Data Analysis for Prediction of Stock Market Performance

Fernando Nakano
George Mason University
fnakano2@masonlive.gmu.edu

Nanditha Pilli
George Mason University
npilli@masonlive.gmu.edu

Melissa Cirtain
George Mason University
mcirtain@masonlive.gmu.edu

Gabrila Turner
George Mason University
gturne3@masonlive.gmu.edu

***Abstract*—If you are followed the news in US between January and February in 2021, an event occurred on the stock market: a group of people on a subreddit forum WallStreetBets a longstanding subreddit that had over 3.5M Reddit users that discusses highly speculative trading ideas and strategies has caused a huge disruption on the market. On Sep 19, 2020 a member of the group posted a case where the company GameStop (GME), that primarily sells video games and consoles, has been in decline due to the shift from physical media to digital paired with the COVID-19 situation.**

***Index Terms*—Machine Learning, Stock, Finance, Data**

I. Introduction

The proposed topic is to gather data from WSB posts process and then explore and analyze the datasets by using an n-gram to find the most relevant stocks on specific days and correlate the actual stock's market performance based on that data as well as produce various analysis and visualizations to fund trends and further understand the data. The performance of the stock is defined on the close market if it was UP or DOWN on that day.

Our Hypothesis is that the temporal performance of specific stocks on the market can be attributed to the mentions of the same on Reddit's WSB posts.

The Alternate Hypothesis is that there is no correlation between stock performance and Reddit's WSB posts, so we should not try to use this data to predict market performance and instead rely on better studied classic methods for investment. Stockwits and other social media platforms are rich source of data for stock analysis [1]–[18] [19] [**?**]. If the hypothesis proves to be true in being efficient in driving the stock market on specific stocks on the specific stocks, it could be then be used for helping on predicting the market for investors, or the other way around, to prevent people from buying such mentioned ones [20]–[28].

In recent years, natural language processing models provide different tools to analyse finance and stock market data etc [29] [30]–[37].With many outlets talking about this on the news recently, the importance of this work is highly relevant because it will allow us to educate the less informed investors to take the right choice for and to support people's decision if it is worth using this data or staying away from this group information when deciding what are their next step to investing. Irresponsible investing based on Hype can damage a person future and mind and indirectly impact the economy.

II. Literature Research

After listening a lot on the news, one news article got attention posted by BBC(2021) Robinhood: US family sue trading app over son's suicide, the young man mistakenly understood that he owned $700.000 dollars by confusion on how the platform displayed the buying power because Robinhood blocked trading for Game Stop(GME) stocks. This article is very shocking and with more research we can find a lot of information about this case and as Kawa(2020) wrote in Bloomberg on his post "Reddit's Profane, Greedy Traders Are Shaking Up the Stock Market", he highlights that the reddit's board may be reshaping how the options market. This article gives a great insight on how the users of this subreddit board made life of investing veterans harder to deal and Dolbec(2021) article WallStreetBets is disrupting financial markets — possibly permanently comparing what WSB's posts did to what fashion and music had accomplished in the past, completely transforming a market just through the power of highly engaged people sharing a passion.

With that in mind, next literature research was on studying how to predict market, there are innumerous study on this topic, however one article by Investopedia is very informative with the basics of it. Yates (2021) mention 4 Ways to Predict Market Performance, the momentum one is the one that caught attention to be used as based of research of this study, the interesting part is that Studies have found that mutual fund inflows are positively correlated with market returns. Momentum plays a part in the decision to invest and when more people invest, the market goes up, encouraging even more people to buy. It's a positive feedback loop. As mentioned by Yates (2021).

Wallstreet bets reddit board, uses a specific language to indicate what are the users actions. On the explanation of WallStreetBets (2020) - Dissecting the Unique Lingo and Terminology used in the Subreddit r/Wallstreetbets, we can briefly understand the language used and use that to help train a natural language SentimentIntensityAnalyzer with words and phrases to understand the sentiment, for example, the article mentions that dd means that the user will double down all of his investments into a specific stock, or yolo(you only live once), means they will sell every single thing and invest every single penny they have on a specific stock, diamond hands means that the user will hold the stock for as long as possible and paper hands means they sold too early and lost

the momentum, and much more terms.

For Sentiment analysis for trading on top of the topics Swarnkar (2020), explain how Vader is more lightweight than regular Machine Learning algorithms as it won't require large training data and perform with a F1 score of (0.96) while human is (0.84) labeling sentiments on text like tweets and social media as positive, negative and neutral providing a valence score, allowing for simple easy and training with a very small set of data added to the lexicon.

III. Proposed Dataset

For this research we have three main source of data that we are planning to use.

The first dataset is WSB posts, that we have multiple ways of collecting on the web, the first and most accurate method would be to retrieve directly from Reddit using the Reddit API, the problem with this method is that it has some limitations. With enough time and resources, it would be possible to download many years of data without filters of all post boards, however Redis API give you only 60 requests per minute and up to 100 items only on each request and to download 8 years of data with hundreds or thousands of posts daily it would take a very long time without having a server to leave this running. For this reason, we decided to use a publicly available WSB dataset from Kaggle.

The dataset selected for this is the Collection of Kaggle WSB posts on *https://www.kaggle.com/gpreda/reddit-wallstreetsbets-posts*, this dataset is updated daily with all the titles, body, dates, and scores of the posts, with this data we can compare the Boom of Game Stop case and validate the recent accuracy against new stocks mentions.

A. Specific Stock Performance Data

For stock data we will be downloading from Yahoo finance using python, we have a method using python yfinance package to download stock data as

IV. Methodology

Methodology on this study will follow the following steps however it is known that data preparation analysis and evaluation can be repeated many times until we reach an interesting result.

It starts with first collecting Data for WSB and start our data preparation, the initial attempt on the data will comprise of 7 steps, which can be repeated depending on the results of the exploration.

- Data Collection
- Data Normalization
- Tokenization
- Sentiment analysis
- Stock Matching
- Filtering

After data collection, the pipeline will find the list of mentioned stocks on WSB posts by tokenizing with an n-gram posts over the entire dataset using the nlkt Python package, then with the identified items we will match with stock keys using a fuzzy matcher to identify posts with possible stock mentioning. Once that is done, we can then pass a sentiment analysis to look forpositive, negative and neutral sentiments. and then we will have to download historical data from the identified stocks matching the WSB time series range. Now we may need to manipulate the data in a way and merge or aggerate the dataset into useful data structures that will make our data exploration easier, then describe the dataset to try and understand by visualizing the data with different charts matching the stock data. The idea here is to generate a time-series dataset that we can plot together to find possible matches and look for a correlation between them over time, comparing the WSB accuracy before and after the "incident" would help us identify how to proceed further and work on predictions.

Once data is prepared, we need to perform Exploration and identify if further process is needed before proceeding with even more exploration.

If the hypothesis is proved to be valid, we will evaluate multiple algorithms to try to predict if stock performance goes high or low depending on the heat of the comments under the Reddit WSB channel.

V. Data Collection & Exploration

On the Data collection we also need to massage some data. this study will start by exploring two datasets the WallStreetBets posts and list of stocks from NASDAQ and NYSE exchanges.

At this point we have collected and merged a complete List of stocks traded on US market, this is how it is the summary after we merged both stock sets: Next step is to read the Reddit WSB posts into a data frame and peek investigate a few records to see if the parsing went fine. Dataset looks fine, now will show a basic summary of the features that we are planning to use first: Started testing with different methods for parsing, lemmatize, is the initial testing, the idea here Is to lemmatize all the words for each day, to help on data normalization for easier analisys. After Lemmatizing the entire dataset and tokenizing it, we found some mentions of stocks, however the data is still very dirty as the following example: So we need to normalize to remove punctuation and maybe lower_case to match and ignore stop words, and we created a helper function to look up stock codes on a sentence, here is an example usage so far, for this case we did not have to use Lemmatizing.

For this Stock search we combined multiple techniques in this set order.

1) Stock search by Key Symbol
2) Stock Search by Stock description Security Name
3) Fuzzy search on remaining not matching stocks based on stock key Symbol and Security Name.

Now that we can both get stock mentions from a sentence and get stock data information, we will start to work on sentiment analysis.

To further filter data, we will use NLTK VaderSentiment analysis to identify positive, neutral and negative feedback, for Vader, Lemmatization is not required, and in fact since

```
In [181]: # lets see a sneak peek of the data,
          # as you can see, the dataset has a lot of information,
          # but for our case we will start exploring on title, body, timestamp and the unique identifier
          df_wsb.head(5)
```

Out[181]:

	title	score	id	url	comms_num	created	body	timestamp
0	It's not about the money, it's about sending a...	55	l6ulcx	https://v.redd.it/6j75regs72e61	6	1.611863e+09	NaN	2021-01-28 21:37:41
1	Math Professor Scott Steiner says the numbers ...	110	l6uibd	https://v.redd.it/ah50lyny62e61	23	1.611862e+09	NaN	2021-01-28 21:32:10
2	Exit the system	0	l6uhhn	https://www.reddit.com/r/wallstreetbets/commen...	47	1.611862e+09	The CEO of NASDAQ pushed to halt trading "to g...	2021-01-28 21:30:35
3	NEW SEC FILING FOR GME! CAN SOMEONE LESS RETAR...	29	l6ugk6	https://sec.report/Document/0001193125-21-019848/	74	1.611862e+09	NaN	2021-01-28 21:28:57
4	Not to distract from GME, just thought our AMC...	71	l6ufgy	https://i.redd.it/4h2sukb662e61.jpg	156	1.611862e+09	NaN	2021-01-28 21:26:56

Fig. 1.

Fig. 2.

```
In [62]: # As we can see, most stocks are unique on each,
         # which is a good initial sign that the dataset is probably good
         # for now we are going to keep them like this and merge into a second dataframe
         # to be used for lookups in the future.
```

```
In [63]: #First we need to rename different columns to match on concat.
         df_other_list.rename(columns={'ACT Symbol': 'Symbol'}, inplace=True)
         #Now we can concat these datasets into a single stocks dataframe.
         stocks = pd.concat([df_nasdaq_list[['Symbol', 'Security Name']],
                             df_other_list[['Symbol', 'Security Name']]])
         stocks.describe()
```

Out[63]:

	Symbol	Security Name
count	9899	9897
unique	9898	9871
top	File Creation Time: 0222202118:02	NYSE Test One Common Stock
freq	2	6

Fig. 3.

Vader is trained using entire sentences and social media data, special characters and emojis does help on identifying positive, negative and neutral feedbacks.

WallStreet Bets has its own language inside of the forum as mentioned on so we need to teach the vader lexicom to understand what those word means to help classify.

We will create a basic lexicom to teach vader how to be an WSB expert with following word:score to be able to filter positive feedbacks, in the sense that we want people that say it will buy or hold stocks.

After initial run with this we can already see we can find some interesting results on positiveness, with score of 1.0 on positive on following Post that we could interpret as good purchasing action.

VI. Data Analysis

Data Analysis Placeholder

VII. Conclusion

Conclusion Placeholder

VIII. Limitations and Further Research

Further Research Placeholder

References

[1] S. Chen, S. Owusu, and L. Zhou, "Social network based recommendation systems: A short survey," in *2013 International Conference on Social Computing*, pp. 882–885, 2013.

[2] S. Lin, C. Liu, and Z.-K. Zhang, "Multi-tasking link prediction on coupled networks via the factor graph model," in *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society*, pp. 5570–5574, 2017.

```
In [64]: # Now we have our initial stock dataset ready.
         # lets keep this dataset to the side now for later usage.

In [65]: # Next step lets get some r/wallStreetBets dataset, import and do some data exploration.

In [66]: # First import into a Dataframe
         df_wsb = pd.read_csv('reddit_wsb.csv')

In [67]: # lets see a sneak peek of the data,
         # as you can see, the dataset has a lot of information,
         # but for our case we will start exploring on title, body, timestamp and the unique identifier
         df_wsb.head(10)
```

Out[67]:

	title	score	id	url	comms_num	created	body	timestamp
0	It's not about the money, it's about sending a...	55	l6ulcx	https://v.redd.it/6j75regs72e61	6	1.611863e+09	NaN	2021-01-28 21:37:41
1	Math Professor Scott Steiner says the numbers ...	110	l6uibd	https://v.redd.it/ah50lyny62e61	23	1.611862e+09	NaN	2021-01-28 21:32:10
2	Exit the system	0	l6uhhn	https://www.reddit.com/r/wallstreetbets/commen...	47	1.611862e+09	The CEO of NASDAQ pushed to halt trading "to g...	2021-01-28 21:30:35
3	NEW SEC FILING FOR GME! CAN SOMEONE LESS RETAR...	29	l6ugk6	https://sec.report/Document/0001193125-21-019848/	74	1.611862e+09	NaN	2021-01-28 21:28:57
4	Not to distract from GME, just thought our AMC...	71	l6ufgy	https://i.redd.it/4h2sukb662e61.jpg	156	1.611862e+09	NaN	2021-01-28 21:26:56
5	WE BREAKING THROUGH	405	l6uf7d	https://i.redd.it/2wef8tc062e61.png	84	1.611862e+09	NaN	2021-01-28 21:26:30
6	SHORT STOCK DOESN'T HAVE AN EXPIRATION DATE	317	l6uf6d	https://www.reddit.com/r/wallstreetbets/commen...	53	1.611862e+09	Hedgefund whales are spreading disinfo saying ...	2021-01-28 21:26:27
7	THIS IS THE MOMENT	405	l6ub9l	https://www.reddit.com/r/wallstreetbets/commen...	178	1.611862e+09	Life isn't fair. My mother always told me that...	2021-01-28 21:19:31
8	Currently Holding AMC and NOK - Is it retarded...	200	l6ub4i	https://i.redd.it/6k2z7ouo42e61.png	161	1.611862e+09	NaN	2021-01-28 21:19:16

Fig. 4.

```
In [68]: #Now lets run some basic summary, to check on these attributes
         print(df_wsb[['title', 'body', 'timestamp','id']].describe()
               .loc[['count','unique','freq']])
```

```
          title   body timestamp     id
count     36668  18534     36668  36668
unique    35795  18295     27008  36668
freq         37     17        14      1
```

Fig. 5.

```
In [77]: y = df_wsb['title'].iloc[0]
         tokenized = nltk.word_tokenize(y)

In [83]: def lemmatize_words(words):
             lemmatized_words = [WordNetLemmatizer().lemmatize(word, 'v') for word in words]
             return lemmatized_words

         lemmatizer = WordNetLemmatizer()
         print(lemmatize_words(tokenized))
```

```
['It', "'s", 'not', 'about', 'the', 'money', ',', 'it', "'s", 'about', 'send', 'a', 'message', '.', '🚀💎🙌']
```

Fig. 6.

[(15147, '.'),(14362, '!'),(9261, ','),(7874, 'I'),(6771, 'GME'),(6640, '?'),(5172, '$'),(4683, ''),(2911, 'buy'),(2644, 'AMC'),(2394, 'THE'),(2228, 'Robinhood'),(1886, 'HOLD'),(1816, '-'),(1743, 'go'),(1693, 'hold'),(1685, ')'),(1675, "'s"),(1663, '('),(1635, 'stock'),(1635, ':'),(1524, 'get'),(1472, 'The'),(1457, 'sell'),(1388, "n't"),(1335, 'TO'),(1317, 'BUY'),(1296, '...'),(1265, 'share'),(1083, '``'),(1076, 'BB'),(1047, '🚀'),(1016, 'like'),(994, 'still'),(946, 'This'),(936, 'We'),(936, 'NOK'),(915, 'short')...

Fig. 7.

```python
def main():
    print("Create Full Stock List")
    stock_dict_list = get_stock_dict()
    sentence = 'Not to distract from GME, just thought our AMC and BlackBerry or Microsoft'
    print('Looking for sentence: ')
    print('->', sentence)
    sentence = sentence.lower()
    sentence = re.sub(r"[-()\"#/@;:<>{}`+=~|.!?,]", "", sentence)
    stop_words = set(stopwords.words('english'))

    for word in nltk.word_tokenize(sentence):
        if word not in stop_words:
            stock = search_stock(stock_dict_list, word)
            if stock is not None:
                print(word, stock)
```

Fig. 8.

```
/Users/fernak/Git/gmu-ait-580/venv/bin/python /Users/fernak/Git/gmu-ait-580/stocks.py
Create Full Stock List
Looking for sentence:
-> Not to distract from GME, just thought our AMC and BlackBerry or Microsoft
gme GME
amc AMC
blackberry BB
microsoft MSFT

Process finished with exit code 0
```

Fig. 9.

[3] M. Heidari, J. H. J. Jones, and O. Uzuner, "Deep contextualized word embedding for text-based online user profiling to detect social bots on twitter," in *IEEE 2020 International Conference on Data Mining Workshops (ICDMW), ICDMW 2020*, 2020.

[4] Y. Chu, F. Huang, H. Wang, G. Li, and X. Song, "Short-term recommendation with recurrent neural networks," in *2017 IEEE International Conference on Mechatronics and Automation (ICMA)*, pp. 927–932, 2017.

[5] C. Yang, X. Chen, T. Song, B. Jiang, and Q. Liu, "A hybrid recommendation algorithm based on heuristic similarity and trust measure," in *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, pp. 1413–1418, 2018.

[6] S. Ji and J. Liu, "Interpersonal ties and the social link recommendation problem," in *2019 6th International Conference on Systems and Informatics (ICSAI)*, pp. 456–462, 2019.

[7] M. Heidari and S. Rafatirad, "Bidirectional transformer based on online text-based information to implement convolutional neural network model for secure business investment," in *IEEE 2020 International Symposium on Technology and Society (ISTAS20), ISTAS20 2020*, 2020.

[8] J. Wang, H. Song, and X. Zhou, "A collaborative filtering recommendation algorithm based on biclustering," in *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, pp. 803–807, 2015.

[9] S. Chen, S. Owusu, and L. Zhou, "Social network based recommendation systems: A short survey," in *2013 International Conference on Social Computing*, pp. 882–885, 2013.

[10] M. Heidari and S. Rafatirad, "Semantic convolutional neural network model for safe business investment by using bert," in *IEEE 2020 Seventh International Conference on Social Networks Analysis, Management and Security, SNAMS 2020*, 2020.

[11] A. Gatzioura, J. Vinagre, A. M. Jorge, and M. Sànchez-Marrè, "A hybrid recommender system for improving automatic playlist continuation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 5, pp. 1819–1830, 2021.

[12] Z. Liao, Y. Song, Y. Huang, L.-w. He, and Q. He, "Task trail: An effective segmentation of user search behavior," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 12, pp. 3090–3102, 2014.

[13] M. Heidari, J. H. J. Jones, and O. Uzuner, "An empirical study of machine learning algorithms for social media bot detection," in *IEEE 2021 International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021*, 2021.

[14] C.-Y. Chi, Y.-S. Wu, W.-r. Chu, D. C. Wu, J. Y.-j. Hsu, and R. T.-H. Tsai, "The power of words: Enhancing music mood estimation with textual input of lyrics," in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pp. 1–6, 2009.

[15] A. Gatzioura, J. Vinagre, A. M. Jorge, and M. Sànchez-Marrè, "A hybrid recommender system for improving automatic playlist continuation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 5, pp. 1819–1830, 2021.

[16] H. Yang, C. He, H. Zhu, and W. Song, "Prediction of slant path rain attenuation based on artificial neural network," in *2000 IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 1, pp. 152–155 vol.1, 2000.

[17] M. Heidari, S. Zad, and S. Rafatirad, "Ensemble of supervised and unsupervised learning models to predict a profitable business decision," in *IEEE 2021 International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021*, 2021.

[18] M. Thompson, G. Alshabana, T. Tran, and A. Chitimalla, "Predict covid-19 cases using opensky data." http://mason.gmu.edu/ ttran81/, 2021.

[19] M. Heidari, J. H. J. Jones, and O. Uzuner, "Offensive behaviour detection on social media platforms by using natural language processing models," 2021.

[20] L. Kawa, "Reddit's profane, greedy traders are shaking up the stock market." https://www.bloomberg.com/news/articles/2020-02-26/reddit-s-profane-greedy-traders-are-shaking-up-the-stock-market, Feb 26, 2020.

[21] T. Yates, "4 ways to predict market performance." https://www.investopedia.com/articles/07/mean_reversion_martingale.asp, Feb 4, 2021.

[22] P.-Y. Dolbec, "Wallstreetbets is disrupting financial markets — possibly permanently." https://theconversation.com/wallstreetbets-is-disrupting-financial-markets-possibly-permanently-154355, Feb 3, 2021.

[23] J. Brownlee, "Random forest for time series forecasting." https://machinelearningmastery.com/random-forest-for-time-series-forecasting/, November 2, 2020.

[24] W. Daniel, "Wallstreetbets traders are pushing risky stocks to all-time highs." https://markets.businessinsider.com/news/stocks/wallstreetbets-traders-equities-all-time-highs-quotes-forum-explain-phenomenon-2021-1-1030022946, Jan 31, 2021.

[25] BBC, "Robinhood: Us family sue trading app over son's suicide." https://www.bbc.com/news/world-us-canada-55990461, Feb 9 2021.

[26] WallStreetBets, "Dissecting the unique lingo and terminology used in the subreddit r/wallstreetbets.." https://www.wallstreetbets.shop/blogs/news/dissecting-the-unique-lingo-and-terminology-used-in-the-subreddit-r-wallstreetbets, Aug 30, 2020.

[27] N. Swarnkar, "Vader sentiment analysis in algorithmic trading." https://blog.quantinsti.com/vader-sentiment/, May 21, 2020.

[28] M. Heidari, J. H. J. Jones, and O. Uzuner, "Fraud detection to increase customer trust in online shopping experience," 2021.

[29] T. W. Russell, J. T. Wu, S. Clifford, W. J. Edmunds, A. J. Kucharski, and M. Jit, "Effect of internationally imported cases on internal spread of COVID-19: a mathematical modelling study," *The Lancet Public Health*, vol. 6, pp. e12–e20, Jan. 2021.

[30] S. Zad, M. Heidari, J. H. J. Jones, and O. Uzuner, "Emotion detection of textual data: An interdisciplinary survey," in *IEEE 2021 World AI IoT Congress,AIIoT2021*, 2021.

[31] A. Adekunle, M. Meehan, D. Rojas-Alvarez, J. Trauer, and E. McBryde, "Delaying the COVID-19 epidemic in australia: evaluating the effectiveness of international travel bans," *Australian and New Zealand Journal of Public Health*, vol. 44, pp. 257–259, July 2020.

[32] M. Heidari, S. Zad, B. Berlin, and S. Rafatirad, "Ontology creation model based on attention mechanism for a specific business domain," in *IEEE 2021 International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021*, 2021.

[33] M. Bielecki, D. Patel, J. Hinkelbein, M. Komorowski, J. Kester, S. Ebrahim, A. J. Rodriguez-Morales, Z. A. Memish, and P. Schlagenhauf, "Air travel and COVID-19 prevention in the pandemic and peri-pandemic period: A narrative review," *Travel Medicine and Infectious Disease*, vol. 39, p. 101915, Jan. 2021.

[34] M. Heidari and J. H. Jones, "Using bert to extract topic-independent sentiment features for social media bot detection," in *2020 11th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, pp. 0542–0547, 2020.

[35] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. P. y Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M. E. Halloran, I. M. Longini, and A. Vespignani, "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak," *Science*, vol. 368, pp. 395–400, Mar. 2020.

[36] S. Zad, M. Heidari, J. H. J. Jones, and O. Uzuner, "A survey on concept-level sentiment analysis techniques of textual data," in *IEEE 2021 World AI IoT Congress,AIIoT2021*, 2021.

[37] M. Heidari and S. Rafatirad, "Using transfer learning approach to implement convolutional neural network model to recommend airline tickets by using online reviews," in *2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMA*, pp. 1–6, 2020.