Deep Learning Equity Trading Model

Shayne Biagi, Andrew Istfan, Franco Medrano

Problem Definition

- Definition
  - Accurately forecasting the stock market has always been a cornerstone of business research and development.
  - Popular trading algorithms rely too much on technical analysis of basic indicators (High, Low, Open, Close, Volume)
  - Research into inclusion of sentiment analysis only leverages official news outlets, which involve many biases [1].
  - State-of-the-art deep learning models have seen very little application in this field [2]. And the models initially explored have seen many incremental improvements since.

Motivation

- Accurate prediction of the stock market is greatly beneficial for a variety of persons:
  - Politicians and Government Officials
    - Can leverage predictions to predict economic health and work to counteract problematic trends
  - Investors
    - Can better profit from trades and gain stability in income from the stock market.
  - Businesses and Corporations
    - Could utilize forecasts to more accurately assess quarterly performance and projected success.

Existing Solutions and Limitations

- Traditional GAN with BERT [1]
  - Model is prone to convergence failures and training instability
  - Fundamental analysis only leverages biased news outlets and does not evaluate sentiment directly from the public.
- Traditional GAN Model [2]
  - Model is prone to convergence failures and training instability
  - No inclusion of fundamental analysis
- PCA-LSTM Model [3]
  - Uses a simple LSTM neural network for predictions based solely on technical indicators
  - No inclusion of fundamental analysis

Our Approach - Pipeline Diagram
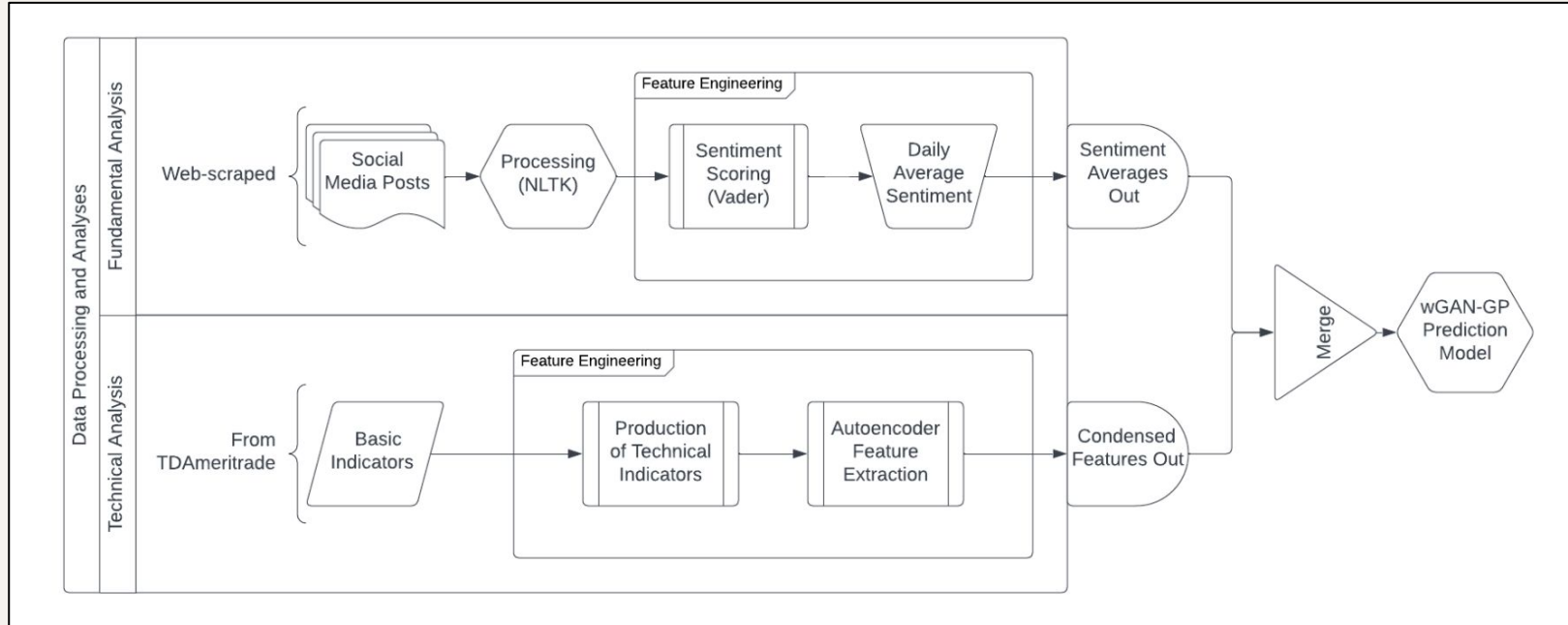
Our Approach - Novelty

- The usage of sentiment analysis of data coming directly from the public, rather than from biased news outlets or similar feeds.
- wGAN-GP is used over traditional GAN networks, for promotion of training stability and convergence. Improved forms of GAN networks have seen little implementation in the field of finance.

Our Approach - The Dataset

- Relying on the TDAmeritrade API, basic financial data for $AAPL and a number of comparative assets are compiled for the past 10 years including: Open, Close, High, Low, and Volume
- Comparative Assets
  - To accurately gauge and incorporate industry and macro trends, we have compiled the basic technical indicators of competitive companies, industry indexes, and
- Tweets mentioning "$AAPL" and "AAPL" are scraped from Twitter.
- The dataset includes data from 05/09/2013 to 03/31/2023

Our Approach - Fundamental Analysis

- Fundamental Analysis
  - Text content from Twitter are web-scraped and processed via the NLTK library.
  - VADER is used upon each media post to compute sentiment scores which are then averaged for any given day.

Our Approach - Technical Analysis

- Feature Creation
  - Technical indicators are computed from our initial set of basic indicators. Some of the technical indicators computed are:
    - Stochastic Oscillators
    - Relative Strength Index
    - Simple Moving Averages for Close and Volume
    - Moving Average Convergence/Divergence
- Feature Extraction (***Note: Development is still in progress***)
  - An autoencoder is used to produce a set of compressed features from the combination of basic and technical indicators.
  - The compressed representation of the full technical indicator set is fed into the prediction model for training and forecasting purposes.

Our Approach - Prediction Model

- After feature engineering, our final set of compressed technical indicators is appended with the sentiment scores from the fundamental analysis and fed into a wGAN-GP model for training.
- wGAN-GP Model Description
  - Wasserstein distance is used as a loss function for promotion of training stability.
  - Gradient norm penalty is enforced in the discriminator in hopes of achieving Lipschitz continuity.
  - Generator is an LSTM with input units equal to the final number of features and 512 hidden units, leading into a final linear layer with a single output for the closing price of each day.
  - Discriminator is a 13-Layer CNN composed of 1-D Convolutional, LeakyReLU, and Dense layers.
- Model is tuned with a predefined grid search.

Generator

Discriminator

Results

- The graph to the right details predictions made by the wGAN-GP model over the test set with a 3-day rolling window to predict the closing price of the 4th day.
- Evaluation Metrics from predictions:
  - RMSE: 4.67
  - MAPE: 0.0255
- The model is able to predict future stock movements with reasonable accuracy.
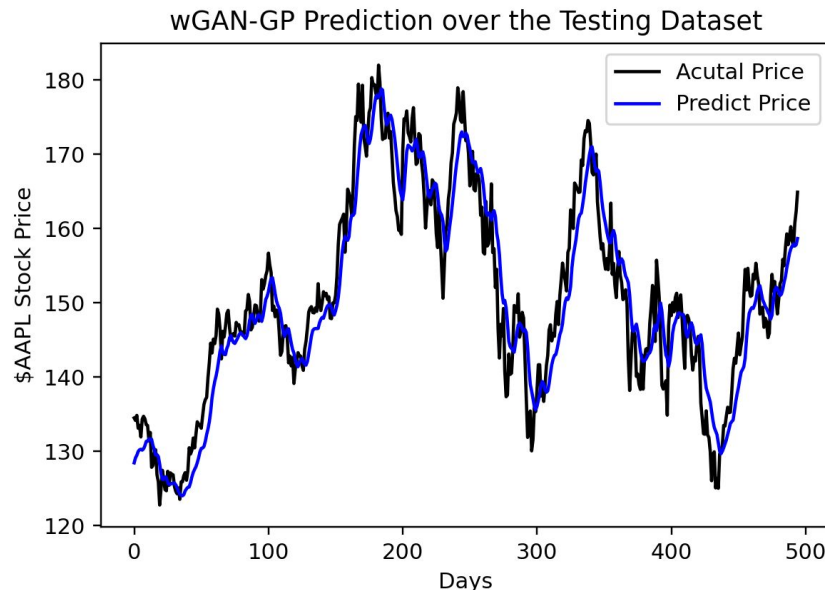
Note: This result was gained from training and testing the model on the full set of technical indicators, without compression from autoencoding. Autoencoding for feature extraction is currently under development in hopes of simplifying the feature space to improve generalization and thus forecasting accuracy.

Future Work

- Implement online learning capabilities
  - Reddit continuous scraping
- Expand data collection of sentiment analysis
  - Reddit
  - Facebook
- Apply model to additional stock securities
  - Meta
  - Gold
- Enhance dataset to pool financial data on intradaily intervals
- Implement model in variable macroeconomic conditions to measure its generalization ability

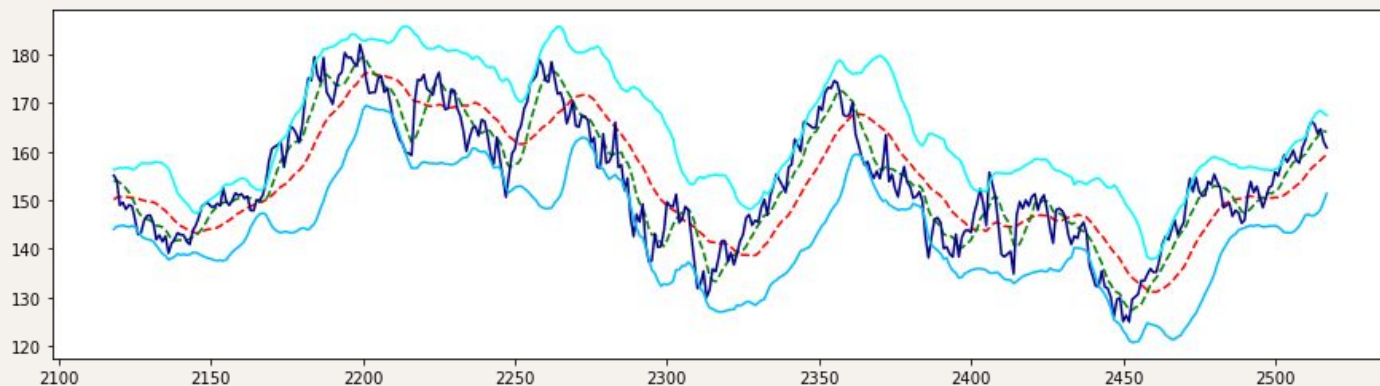# Team Contributions

Shayne Biagi

- Developed wGAN-GP Stock Prediction Model
  - Generator
  - Discriminator
  - Tuning

Discriminator Loss with Gradient Penalty

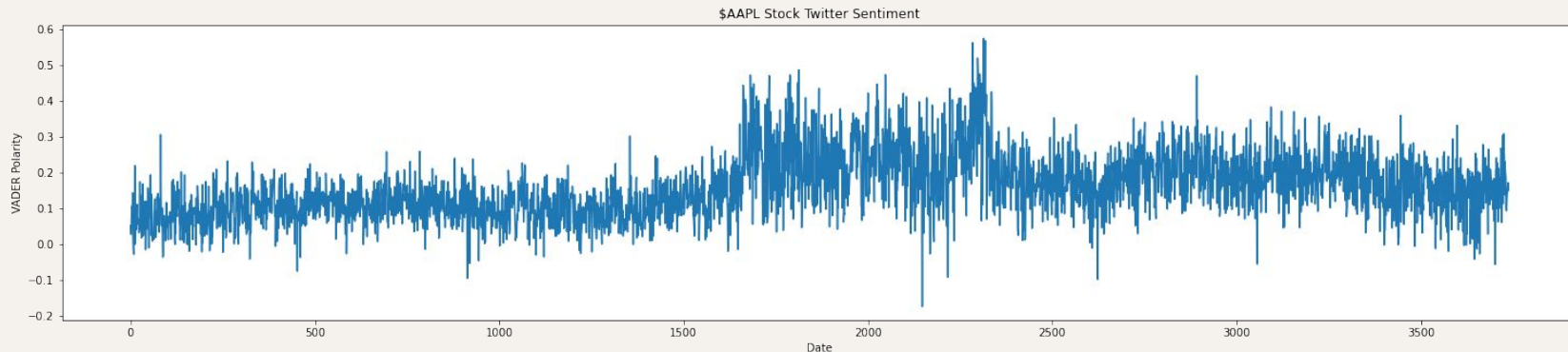Generator Loss
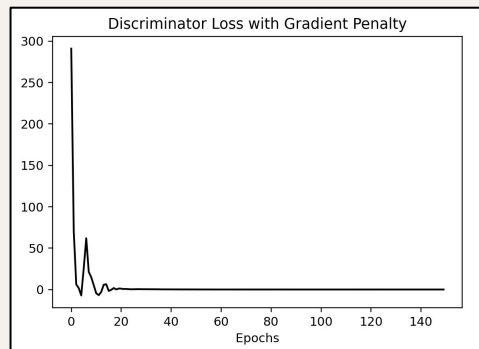
Team Contributions

Andrew Istfan

- Extracted basic stock indicators
- Calculated technical indicators
- Feature reduction
- Autoencoding

Team Contributions

Franco Medrano

- Created Reddit webscraper
- Scraped Twitter data
- Sentiment analysis
- Variable Importance with XGBoost

Tools, Software, Environment

- Python
- Python libraries
  - Pandas, Numpy, Matplotlib, Sklearn, NLTK
  - PRAW, Scweet
  - VarClusHi Clustering Algorithms
- wGAN-GP
  - Implemented in PyTorch
- Jupyter
- TDAmeritrade API

References

[1] Priyank Sonkiya, Vikas Bajpai, and Anukriti Bansal. Stock price prediction using BERT and GAN. Technical report, July 2021.

[2] Ricardo Alberto Carrillo Romero. Generative Adversarial Network for Stock Market Price Prediction, 2019.

[3] Yulian Wen. Research of Stock Price Prediction Based on PCA-LSTM Model. April 2020.