NewsWorthy

Erdos
Data Science
Project
May
2024

Jem Aizen Guhit
Tim Alland
Nawaz Sultani
Ogonnaya Michael Romanus
Kenneth Anderson
Sarasi Jayasekara

Central Question

Can we predict the stock price movements of a company using the sentiment scores of financial news headlines?

Approach

- Explore news sentiment data to find features that correlate with stock price movement.
- Develop and validate predictive models that use data from financial news headlines, alongside features from stock data, to predict the stock price movement of S&P 500 companies (focusing on 15 of them).

Testing

Run a simulation of an investment portfolio directed by the best model, and evaluate how it performs during the test period.

Data Gathering

We used Stock News API to collect news from sources including

The Motley Fool
Investor's Business Daily
Zacks Investment Research
Market Watch
24/7 Wall Street
Reuters
CNBC
Business Wire
Forbes
The Guardian
Fox Business
NY Times
... and more ...

Data Gathering

We collected market data using yFinance, and selected the top 3 sectors based on market weight.

Healthcare	Technology	Finance

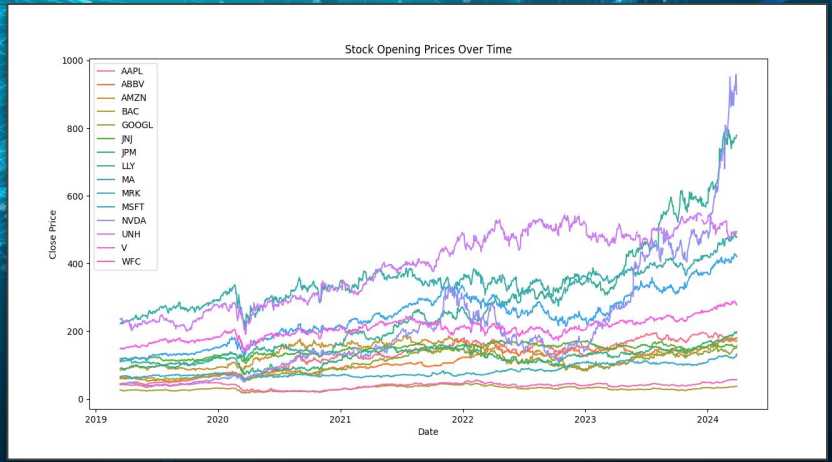
From each selected sector we picked 5 large-cap companies that generate news on a consistent frequency.

Healthcare	Technology	Finance
Eli Lilly & Co	Apple	J P Morgan Chase & Co
United Health	Microsoft	Visa
Johnson & Johnson	Nvidia	Mastercard
Merck & Co Inc	Google	Bank of America
AbbVie Inc	Amazon	Wells Fargo

Data Gathering

We gathered 63704 articles,
along with
Market data spread throughout the timespan of 5 years

March 2019 - March 2024



Initial Decisions

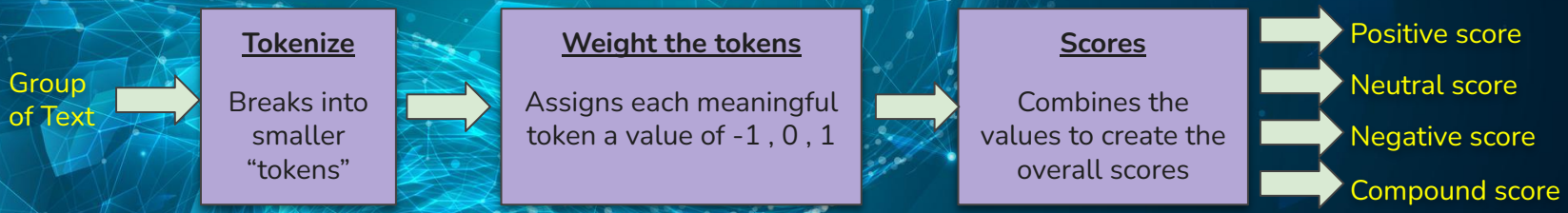
Train - Test - Split

Our dataset included 5 years of stock prices and news headlines.

- We set aside the last year of data as our test set (March 2023 - March 2024)
- From the 4 years in our training set, we used the last year, broken into 4 3-month increments as validation sets.

Sentiment Analysis Tools

Sentiment analysis is the process of analyzing groups of texts (such as sentences or articles) to assign a value to it that reflects how positive, negative, or neutral the overall sentiment seems).

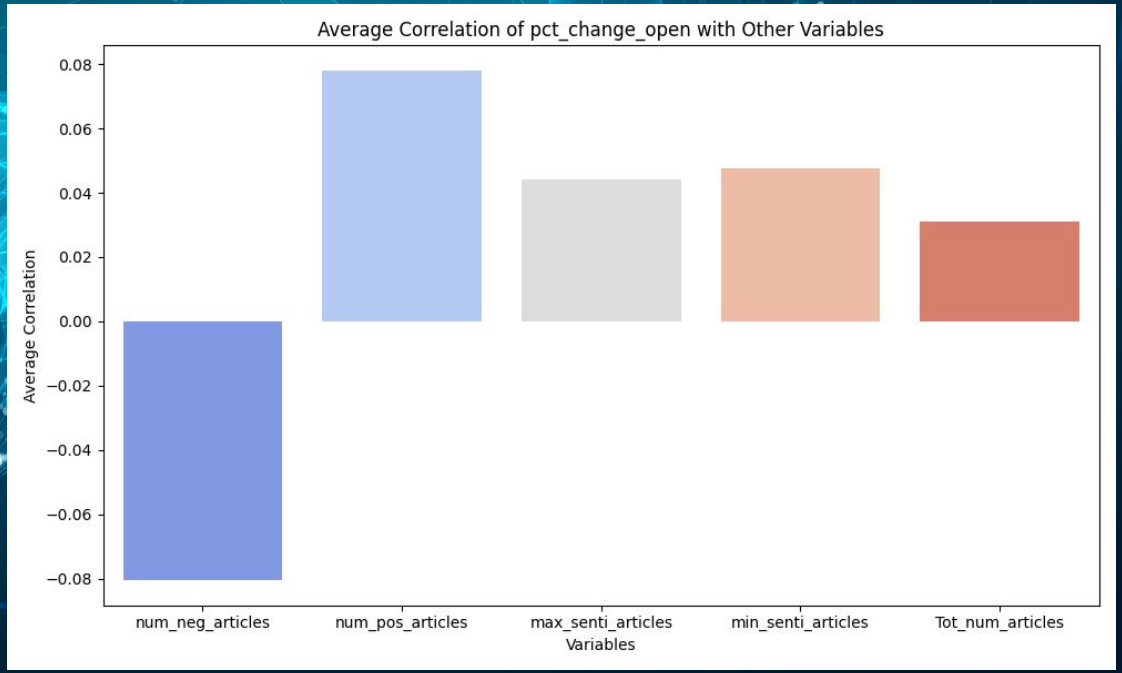


- Vader (Valence Aware Dictionary and sEntiment Reasoner)
  Is an open sourced rule-based sentiment analysis tool. While it's meant to work well with text from social media, it displays mixed performance in domain-specific contexts.
- FinVader
  We used, FinVader, a variant of Vader, which includes finance lexicons.

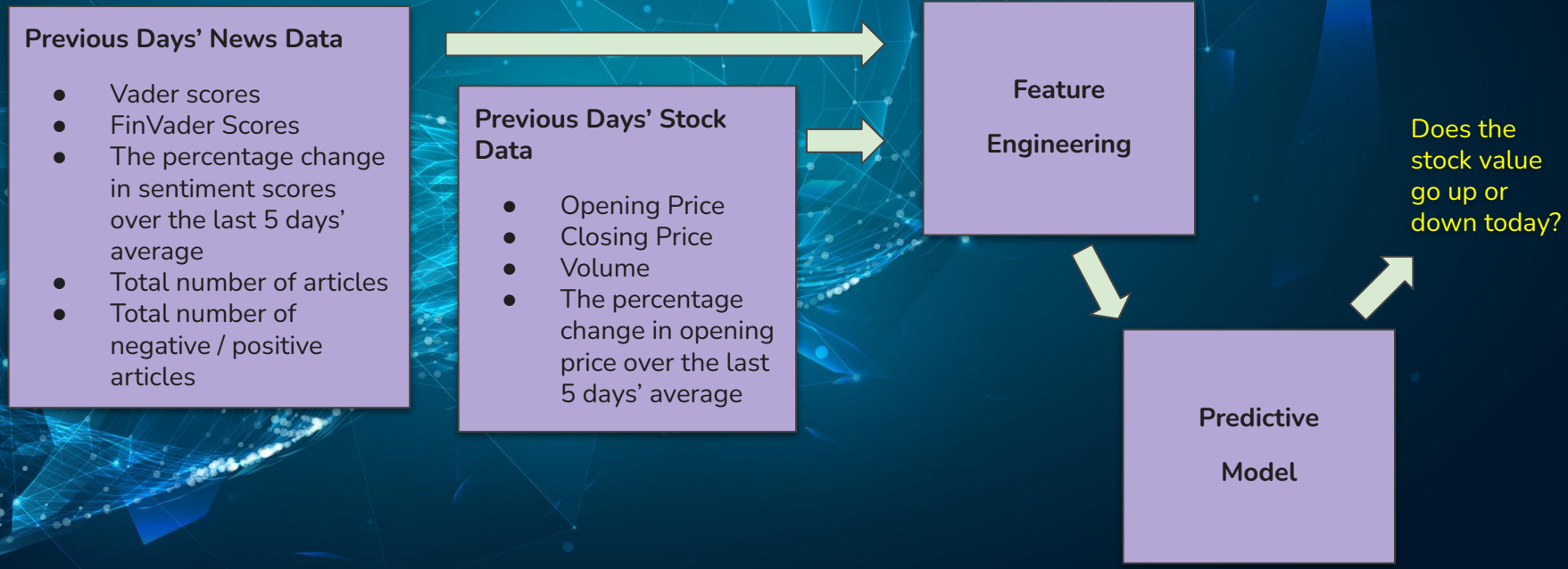
Exploratory Data Analysis

Observations

But in several companies, there are other combinations that seem to show a significant correlation, such as percentage change of opening price vs number of positive articles and number of negative article.

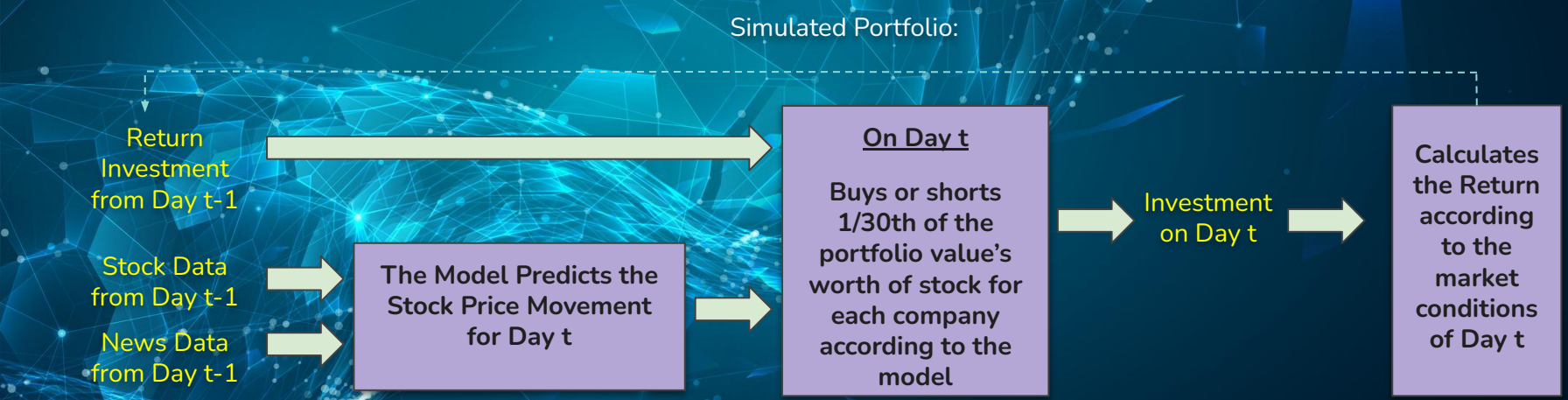


Modeling Pipeline



Stock Portfolio Simulation

KPI: Did we make beyond a —% profit in the simulated scenario?



Models Used

Baseline Model: ARIMA

Model 1

Logistic Regression

Model 2

Gradient Boosted Decision Trees

Model 3

Long Short Term Memory (LSTM)

Stock Portfolio Simulation

KPI: Did we make beyond a —% profit in the simulated scenario?

Model	Best Average Growth (on 3 month validation sets)
Logistic Regression	1.5%
Grad Boosted Trees	2%
XGBoost	1.5%
LSTM	1%
Buy and Hold	1%

Conclusions

And Observations

News Sentiment Matters!

- Our models show improvement in accuracy of stock selection over making random guesses.
- Ubiquity of news presents some challenges with determining how to capture and model sentiment.
- Our study demonstrates the usefulness of news sentiment and highlights the need for future research.

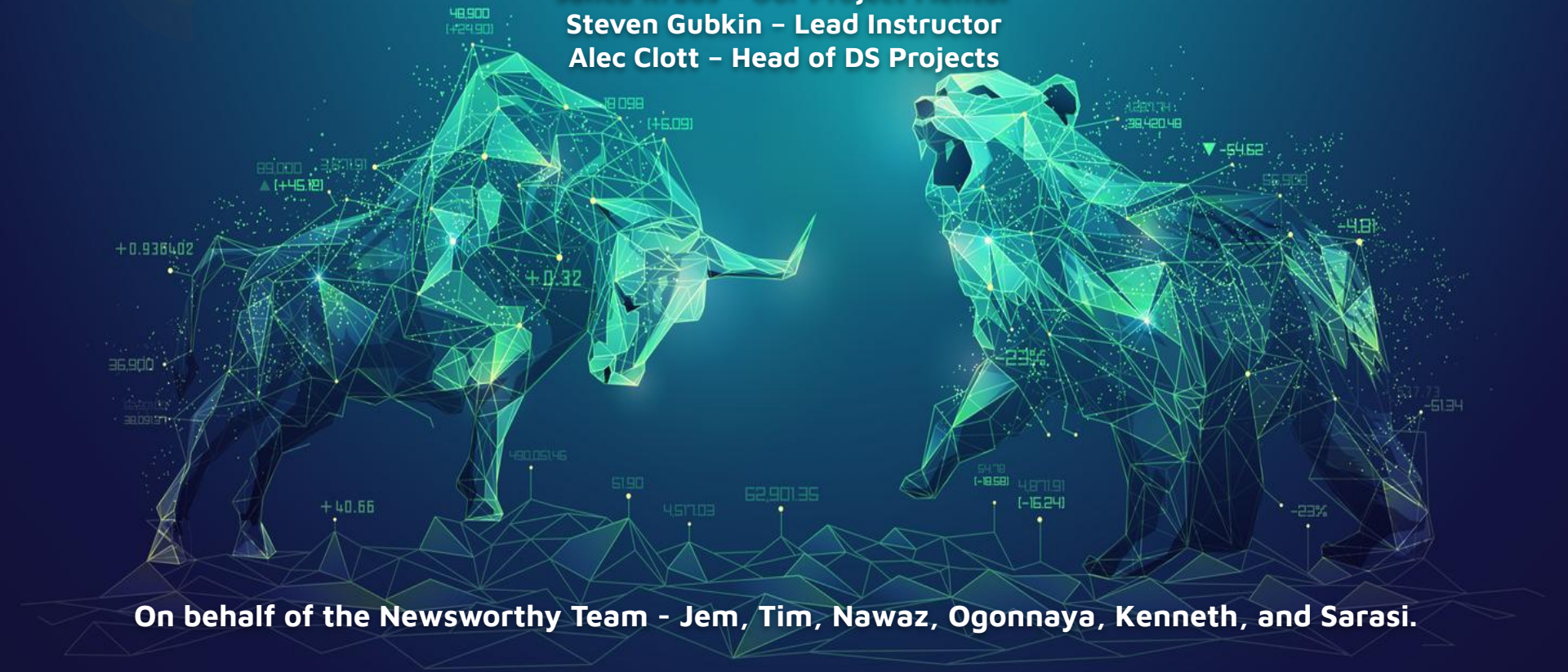
Further Questions

And Other Possible Research Directions

- Future studies should examine the interplay between financial news and social media commentary on stock movement
- Simulations and additional feature engineering that assess the recency and length of news cycles (7 days, monthly, quarterly, yearly, etc) on stock movement may improve model performance

Special Thanks to:

Roman Holowinsky – Director of the Erdös Institute
Janco Kraus – Our Project Mentor
Steven Gubkin – Lead Instructor
Alec Clott – Head of DS Projects



On behalf of the Newsworthy Team - Jem, Tim, Nawaz, Ogonnaya, Kenneth, and Sarasi.