTM model has increased by **30%** after adding sentiment features
2. The performance of **LSTM model** is better than ARIMA model.
3. Through prediction, we can understand the possible stock price trends of Taiwan's overall and various industries under the impact of the epidemic and can apply them when encountering major events in the future.

Reference

[1] TWSE–Indices information https://www.twse.com.tw/zh/page/products/indices/series.html
[2] Open government data platform – Covid-19 inspection number data https://data.gov.tw/dataset/120451
[3] WHO Director-General's opening remarks at the media briefing on COVID-19, 11 March 2020. Retrieved from https://reurl.cc/8WX4dd
[4] Anue News https://www.cnyes.com
[5] CKIP Tagger https://github.com/ckiplab/ckiptagger