



Retail Trading During the COVID-19 Pandemic

Team 26: Iris Chang, Ruqi Gao, Ga Yeong Lee, Su Min Lee, Sharne Sun

Overview

- Goal: Understand and Predict the Behavior of Retail Traders
- Acquisition and Preprocessing of Data
- Analysis
 - COVID-19 news
 - Financial news
 - Social media (Reddit)
- Conclusion





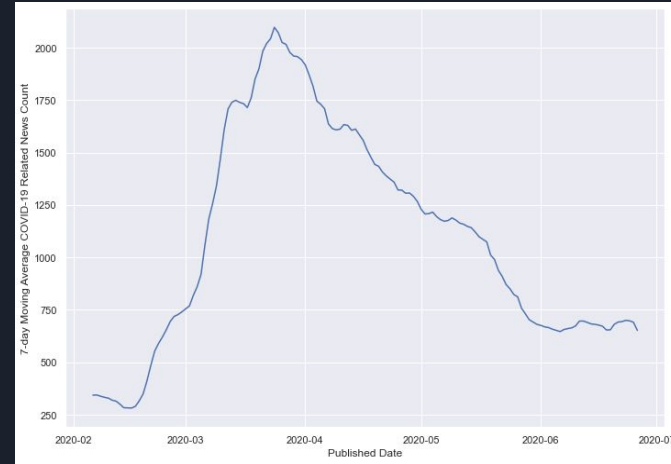
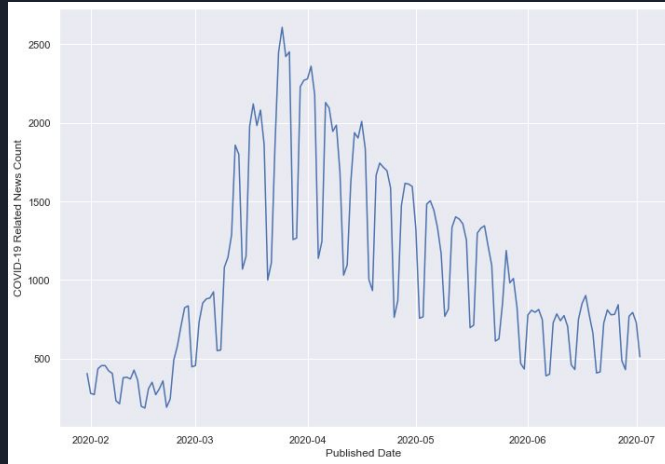
Data and Models

- Data:
 - Robinhood Popularity
 - News Headlines
 - COVID-19 (Kaggle)
 - Financial (finnhub stock API)
 - Reddit (scraped from subreddit WallStreetBets)
- Models:
 - VADER, OLS, SHAP, XGBOOST

COVID-19 News



COVID-19 Related News Headline Count



- The COVID-19 related news count peaks in mid-March
- News count shows weekly periodicity

COVID-19 Related News Headline Sentiment Count and S&P 500 Price Level



- VADER Package
- Increase/Decrease in positive and negative COVID-19 news count are concurrent, and negatively correlated with S&P 500 daily close price
- Average Sentiment Score doesn't explain much about S&P 500 daily close price

Attention-Driven Retail Trading

$$\ln(\text{SPY Popularity}) \sim a + b_1 * \text{Coverage} + b_2 * \text{Lockdown} + b_3 * \text{Reopen} + b_4 * \text{Coverage} * \text{Lockdown} + b_5 * \text{Coverage} * \text{Reopen}$$

$\ln(\text{SPY popularity})$	A: Coverage Only	B: Lockdown	C: Reopen	D: Lockdown and Reopen
<i>Coverage</i>	0.010***[21.6]	0.016***[25.4]	0.009***[21.3]	0.016***[18.6]
<i>Lockdown</i>		14.56***[5.9]		14.55***[7.1]
<i>Reopen</i>			14.55***[6.1]	14.50***[8.3]
<i>Coverage</i> \times <i>Lockdown</i>		-0.016***[-10.4]		-0.016***[-11.1]
<i>Coverage</i> \times <i>Reopen</i>			-0.090***[-3.2]	-0.016***[-7.4]
<i>Adj.R</i> ²	0.756	0.879	0.845	0.918

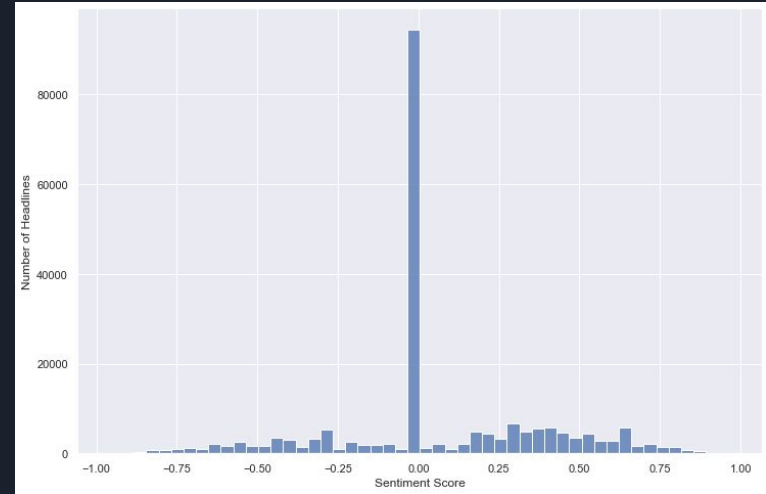
Table 5: This table reports OLS regression results of log number of retail user stock holding accounts on the contemporaneous ratio of COVID-19 related news coverage for the sample from February 1st, 2020 through July 1st, 2020. The dependent variable is the daily log number of Robinhood user accounts holding for S&P 500. Results based on the entire sample period, lockdown period, and reopen period are reported in Panel A, B, and C, respectively. Lockdown is a dummy variable equal to one between March 16th and May 7th. Reopen is a dummy variable equal to one since May 8th. Coverage is the frequency count of COVID-19 related new headlines. The t-statistics reported in square brackets are based on standard errors clustered at firm and day levels. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Financial News



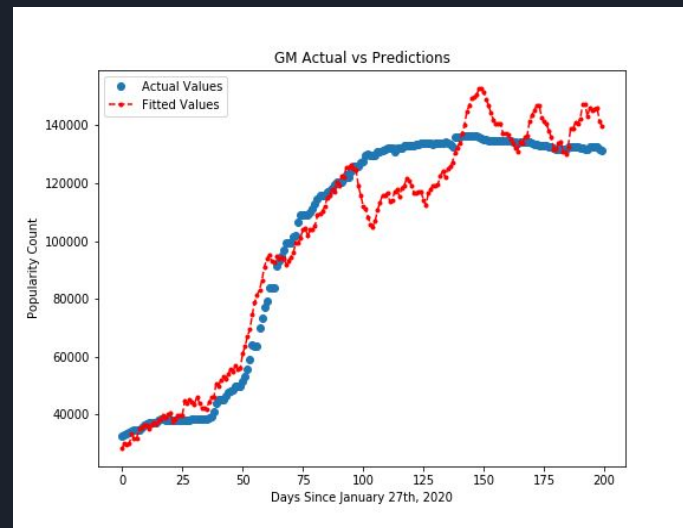
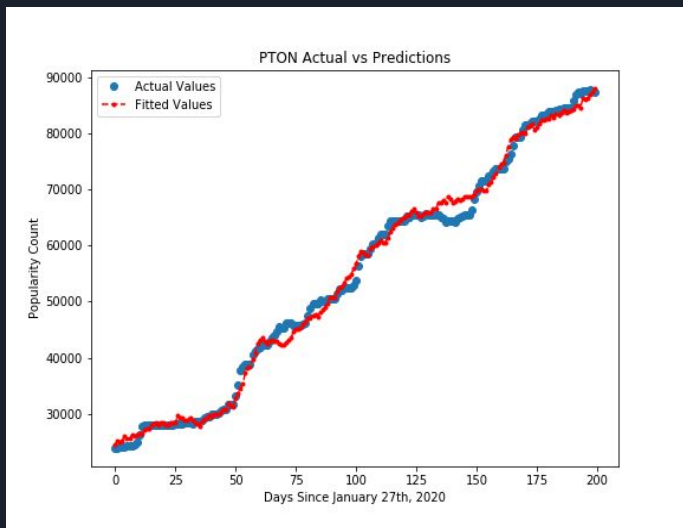
News Headline Sentiment Analysis - Predicting Stock Popularity on Robinhood

- Finnhub Data
 - Headlines labeled with time released and related stock
 - 30 most popular reddit stock
 - 1/15/2020 - 8/15/2020
- Vader Sentiment Score Neutrality
- Robinhood Shareholder Growth



News Headline Sentiment Analysis - Predicting Stock Popularity on Robinhood

$$Popularity_{i,t} = \alpha_i + shareholders_t + neg_{i,t} + pos_{i,t} + \sum_{t-7}^{t-1} \mu_{i,t}$$

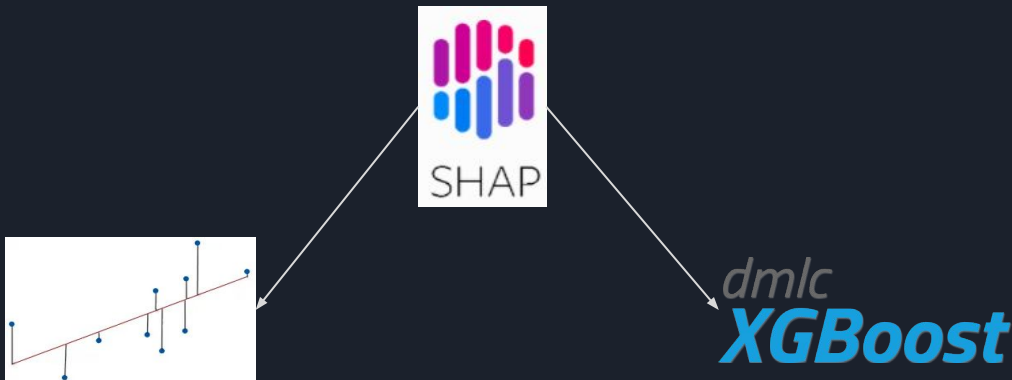


Reddit



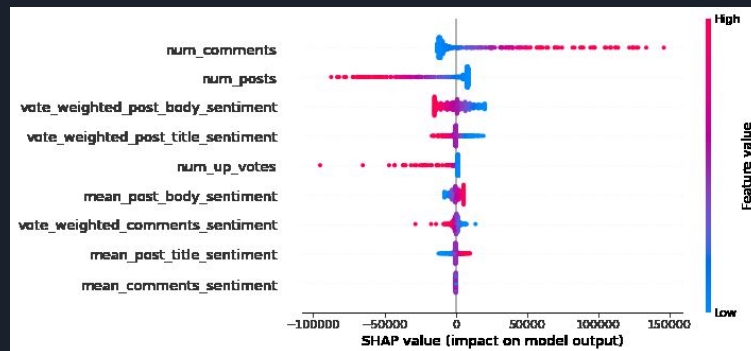
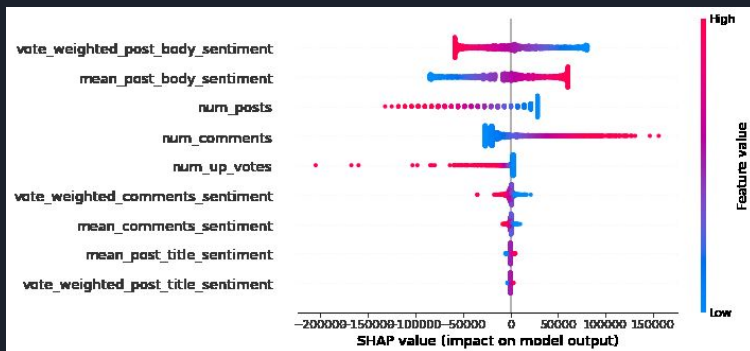
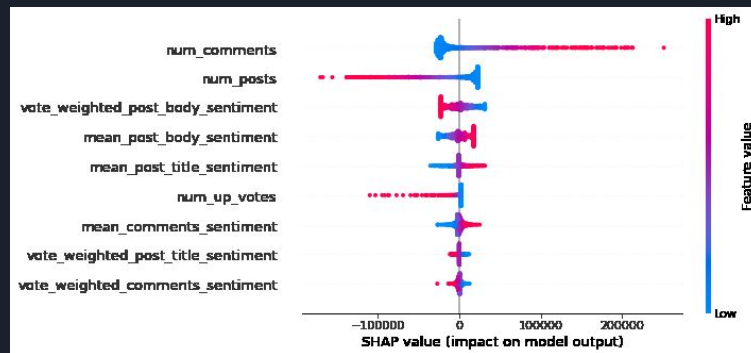
Reddit /r/wallstreetbets Sentiment Analysis - Predicting Stock Popularity on Robinhood

- Analysis
 - Sentiment analysis using VADER
 - SHAP analysis on OLS and XGBoost models for 1day/3day/7day time window averages



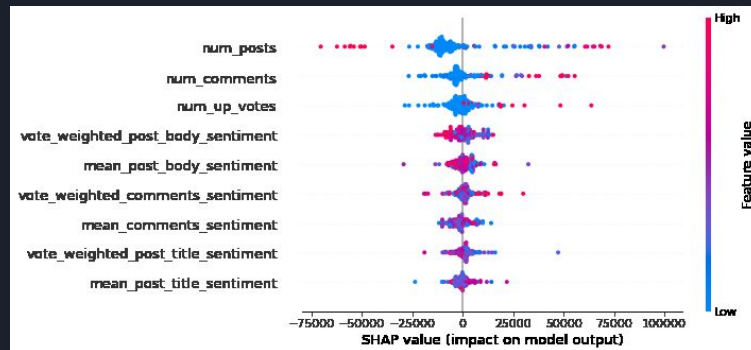
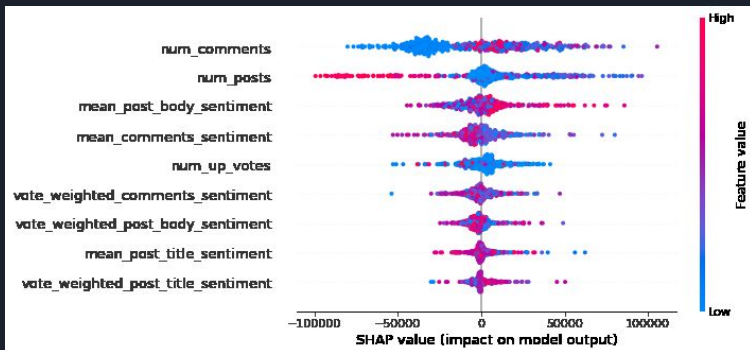
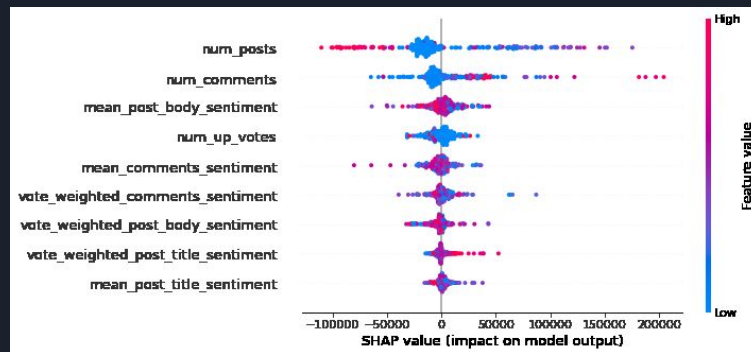
Reddit /r/wallstreetbets Sentiment Analysis - Predicting Stock Popularity on Robinhood

- SHAP summary plots for OLS:
 - Daily (bottom left)
 - 3-day (top right)
 - Weekly (bottom right)



Reddit /r/wallstreetbets Sentiment Analysis - Predicting Stock Popularity on Robinhood

- SHAP summary plots for XGBoost:
 - Daily (bottom left)
 - 3-day (top right)
 - Weekly (bottom right)



Conclusion





Goal Revisited

- Study retail trading during COVID-19 pandemic by investigating the stock popularity on Robinhood



Major Findings

- The S&P 500 Index popularity significantly corresponds to the COVID-19 news headlines, especially during the lockdown period.
- The volume of financial news headlines is more important to the stock popularity on Robinhood, rather than the exact sentiment.
- The number of posts on Reddit are important factors in determining the stock popularity on Robinhood, while the number of upvotes and sentiment scores are not necessarily strong indicators.



Acknowledgements

- Special thanks to our mentor Cristina, our TA Natalie and Correlation One for the Women's Summit experience!



Thank you!



Q & A



References

- CHE Gilbert and Erric Hutto. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) <http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf>, volume 81, page 82, 2014.
- Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pages 785–794, 2016.
- Erik Štrumbelj and Igor Kononenko. Explaining prediction models and individual predictions with feature contributions. Knowledge and information systems, 41(3):647–665, 2014.
- Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In Advances in neural information processing systems, pages 4765–4774, 2017.
- Brad M Barber and Terrance Odean. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. The review of financial studies, 21(2):785–818, 2008.
- Josef Lakonishok, Andrei Shleifer, and Robert W Vishny. The impact of institutional trading on stock prices. Journal of financial economics, 32(1):23–43, 1992.
- Mark Grinblatt, Sheridan Titman, and Russ Wermers. Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior. The American economic review, pages 1088–1105, 1995.
- Richard W Sias. Institutional herding. The Review of Financial Studies, 17(1):165–206, 2004.



Ideation Revisited

- Goal:
Study retail trading during COVID-19 pandemic by investigating the stock popularity on Robinhood.
- Hypotheses formulation:
 - COVID-19 news and attention-driven retail trading
 - Financial news and retail trading on attention-grabbing stocks
 - Social media (Reddit) posts and retail trading on attention-grabbing stocks



Next Steps and Improvements

- Incorporate more stocks
- What can be improved?
 - Sentiment score construction: explore different NLP models
 - Feature extraction: for each feature, include a weight representing contribution of feature to prediction.
 - Time series momentum: for each stock, include its past week returns to control for the tendency that retail traders tend to buy stocks exhibiting extreme returns.

Conclusion

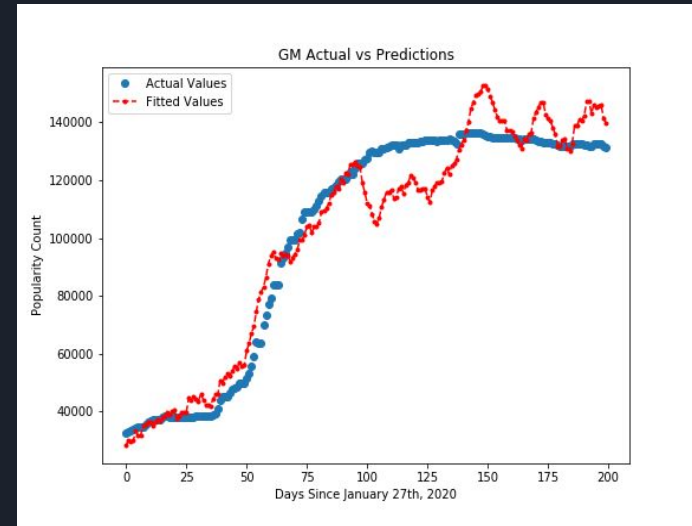
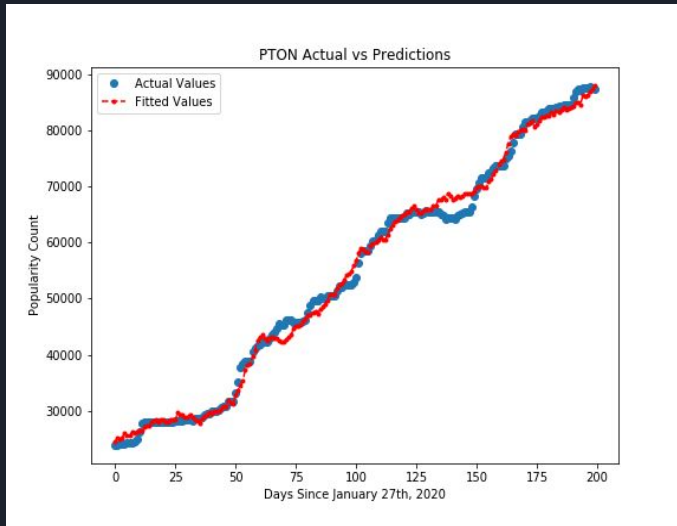
- The S&P 500 Index popularity significantly corresponds to the COVID-19 news headlines, especially during the lockdown period.

ln(SPY popularity)	A: Coverage Only	B: Lockdown	C: Reopen	D: Lockdown and Reopen
<i>Coverage</i>	0.010***[21.6]	0.016***[25.4]	0.009***[21.3]	0.016***[18.6]
<i>Lockdown</i>		14.56***[5.9]		14.55***[7.1]
<i>Reopen</i>			14.55***[6.1]	14.50***[8.3]
<i>Coverage × Lockdown</i>		-0.016***[-10.4]		-0.016***[-11.1]
<i>Coverage × Reopen</i>			-0.090***[-3.2]	-0.016***[-7.4]
<i>Adj.R²</i>	0.756	0.879	0.845	0.918

Table 5: This table reports OLS regression results of log number of retail user stock holding accounts on the contemporaneous ratio of COVID-19 related news coverage for the sample from February 1st, 2020 through July 1st, 2020. The dependent variable is the daily log number of Robinhood user accounts holding for S&P 500. Results based on the entire sample period, lockdown period, and reopen period are reported in Panel A, B, and C, respectively. Lockdown is a dummy variable equal to one between March 16th and May 7th. Reopen is a dummy variable equal to one since May 8th. Coverage is the frequency count of COVID-19 related new headlines. The t-statistics reported in square brackets are based on standard errors clustered at firm and day levels. *, **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Conclusion

- The volume of financial news headlines is more important to the stock popularity on Robinhood, rather than the exact sentiment.



Conclusion

- The number of posts on Reddit are important factors in determining the stock popularity on Robinhood
- The number of upvotes and sentiment scores are not necessarily strong indicators.

