UTSA
CREATING
BOLD FUTURES

Stock Prediction

Using Elon Musk's tweets

CS 3793/5233 Artificial Intelligence

Project **Overview**

Stock Prediction

Does Elon Musk drive significant changes in the stock market with his tweets?

- Tesla shares plummeted when Musk tweeted that prices were “too high” in 2020
- Do less direct tweets sway market sentiment?

AI stock prediction model that leverages **sentiment analysis** for **feature extraction** in traditional neural network architecture.

- Sentiment analysis on Elon Musk tweets
- Focus on Tesla stock
- Time series prediction

The aim of this project is to build a strong foundation in the field of cutting-edge **time series forecasting** as well as discover how much impact on the market Elon Musk has.

Problem Statement

Tesla Stock Prediction and Sentiment Analysis

By amalgamating sentiment analysis with the existing AI model, we aim to augment the prediction performance and capture the dynamic correlation between Musk's tweets and Tesla's stock movement.

- Can we achieve a high level of accuracy using only common-stock-associated metrics (daily open, daily close, volume, etc.)?
- Can we enhance this model's accuracy by integrating sentiment analysis on Elon Musk’s tweets?

Methods

Explored Architectures

LSTM Time series prediction

- Trained on “Open” data related to TSLA stock price

Transformer-based: BERT

- Pre-train word relationships on unlabeled data
- Fine-tune with sentiment labels (stock response)
- Encoding and tokenizing text
- Sentiment classification

Time series prediction on stock

Chosen Approach

Two-part System

Feature extraction

- BERT for positive/negative sentiment analysis on Elon Musk's Tweets and the use of traditional stock features

LSTM Time Series Prediction

- Perform high-dimensional nonlinear function approximation, serving as a mechanism for predicting stock processes based on the extracted features.
- X_Train , X_val, X_test (Data & Open)
- Y_train, Y_val, Y_test (Close)

Experimental Setup

System performance was assessed by the following metrics:

- Root Mean Square Error (RMSE)

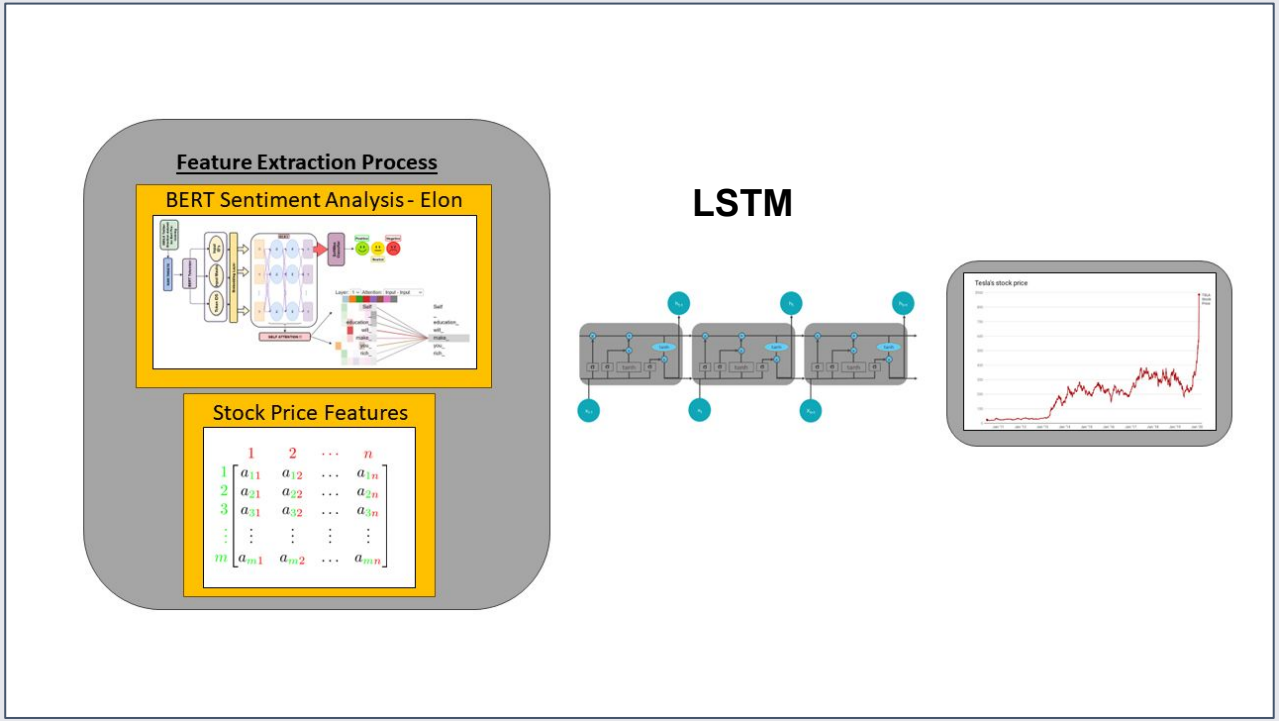
Two Experiments:

- Without Sentiment analysis
- With Sentiment analysis (Future Work)

Dataset was comprised of the following data

- Elon Musk Tweets (Kaggle)
- Tesla stock prices (import yfinance)

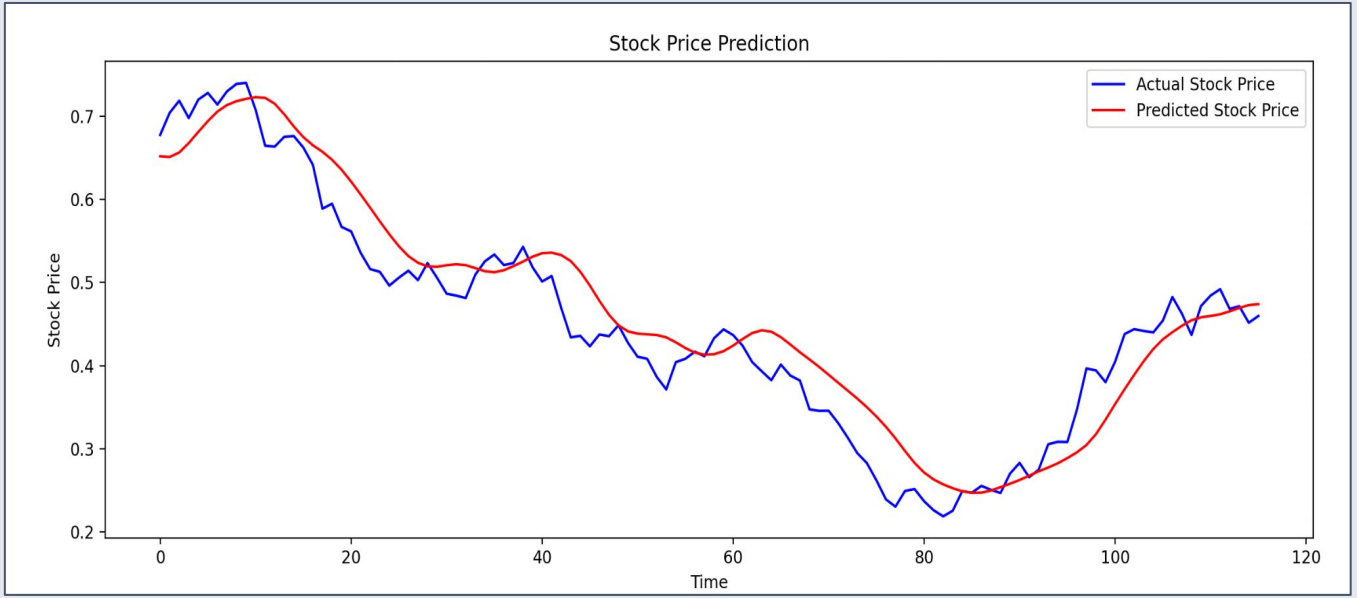
Proposed System Architecture



UTSA

Time Series Prediction using LSTM

- Experiment 1 - no sentiment analysis
- RMSE: 0.052
- 11 Months



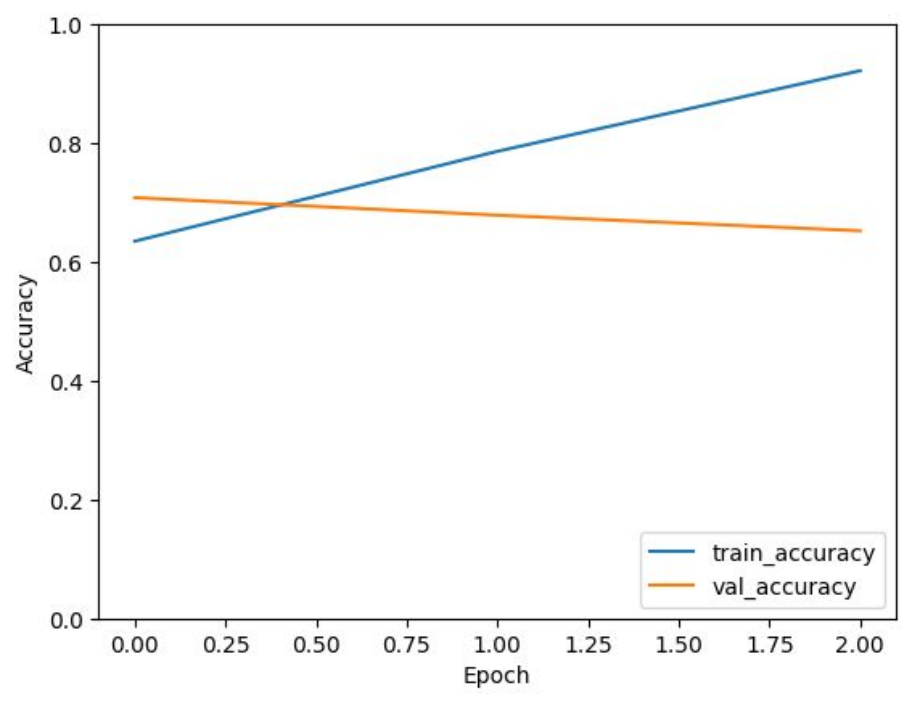
Sentiment Analysis using BERT

Dataset Labeling:

- Several tweets per day
- Tweets on days the market was closed
- Market metrics to measure impact
- Timespan to calculate market impact
- Where is the correlation if it exists?

BERT model:

- DistilBERT base model used
- TwHIN-BERT and BERTweet pre-trained for Twitter



Challenges

- **Schedule**
- Identifying an appropriate model based on the current scientific literature that would:
  - provide adequate challenge
  - strong results
  - aligned with current technical ability
- Selecting features for use in our proposed system
- Scheduling and handling time management

References

- [1] Selvin, Sreelekshmy, et al. "Stock price prediction using LSTM, RNN and CNN-sliding window model." 2017 international conference on advances in computing, communications and informatics (icacci). IEEE, 2017.
- [2] https://www.kaggle.com/datasets/andradaolteanu/all-elon-musks-tweets
- [3] Braig, Niklas, et al. "Machine Learning Techniques for Sentiment Analysis of COVID-19-Related Twitter Data." IEEE Access 11 (2023): 14778-14803.
- [4] Kiranyaz, Serkan, et al. "1D convolutional neural networks and applications: A survey." Mechanical systems and signal processing 151 (2021): 107398
- [5] Hutto (2014) vaderSentiment [Source code] https://github.com/cjhutto/vaderSentiment.
- [6] Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. (2018). Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing. [Online]. Available: https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html