NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

School of Computer Science and Engineering
College of Engineering

Natural Language Processing in Finance
Predicting Stock Market Trends Using a GRU-based BERT Sentiment Index
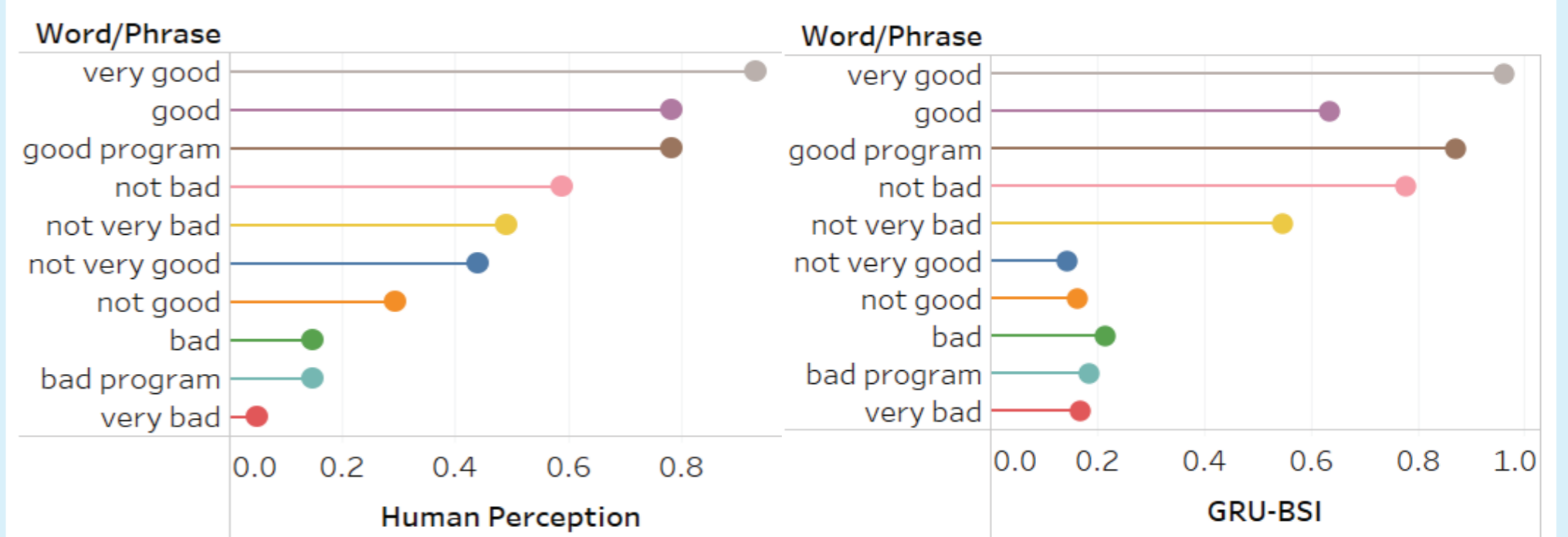
Objectives

Despite the widespread recognition that BERT model has outperformed most NLP models, BERT remains an underexplored terrain in the field of stock trend prediction.

Aim to construct a novel sentiment polarity index ("GRU-BSI") as a measure of market sentiments that were extracted from BERT word embeddings
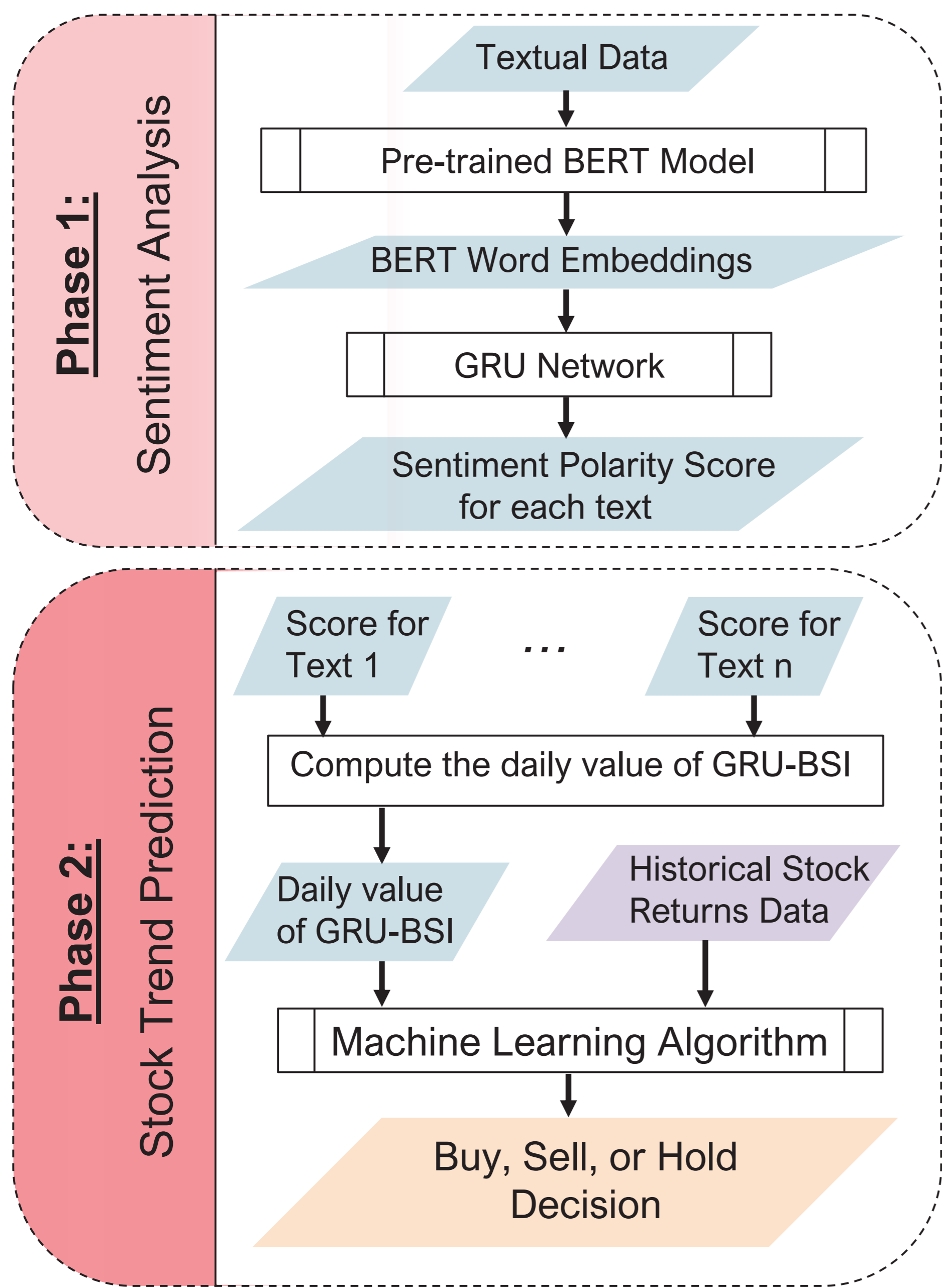
Utilise GRU-BSI and stock returns to predict stock trends

Methodology

Phase 1: Sentiment Analysis

Textual Data → Pre-trained BERT Model → BERT Word Embeddings → GRU Network → Sentiment Polarity Score for each text

Phase 2: Stock Trend Prediction

Score for Text 1 … Score for Text n → Compute the daily value of GRU-BSI → Daily value of GRU-BSI

Historical Stock Returns Data

Daily value of GRU-BSI + Historical Stock Returns Data → Machine Learning Algorithm → Buy, Sell, or Hold Decision

Comparison of Sentiment Polarity Scores

The sentiment polarity scores derived from BERT word embeddings (Right Fig.) *bore close resemblance* with the scores assigned manually by a human (Left Fig.), despite the presence of slight discrepancies.

Word/Phrase: very good, good, good program, not bad, not very bad, not very good, not good, bad, bad program, very bad — Human Perception (0.0–0.8) / GRU-BSI (0.0–1.0)

Stock Trend Prediction Results

**1** GRU-BSI *outperformed* VADER and TextBlob, which are well-established sentiment analysis tools, in *three out of six machine learning algorithms* utilised for stock trend prediction.

- The three algorithms are KNN, MLP, and LSTM.

Best-performing LSTM Model for AMAZON
Accuracy vs Window Size (1–8)

Best-performing MLP Model for AMAZON
Accuracy vs Window Size (1–8)
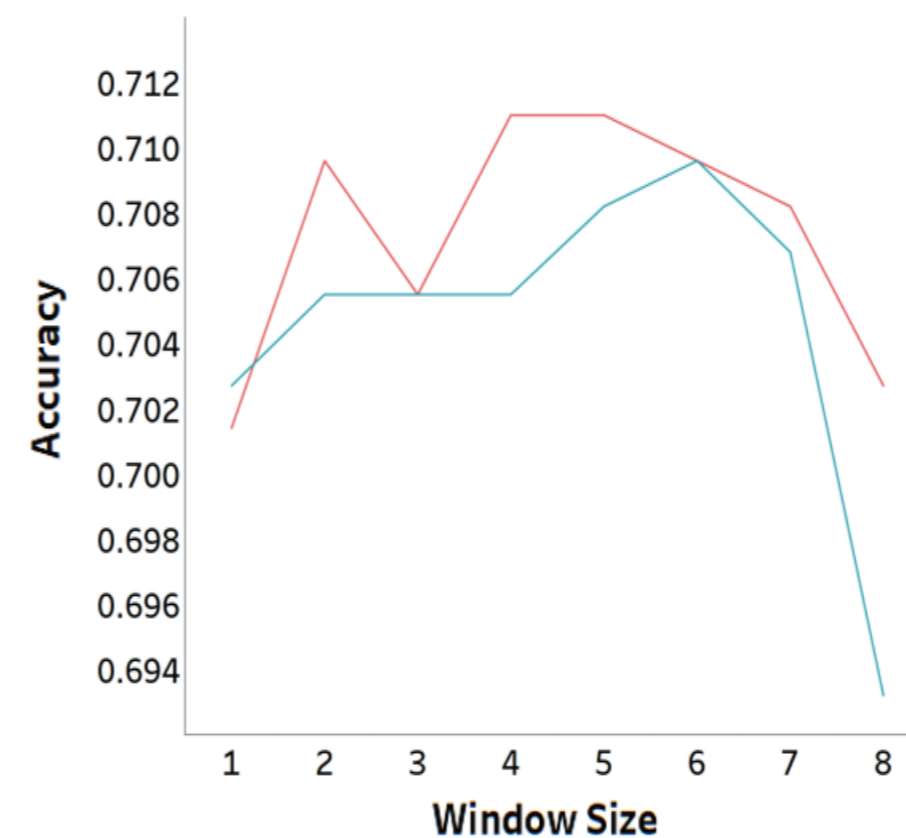
■ Baseline LSTM Model for Amazon (Input variables are historical stock returns. No sentiment data included)
■ Best-performing LSTM Model for Amazon (Input variables are historical stock returns and GRU-BSI constructed using social media comments)
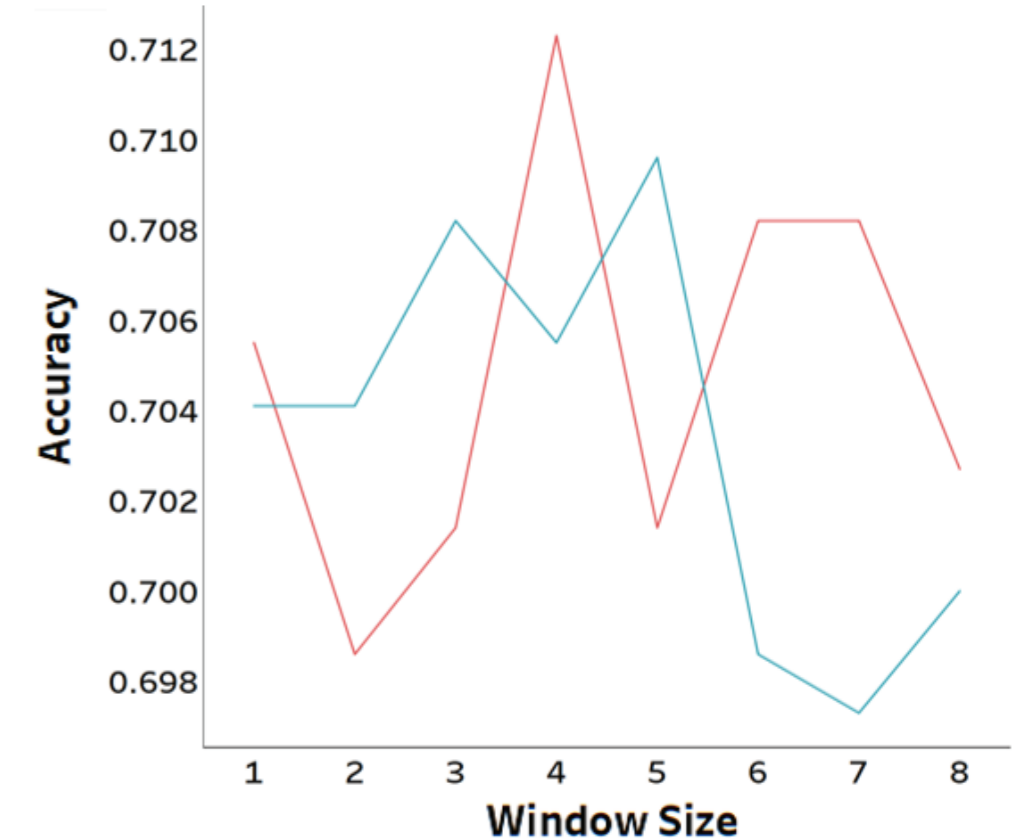
■ Baseline MLP Model for Amazon (Input variables are historical stock returns. No sentiment data included)
■ Best-performing MLP Model for Amazon (Input variables are historical stock returns and GRU-BSI constructed using social media comments)

2 *Social media comments* have stronger predictive power than news headlines.

3 Optimal Input Window Length for Stock Trend Prediction: *Between 3 and 6*

Student: Pun Shu Fen   **Supervisors: Dr. Li Fang & Dr. Wang Zhaoxia**