

Reddit: A Novel Online Financial Consultancy

Lydia Manikonda, Shailesh Divey

Rensselaer Polytechnic Institute
{manikl, sdivey}@rpi.edu

Abstract

Social media forums are becoming rich and valuable data sources due to their features and the level of user engagement on these venues. This paper investigates data from online social media forums – which are popular finance-related subreddits to provide answers to five different research questions proposed in this paper. Our analysis revealed five main interesting insights. 1) We found that users primarily seek information and share information on these subreddits. 2) We found that finance-related subreddits are highly negative in sentiment contrary to the popular belief that Reddit posts are in general positive in sentiment. 3) Users on these subreddits do not focus much on their past nor future but their present situation. 4) Even though overall sentiment is negative across the subreddits we considered, word-wise, positive emotion dominates negative emotion. 5) All these four subreddits have large instances of different set of pronouns which provide insights about whom the users are addressing in their posts. Overall, we found the patterns of seeking personal advice as well as sharing experiences of financial decision-making are dominating compared to any other interests.

1 Introduction

Social media platforms have become the de facto medium to seek information, share opinions, and connect with other individuals (Benevenuto et al. 2009; Wilson et al. 2009). The user communication over these platforms for different purposes is generating enormous amounts of rich valuable data that has been used for various applications. Some of these applications include – predicting stock market prices (Bollen, Mao, and Zeng 2011), predicting flu outbreak (Achrekar et al. 2011), real-time analysis about politics (Wang et al. 2012), etc. In order to make these kinds of predictions, analyzing the data to first identify important predictors and variables is crucial. Especially, given the enormity as well as the rich diverse content present in the online social media data helps in better understanding of the offline behavior of individuals and thereby, society at large.

Despite the growing interest in intersection of social networks and capital markets, much of this limited work has been focused on platforms that are not forums (Chen et al. 2014; Hazari and Richards 2011; Way, Wong, and Gibbons 2011; Xu and Zhang 2009). Investigating these do-

main on social media forums designed for discussions have a great potential to reveal important patterns. Reddit (<http://www.reddit.com>) is an online social news website and forum wherein the content is organized in subreddits by the areas of interest. In each subreddit, users can make a post in the form of natural language, web urls, visual entities like images and videos. According to a recent Pew Research report ¹, 11% of adults in United States are active users of Reddit. The widespread usage along with the rich discussions help best answer the research questions we propose.

Using quantitative analysis, we investigate five research questions as follows – **RQ1**. Can we infer the main motivations for users to participate in these subreddits about finance?; **RQ2**. What topics do users discuss on these forums?; **RQ3**. How are their posts aligned with regard to emotions and other psycholinguistic metrics?; **RQ4**. Are these discussions about their past, present or future plans?. Finally, **RQ5**. what are their sentiments expressed through the online posts shared on these subreddits?

We summarize our contributions as follows:

- Main aim of users on these forums is to seek information or share information.
- We found that longer posts (#words/characters) do not necessarily receive higher social engagement. We measured this phenomenon both in terms of active (posts and their replies) and passive (likes) engagement.
- Contrary to the existing literature, posts made on finance-related subreddits are highly negative in sentiment.
- Users on these subreddits do not focus much on their past or future but only their present situation.
- Even though we found the overall sentiment negative across the posts on the subreddits, word-wise, positive emotion dominates negative emotion.
- Topic modeling and word embeddings revealed that users talk about various issues surrounding their personal lives including credit card debt, insurance related to medical, auto, home, HVAC-related issues, student loans, etc. Users emphasize on setting realistic goals to achieve their financial independence and to invest smarter.
- Each subreddit contains content focusing on different types of pronouns which suggest that some forums are

Subreddit	Max Responses	Mean Responses
FIND	497	125
INV	490	25
PF	475	10
SMKT	482	15

Table 1: **Active engagement:** Maximum number of responses received for a post and average number of responses per post in a given subreddit

primarily to seek information and other forums are exclusively to share information.

1.1 Dataset

We crawl data from four subreddits that focus on efficient financial decision-making which aim at different goals – 1) approaches to retiring early (*/r/financialindependence/*), 2) actively participating in financial markets (*/r/investing/*), 3) asset management and financial planning (*/r/personalfinance/*), and 4) discuss market news and strategy (*/r/stockmarket/*). We utilize the Python wrapper for Reddit API (PRAW²) to crawl the Reddit posts and their metadata. The final data has 698 discussion threads from */r/financialindependence* (we denote this as FIND in the rest of this paper), 852 discussion threads from */r/investing* (INV), 908 discussion threads from */r/personalfinance* domain (PF), and 712 discussion threads from */r/stockmarkets* (SMKT).

2 Content Analysis

Social media discussion threads are commonly studied from the perspective of social dynamics (Backstrom et al. 2013). There exists literature (Rubin and Spaht 2010; Spaht and Rubin 2012) in the area of finance about various strategies and options to achieve financial independence as well as ideas to retire early. Alongside, there are numerous financial advising and consulting firms that help individuals personally to achieve their financial goals. However, the subreddits we considered here are popular among online users, and our main goal in this research is to understand what users talk on these subreddits related to finance and how they help each other to achieve their financial goals. We investigate various factors of the users’ language on these communities and extract different patterns to draw insights.

Social Engagement: Longer the post doesn’t mean higher engagement To understand the engagement on these subreddits, we consider both passive (the upvotes and downvotes) and active (responses) engagement metrics (shown in Table 1) on these forums. We also consider the post length statistics in Table 2 in the same context which suggests that indeed FIND is the most active subreddit with not only passive engagement but also in terms of the active engagement even though other subreddits do attract large number of responses only for a specific set of posts.

Motivations of these communities: Acting as online financial consultancies To understand the main motivations of these communities, at a user-level, we first extracted

the other subreddits that these users from our dataset are actively participating in. We found that overall *”/r/AskReddit”* is the most popular subreddit where these users (23.75% of them) are actively making posts. Among the top-10 (AskReddit, personalfinance, Showerthoughts, funny, investing, pics, financialindependence, aww, wallstreetbets, todayilearned) of the active subreddits, 6 of them are information seeking subreddits where most of them are related to finance.

At a post-level, to draw insights about the interests of these communities we first conduct n -gram analysis. N -gram is a sequential list of n words. We use them to encode the likelihood that a phrase reveals the purpose of these subreddits. As part of this analysis, a given sentence has to be split into chunks of consecutive words that are of length ‘ n ’. Due to space restrictions, we show only the list of top-10 *trigrams* in Table 3 that are extracted from the posts made on these four different subreddits separately. The n -grams suggest that users on these forums are primarily seeking information and sharing information. For example, phrases such as *non-investing discussion, helpful retirement accounts, high yield savings*, etc. suggest that these subreddits might be acting as online financial consultancies.

Topics of discussion: Wide spectrum of inquiries regarding advice to earning, spending, saving, budgeting, asset management, trading know-how, etc We utilize Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan 2002) to perform topic modeling. LDA assumes that the document is a mixture of topics where each topic is a mixture of words assigned to a topic, where the topic distribution is assumed to have a dirichlet prior. We use the Python package *gensim* (Řehůřek and Sojka 2010) to perform topic modeling over the online reviews in our dataset. Table 4 shows that users sharing their personal financial details such as loans, insurances, and at the same time posing questions to other users by asking questions such as – “rent or buy a house?”. These kinds of questions as well as sharing their personal financial details along with the concerns are quite evident in these topics. These observations are similar to the other insights we obtained from the above analyses that users are coming to these platforms to seek advice extensively.

Posts are highly negative in sentiment Towards conducting sentiment analysis, we utilize the Stanford CoreNLP library (Manning et al. 2014) to automatically classify the sentiment of subreddit posts. For a given sentence, this package produces the percentage distribution of the sentiment present in the text to generate percentage distributions across the 3 categories of sentiments. Table 5 shows interesting results that majority of the posts shared in these subreddits are overall negative in sentiment. Existing literature (Manikonda et al. 2018; Schrading et al. 2015; Steinbaur 2012) in social media analysis shows that Reddit posts are generally positive or neutral but the sentiments of the posts in our dataset contradict the past insights about sentiments from other subreddits.

Subreddits are all uniquely focusing on a specific pronoun where words in a given post express positive emotion irrespective of overall negative sentiment To investigate different linguistic emotions, we employ the psycholinguistic lexicon LIWC (<http://liwc.wpengine.com/>) on the reviews that are posted online towards different cate-

²<https://praw.readthedocs.io/en/latest/index.html>

Subreddit	Total num of posts	TC	TW	TW. $\mu(\sigma^2)$	CC	CW	CW. $\mu(\sigma^2)$
FIND	87,682	299 (8)	68 (1)	11.8 (56)	18,650 (0)	4062 (1)	67.9 (9498.8)
INV	22,080	297 (4)	58 (1)	12.2 (60.4)	8435 (0)	1761 (1)	52.9 (6100.9)
PF	10,287	295 (15)	71 (2)	11.9 (57.3)	6016 (0)	1357 (1)	76.3 (8412.9)
SMKT	11,021	300 (4)	69 (1)	12.1 (59)	38,908 (0)	9745 (1)	47.0 (16228.0)

Table 2: Aggregated statistics about the length of the post’s title and its content. *TC* #characters in the Title; *TW* #words in the Title; *CC* #characters in the content; *CW* #words in the content. Each cell represents the max (min) of that particular column name unless it is the mean value (variance).

FIND	INV	PF	SMKT
work life balance, part time job, full time job, 401k roth ira, credit card debt, student loan debt, households averse spending	non investing discussion, subprime auto loans, blow day delinquencies, financial crisis peak	credit card debt, helpful retirement accounts, high yield savings, 401 fund selection	high income millenials, long term investment, company last earnings, finance yahoo news, trump tariffs set

Table 3: *tri*-grams extracted from the four subreddits

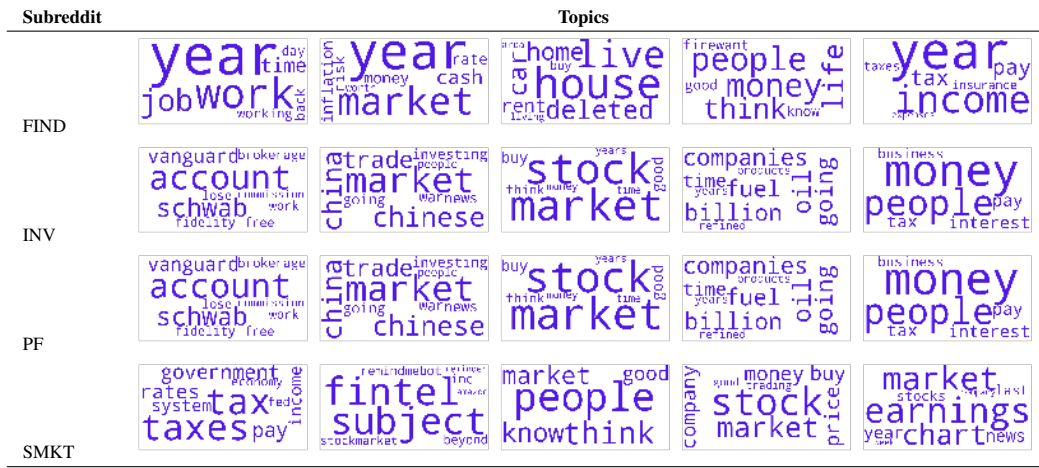


Table 4: Topics compiled, and associated overall posts sentiments for the subreddits

Subreddit	Positive	Neutral	Negative
r/financialindependence	9.6%	14.2%	76.2%
r/investing	8.1%	17.2%	74.7%
r/personalfinance	10.3%	12.9%	76.8%
r/stockmarket	8%	20.4%	71.6%

Table 5: Overall average sentiments scores measured for the posts shared on the 4 Subreddits in our dataset

gories and various brands that we consider as part of our dataset. We focus on mainly 4 different emotional attributes, motivated from the type of interests users exhibit in these domains – *pronouns* (i (1st person singular), we (1st person plural), you (2nd person), she/he (3rd person singular), they (3rd person plural), ipron (impersonal pronouns like it, its, those)), *time orientations* (past, present, future tenses), *affective mechanisms* (Positive emotions, negative emotions, anxiety, anger) and *personal concerns* (Achievement, Leisure, Money). Affect displays the affective processes in terms of how the product in our scenario affects the users. Cognitive mechanisms or processes provides insights about aspects related to cognition such as insight, causation, discrepancy, tentative, certainty, inhibition, etc.

These results show that users on these subreddits do not focus much on either their past or future, but their present situation which we found was very interesting as one of the main goals of these forums is to help users achieve financial independence which happens in the future. Since we are utilizing a word-level psycholinguistic metric, we suspect that this might be due to the fact that users are sharing their present living conditions or income to seek information about achieving financial freedom to retire early. This also shows that users are aggressively focused on seeking feedback on financial conditions.

Co-occurring words: Sharing personal financial details in the community to set realistic goals To detect similar words based on their semantics and syntax, we trained Word2Vec model (Mikolov et al. 2013) using the subreddit posts. The word embeddings suggest interesting insights about how users are openly sharing their financial information to the online community. For example, words such as *fire* are highly associated across all the subreddits with words such as 'goal', 'achieve', 'realistic', 'reached', 'confident' revealing that users want to have realistic goals in order to achieve financial independence and retire early. Similarly,

Lingusitic Dimensions	FIND	INV	PF	SMKT
Pronouns	9.09	8.21	8.91	8.62
i	2.35	1.62	2.13	1.78
we	0.35	0.32	0.24	0.35
you	1.829	1.38	2.01	1.52
shehe	0.3	0.28	0.33	0.31
they	0.42	0.68	0.54	0.67
ipron	3.83	3.91	3.66	3.98
Past Tense	1.73	1.64	1.61	1.66
Present Tense	6.73	6.71	6.99	7.06
Future Tense	0.75	0.83	1.07	0.90
Positive Emotions	3.65	3.23	2.94	3.54
Negative Emotions	1.02	1.08	0.81	1.11
Anxiety	0.15	0.12	0.14	0.12
Anger	0.34	0.38	0.18	0.43
Achievement	1.42	1.34	1.29	1.28
Leisure	0.46	0.34	0.30	0.37
Money	1.85	2.58	3.14	2.61

Table 6: Metric values for different psycholinguistic attributes extracted from the posts shared on the four subreddits.

users share that *401K* should be maximized and they are also contributing to *403b* while discussing various issues pertaining to these retirement plans. We also found that users on these platforms are skeptical in terms of investing in certain places due to their concerns about market crashes and recessions. They also provide details into personal issues such as credit-card debt, and insurance related to medical, auto, home, HVAC-related issues, student loans, etc.

3 Conclusions

We conduct a quantitative analysis on four subreddits related to finance and investments. The results highlight in various ways (user interest level and content-wise) that individuals are coming to these subreddits to primarily seek advice. With over 16 million subscribers, we observe, the topics being discussed cover a wide spectrum of inquiry regarding advice and information to earning, spending, saving, budgeting, asset management, investing in markets, trading know-how, and even the more technical and financial analytics. In addition, financial conversations highlight the feelings of insecurity, which could help explain why we obtain an overall negative sentiment from the financial subreddits. Anonymous forums such as these will allow people interested in pursuing long-term financial independence and stability to learn from shared experiences of other subscribers, thereby leading a wide spectrum of topics of discussion.

References

Achrekar, H.; Gandhe, A.; Lazarus, R.; Yu, S.-H.; and Liu, B. 2011. Predicting flu trends using twitter data. In *2011 IEEE conference on computer communications workshops (INFOCOM WKSHPS)*, 702–707. IEEE.

Backstrom, L.; Kleinberg, J.; Lee, L.; and Danescu-Niculescu-Mizil, C. 2013. Characterizing and curating conversation threads: expansion, focus, volume, re-entry. In *Proceedings of the sixth ACM international conference on Web search and data mining*, 13–22. ACM.

Benevenuto, F.; Rodrigues, T.; Cha, M.; and Almeida, V. 2009. Characterizing user behavior in online social networks. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement*, 49–62. ACM.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2002. Latent dirichlet allocation. In *Advances in neural information processing systems*, 601–608.

Bollen, J.; Mao, H.; and Zeng, X. 2011. Twitter mood predicts the stock market. *Journal of computational science* 2(1):1–8.

Chen, H.; De, P.; Hu, Y. J.; and Hwang, B.-H. 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies* 27(5):1367–1403.

Hazari, S., and Richards, A. R.-A. 2011. A qualitative study of adoption of social media for personal finance investing. In *Proceedings of the Southern Association for Information Systems Conference, Atlanta, GA*. Retrieved from <http://sais.aisnet.org/2011/HazariRichards.pdf>.

Manikonda, L.; Beigi, G.; Liu, H.; and Kambhampati, S. 2018. Twitter for sparking a movement, reddit for sharing the moment: #metoo through the lens of social media. *CoRR* abs/1803.08022.

Manning, C. D.; Surdeanu, M.; Bauer, J.; Finkel, J.; Bethard, S. J.; and McClosky, D. 2014. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, 55–60.

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, 3111–3119.

Řehůřek, R., and Sojka, P. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50. Valletta, Malta: ELRA. <http://is.muni.cz/publication/884893/en>.

Rubin, H., and Spaht, C. 2010. Financial independence: Attainable, maintainable. *Journal of Applied Business and Economics* 11(3):11–18.

Schrading, N.; Alm, C. O.; Ptucha, R.; and Homan, C. 2015. An analysis of domestic abuse discourse on reddit. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2577–2583.

Spaht, C., and Rubin, H. 2012. An investment strategy for financial independence. *Journal of Accounting and Finance* 12(3):38–53.

Steinbaur, T. 2012. Information and social analysis of reddit. Retrieved from *TROYS-TEINBAUER@CS.UCSB.EDU*.

Wang, H.; Can, D.; Kazemzadeh, A.; Bar, F.; and Narayanan, S. 2012. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 system demonstrations*, 115–120. Association for Computational Linguistics.

Way, W. L.; Wong, N.; and Gibbons, D. 2011. Online talk about money: An investigation of interactions around personal finance in social media. *Madison, WI: Center for Financial Security*.

Wilson, C.; Boe, B.; Sala, A.; Puttaswamy, K. P.; and Zhao, B. Y. 2009. User interactions in social networks and their implications. In *Proceedings of the 4th ACM European conference on Computer systems*, 205–218. Acm.

Xu, S. X., and Zhang, X. M. 2009. How do social media shape the information environment in the financial market? *ICIS 2009 Proceedings* 56.