Dec 11th, 12:00 AM

Detection of Stock Manipulation Influencer Content using Supervised Learning

Frederic Haase
University of Cologne, haase@wim.uni-koeln.de

Oliver Rath
Cologne Institute for Information Systems, rath@wim.uni-koeln.de

Julia Lauten
University of Cologne, lauten@wim.uni-koeln.de

Detlef Schoder
University of Cologne, schoder@wim.uni-koeln.de

Detection of Stock Manipulation Influencer Content using Supervised Learning

Completed Research Paper

Frederic Haase
University of Cologne
Cologne, Germany
haase@wim.uni-koeln.de

Oliver Rath
University of Cologne
Cologne, Germany
rath@wim.uni-koeln.de

Julia Lauten
University of Cologne
Cologne, Germany
lauten@wim.uni-koeln.de

Detlef Schoder
University of Cologne
Cologne, Germany
schoder@wim.uni-koeln.de

Abstract

In recent years, social media influencers have emerged as key players in stock manipulation schemes. Despite their growing impact, methods to detect such activities remain scarcely explored. In this study, we examine the social media content of stock manipulation influencers (SMIs) implicated in a $100 million fraud case by the U.S. Securities and Exchange Commission (SEC) in 2022. Leveraging natural language processing (NLP) techniques, we first investigate the linguistic characteristics present in the social media content published by SMIs. Next, we develop and evaluate supervised learning models to detect manipulative content. Our results have significant implications for investors, regulators, and the broader financial community. They reveal the unique linguistic characteristics of SMI content and demonstrate the potential of machine-learning and deep-learning-based techniques in advancing fraud detection systems.

Keywords: stock manipulation influencers, financial fraud, supervised learning, finfluencers

Introduction

In 2022, the U.S. Securities and Exchange Commission (SEC) charged several social media influencers involved in a $100 million securities fraud that exploited the social media platforms Twitter (now: X) and Discord (SEC, 2022). These individuals accumulated hundreds of thousands of followers by posting information on social media platforms (see illustrative tweets in Table 1). They strategically purchased small cap stocks and then encouraged their followers to buy those stocks by posting price targets (signals) or indicating that they would buy or hold. However, when stock prices increased, the individuals sold their shares without disclosing their intent. Ultimately, these individuals used social media to lure investors, resulting in approximately $100 million in fraudulent profits.

As the fraud case of the SEC against social media influencers demonstrates, stock manipulation continues to be a problem in financial markets. The rise of social media has intensified this issue by providing an accessible medium for such schemes to be initiated and spread (Siering, 2019). In a stock manipulation, individuals or groups deliberately influence the market to profit financially by spreading information or fabricating demand, artificially inflating the price of a stock (Siering et al., 2017). Social media platforms such as Twitter, Discord, and Reddit have become prime platforms for sharing financial news and information, as well as engaging in discussions about financial markets (Hu & Tripathi, 2016). Consequently, these platforms have become major targets for individuals seeking to initiate stock manipulation, a development also noted in an SEC investor alert and bulletin (SEC, 2015).

"And I'm not pumping and dumping on anyone. Why would I sell when I truly believe $RIBT gets over $1.50-$2. Haters mad to see anyone winning and come out if the stock is down 3%. All I post is information about it."	*"$CEI I added for a swing. Volume is starting to come into pennies this week and this is very cheap. On the daily chart we have formed a double bottom. Last time it was at these prices we went to $3. So that leaves a lot of range"*
"ALZN great daily volume here. I think this one is setting up to breakout over 12 today. I mentioned my swing adds yesterday and I added more this morning."	*"NO sleep just waking up too lol $ABVC looking strong still!!$4/$4.50 breakouts next."*

Table 1. Illustrative SMIs tweets (adapted from SEC, 2022).

Previous research on detecting stock manipulation has effectively utilized linguistic features of textual content in traditional news sources, such as news articles or newsletters (e.g., Clarke et al., 2020; Siering, 2019; Siering et al., 2021). However, the increasing involvement of social media influencers in stock manipulation schemes (e.g., SEC, 2022) highlights the need for deeper investigation into deception for stock manipulation within social media content. While machine learning (ML) techniques have been explored to some extent in this area (e.g., Clarke et al., 2020; Siering et al., 2021), deep learning (DL) models, proven valuable in financial text classification (Mishev et al., 2020), remain underexplored for stock manipulation detection on social media platforms. These factors call for a more comprehensive investigation into the linguistic features of stock manipulation influencer (SMI) content and a systematic comparison of ML and DL models for detecting SMI content. Therefore, our study investigates the following research questions:

RQ1: *How do linguistic characteristics in manipulative social media content of SMIs differ from those present in non-manipulative social media content?*

RQ2: *Which ML or DL techniques prove most effective in detecting manipulative social media content of SMIs?*

By investigating these research questions, we aim to provide insights that allow to design, develop, and configure efficient fraud detection systems that can safeguard investors, inform regulators, and ensure the integrity of the broader financial market in an evolving digital landscape. To address these objectives, we assembled a tweet dataset derived from the SEC (2022) case for analysis. Leveraging natural language processing (NLP) techniques, we first investigated the linguistic characteristics present in manipulative social media content of SMIs. Subsequently, we developed and evaluated supervised learning models based on ML and DL techniques for detecting manipulative content.

Addressing **RQ1**, our analysis identified significant differences in the linguistic characteristics employed by SMIs compared to a sample of non-manipulative content. For instance, we observed that SMIs tend to use significantly fewer words, employ more informal and personal language, and utilize more language indicative of leadership and status. While some of these linguistic characteristics align with prior research on deceptive content, others appear to be distinct to SMI content. Regarding **RQ2**, our findings suggest that models incorporating recent advancements in NLP based on DL achieve superior performance in detecting SMI content, an aspect that has not been sufficiently explored in previous research.

Literature Background

Stock manipulation influencers

The impact of financial influencers on social media and their influence on financial markets, stocks, and other assets has been an area of growing interest. Some well-known cases of individuals whose social media activities affected the financial markets include for example Elon Musk, whose tweets have been found to affect the market and Twitter stock (Strauss & Smith, 2019), as well as Tesla's stock price (McCabe, 2022). Another prominent example is John McAfee, who was charged with securities fraud for promoting false and misleading messages to encourage followers to invest in cryptocurrencies, which he later sold after the prices increased (Robertson, 2021). The New York Times also reported that many celebrity marketers for cryptocurrencies failed to disclose their financial interests in these assets (Yaffe-Bellany, 2022).

Recently, scholars have started to conceptualize the role of social media participants with influence on financial markets as "Finfluencers" (Guan, 2023; Haase et al., 2023; Pflücke, 2022). Finfluencers are

individuals who exert significant influence in financial markets through their social media presence. We consider SMIs a specific subset of Finfluencers, as it is important to note that not all Finfluencers share the same intentions. While some Finfluencers are motivated exclusively by stock manipulation for profit, others strive to maximize popularity, provide entertainment, "grow their brand", and contribute positively to market discussions (Guan, 2023). As such, Finfluencers encompass a more diverse range of individuals, with SMIs representing a specific subset who concentrate on stock manipulation activities.

Regulatory oversight has started to address the potential misuse of social media by Finfluencers. Discussions on how to regulate the activities of Finfluencers are underway, and it is investigated whether the current legal frameworks sufficiently protects investors (Pflücke, 2022). As social influence and social bots can be combined to orchestrate fraudulent schemes like pump and dump (Tardelli et al., 2022; Mirtaheri et al., 2021), it is crucial for regulators to develop effective measures to safeguard investors from the potentially harmful effects of these financial influencers' activities.

Detection of financial fraud based on linguistic characteristics

The detection of financial fraud has been an area of extensive research, with data mining techniques being applied to address various types of financial fraud. Ngai et al. (2011) provided an overview of these types, identifying major categories such as bank fraud, insurance fraud, securities and commodities fraud, and other finance-related fraud. Among these categories, our study focuses on securities and commodities fraud, which has received less attention from scholars (Ngai et al., 2011; Siering et al., 2021).

Allen & Gale (1992) defined three categories of stock market manipulation schemes: information-based, trade-based, and action-based manipulations. While trade-based manipulation (e.g., Felixson & Pelli, 1999) and action-based manipulation (e.g., Öğüt et al., 2009) have been extensively studied, information-based manipulation has gained increasing attention in recent years due to the spread of fraudulent stock recommendations on the internet (Siering et al., 2017). Our study investigates a special case of information-based manipulation by focusing on the role of social media influencers who use their online presence to spread information about stocks. Unlike traditional manipulations involving the dissemination of information through news articles or newsletters (e.g., Clarke et al., 2020; Siering, 2019), our focus is on the communication of these individuals provided through their social media presences. Social media content differs from traditional news articles or newsletters, as it does not pretend to be a legitimate news source, unlike financial fake news (Clarke et al., 2020). Instead, social media content often reflects personal opinions and insights (Boyd & Ellison, 2007). As a result, the detection of stock manipulation in social media communication presents unique challenges compared to traditional information-based manipulation techniques.

The use of linguistic features has shown to be useful in the detection of financial fraud. By examining the textual content from various sources, researchers have been able to identify linguistic characteristics indicative of fraudulent activities. Table 2 presents a summary of exemplary studies that have utilized linguistic features of various content in detecting different types of financial fraud. Clarke et al. (2020), Siering et al. (2021) and Siering (2019) primarily relied on traditional news sources, such as news articles and stock-promoting newsletters, which differ substantially from social media content as described earlier. Siering et al. (2021) made use of ML models to demonstrate that a combination of linguistic features and bag-of-words (BOW) models can be used to robustly detect stock manipulation fraud. Polarity and entropy, i.e., high information content, stand out as linguistic features identifying fraudulent messages. Craja et al. (2020) showed that their DL model, a hierarchical attention network, outperforms the BOW models in the context of fraud detection in annual reports. The model identifies "red flag" sentences in these reports that hint at potential fraud. The ML model introduced by Dong et al. (2018) uses a classifier for fraud detection utilizing financial ratios and messages from internet financial news boards. The team highlights that linguistic features increase the performance of their classifier.

Research teams also investigated more diverse types of content. Lauwerse et al. (2010) predicted fraudulent events by analyzing linguistic features in email conversations between employees. They found among other things, that a higher frequency of adjectives used in conversations can be linked to fraudulent events. Throckmorton et al. (2015) used a set of ML models as classifiers for fraud detection based on the feature categories accounting risk, acoustic features, and linguistic features derived from earnings conference calls. The team did not find support for the usefulness of linguistic features in detecting fraud. Self-disclosed information by financial intermediaries on LinkedIn was analyzed by Lausen et al. (2020). More ambiguous

content and longer profile summaries were identified as linguistic characteristics helpful in identifying misconduct.

The work by Clarke et al. (2020) is closely related to our research. The team used features based on the LIWC 2015 dictionary (Boyd et al. 2022), demonstrating that a set of ML models can be trained for detecting disinformation. They found the gradient boosting classifier to be most effective with an F1 score of 88.7%, with word count and words per sentence being highly relevant predictors. We intend to expand on the findings above in two ways. First, by analyzing social media content with distinctive features and the role of influencers. Second, the exemplary studies listed in Table 2 mainly employ ML techniques leveraging textual content, but do not provide a comparison against DL techniques. DL has shown to be highly effective in other domains, particularly for text classification in NLP tasks (Devlin et al., 2019; Mishev et al., 2020). Therefore, in addition to previous research, our research aims to provide a systematic comparison of ML models against DL models.

Exemplary References	**Type of Financial Fraud**	**Content**	**ML**	**DL**
Craja et al. (2020)	Financial Statement Fraud	Annual Reports	X	X
Dong et al. (2018)	Corporate Fraud	News articles (SeekingAlpha)	X	
Louwerse et al. (2010)	Accounting Fraud	Emails	X	
Lausen et al. (2020)	Intermediary Fraud	Social media data (LinkedIn)	X	
Throckmorton et al. (2015)	Corporate Fraud	Earnings conference calls	X	
Clarke et al. (2020)	Stock Manipulation Fraud	News articles (SeekingAlpha)	X	
Siering et al. (2021); Siering (2019)	Stock Manipulation Fraud	Newsletters promoting stocks	X	
This study	*Stock Manipulation Fraud*	*Social media data (Twitter)*	X	X

Table 2. Exemplary financial fraud detection methods based on linguistic features.

Theoretical Background

Language is a crucial component of communication, enabling individuals and communities to express information and ideas. In deceptive communication however, individuals may manipulate language to conceal or distort the truth. Therefore, analyzing textual content is essential in detecting linguistic characteristics of deceptive communication. Siering et al. (2016) offer an in-depth review of theories and models related to fraud detection based on communication. For further details, we encourage readers to refer to their work.

One theoretical framework used to explain deceptive communication is information manipulation theory (IMT) (McCornack, 1992). According to IMT, individuals who engage in deceptive behavior violate four basic communication principles: they may exaggerate or understate information to hide or distort the truth, alter the quality of information or lie, take information out of context, or intentionally communicate ambiguously to confuse the receiver. Such tactics can result in language use distinct from non-deceptive communication, which can be identified through linguistic analysis. Our research takes the theoretical lens of IMT, and consequently, we expect differences in the linguistic characteristics between stock manipulator content and non-manipulative content. These communication deviations are anticipated to result in distinctive linguistic patterns that can be used to identify SMI content.

Alongside deceptive communication, previous research has investigated the use of language within financial social media communities, focusing on the unique jargon and terminology employed by members of different subcommunities. For instance, prior studies have discovered that members of financial subcommunities use specialized language to express their identity and reinforce their sense of belonging within the community (e.g., Mancini et al., 2022). Agrawal et al. (2022) found that certain financial subcommunities on social media platforms exhibit distinctive characteristics, such as a high level of emotionality and high levels of informal language. The intergroup distinctiveness theory (IDT) (Tajfel & Turner, 1979) suggests that distinctive language is used to maintain intergroup distinctiveness and differentiation from the out-groups, which explains the distinctive linguistic characteristics observed

between different groups. IDT suggests that SMIs may adopt a unique language or communication style to distinguish themselves from other social media users or financial communities. We consider SMIs as a subcommunity and consequently anticipate distinctive linguistic characteristics from other individuals in their communication.

In summary, drawing on both IMT and IDT theories, we anticipate differences in the linguistic characteristics of SMI content compared to non-manipulative content. These differences will be subject to empirical investigation in our subsequent analysis. Our conceptual model described before is illustrated in Figure 1.

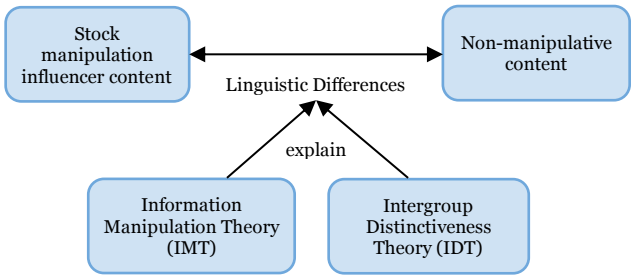

Figure 1. Conceptual model.

Research Methodology

To examine our research questions, we employed the well-established knowledge discovery from databases (KDD) process proposed by Fayyad et al. (1996). This process model is particularly suited for academic research settings and data mining tasks that require significant data preparation (Kurgan & Musilek, 2006). Our methodology involves data acquisition to construct a dataset of SMI tweets and control tweets. Subsequently, we pre-process our dataset of tweets and select appropriate techniques to evaluate supervised learning classifiers derived from the data. A summary of the entire data mining process is illustrated in Figure 2.

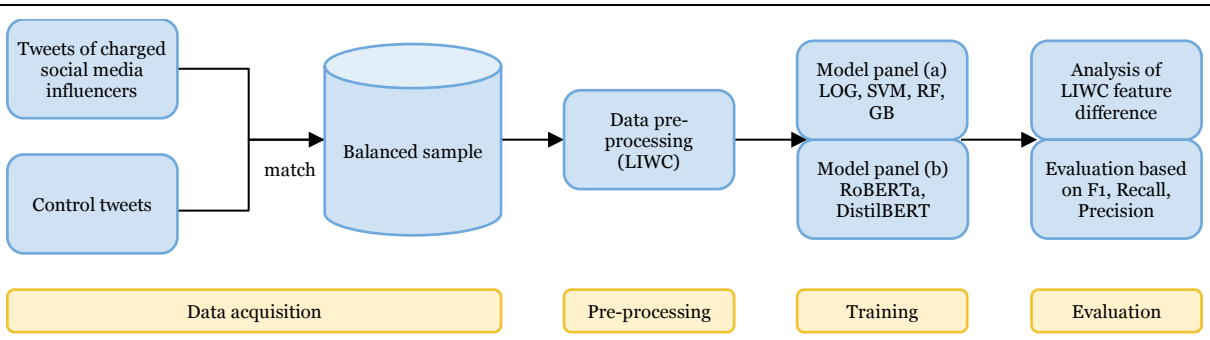

Figure 2. Data mining process based on KDD.

Data Acquisition

For dataset acquisition, we collaborated with Stockpulse, a social media analytics company specializing in financial data. The company continuously collects and analyzes data from major social media platforms, e.g., Twitter, Reddit and Discord, providing a comprehensive source for our analysis. Importantly, we were granted access to the historical archive of Stockpulse, enabling us to retrieve data for tweets that have already been deleted, which is particularly relevant for our study, as many of the manipulative tweets have been removed from the Twitter platform.

We identified all stock manipulating Twitter users from the SEC complaint (SEC, 2022) and gathered 14,625 tweets published from early 2020 through the end of 2022. We retained a total of 3,720 tweets after filtering out retweets and tweet responses, to only keep original tweet creations. To further refine our sample, we kept only those tweets containing a single cashtag mention (n=2028). A cashtag is a specialized term used on social media platforms like Twitter, consisting of a stock's ticker symbol preceded by a dollar sign (e.g., $AAPL for Apple Inc.). This selection allowed us to filter out cashtag piggybacking (Cresci et al., 2019), where multiple cashtags are mentioned in one tweet, making it challenging to sample control tweets.

For each tweet in the resulting dataset of SMI tweets, we sampled a random control tweet that mentioned the same cashtag and was not a retweet or response, and not from a user listed in the SEC complaint (SEC, 2022). Moreover, we ensured that the control tweet was published within one day before or after the tweet's publication date to minimize temporal effects on the sampling. The control tweet was only selected if its creator had at least the same number of followers as the matched SMI tweet, a crucial factor in eliminating bots from our control tweets. If no suitable control tweet was found with the specified requirements, we excluded the original tweet from our sample. The matching of control tweets resulted in a balanced sample of tweets, comprising two sets of tweets with 564 tweets each.

For our sampling of random control tweets, we assumed that tweets not flagged by the SEC's charge list are non-manipulative (backed by a similar approach from Clarke et al. (2020)). This could potentially introduce a noisy label problem, as control tweets might indeed be manipulative. However, considering the SEC's rigorous oversight, the manipulated tweets targeting small cap stocks with low to medium tweet volume, and the close temporal proximity to SMI tweets, we supposed that comprehensive checks were conducted not only on content from the charged influencers but also on our control tweets. In addition, our matched control tweets consisted only of content that is still accessible and has not been deleted, unlike most SMI tweets, which have already been deleted (518/564=92%) By focusing only on non-deleted tweets, we minimize the risk of including tweets that might have been identified as manipulative by the SEC, other market authorities, or the Twitter platform.

Table 3 presents the statistics and properties of the matched control tweets compared to those from the SMIs. The proposed sampling approach ensured two sets of tweets with similar characteristics.

Metric	**SMI tweets (n=564)**	**Control tweets (n=564)**
Average number of followers at tweet creation	121,606	146,212
Average number of retweets at data acquisition	2.52	2.74
Average number of friends at tweet creation	251.83	243.44

Table 3. SMI and control tweets statistics.

Pre-processing

To pre-process our dataset, we employed the Linguistic Inquiry and Word Count (LIWC 2022) tool (Boyd et al. 2022). The LIWC 2022 tool assesses the key psychometric features of language in a given text using a comprehensive dictionary that encompasses more than 12,000 words, word stems, and emojis. It was developed on a corpus that includes, among others, tweets (Boyd et al. 2022), making it particularly suitable for analyzing our data. The output generated by LIWC 2022 comprises a distinct set of 118 variables. To enhance the effectiveness of the applied ML techniques, as described in the subsequent sections, we standardized the numerical output variables of the LIWC 2022 tool by adjusting all variables to have a zero mean and unit variance. This standardization is necessary, as some ML techniques tend to estimate larger effects for variables on a larger scale if the data is not standardized.

Training

We trained both ML and DL models to investigate the extent to which linguistic characteristics can contribute to the detection of SMI content. To conclude, we determined which technique performed best on the detection task. We formulated the detection of SMI content as a binary classification task to differentiate between SMI tweets and non-manipulative tweets. As techniques, we (a) built a panel of

models based on the standardized LIWC 2022 output variables, and (b) built a panel of models leveraging recent advancements in NLP that incorporate raw texts, based on pre-trained Transformer models (Vaswani et al., 2017; Devlin et al., 2019). We refer to (a) as ML-based classifiers and (b) as DL classifiers.

For model panel (a), we applied logistic regression (LOG), support vector machine (SVM), random forest (RF) and gradient boosting (GB) as ML techniques. More comprehensive reviews, providing detailed discussions on these ML techniques for financial fraud detection, are provided by Ngai et al. (2011) and West & Bhattacharyya et al. (2016). We utilized the scikit-learn implementation of these algorithms (Pedregosa et al., 2011) and used the XGBoost implementation for GB (Chen & Guestrin, 2016).

For model panel (b), we made use of DL-based Transformer models (Vaswani et al., 2017), which are pre-trained on general text corpora and then fine-tuned for domain-specific tasks, such as text classification (Devlin et al., 2019). Recent studies highlight the benefits of Transformer-based models compared to lexicon-based methods, such as LIWC 2022, including the domain of financial text classification (Mishev et al., 2020). Consequently, we decided to examine these models and compare them with the model panel (a). Specifically, we used DistilBERT (Sanh et al., 2020) and RoBERTa (Liu et al., 2019) based on the Transformers library (Wolf et al., 2020).

We trained all ML and DL techniques, evaluated their performance as described below, and tuned important hyperparameters using a grid search approach to optimize the F1 score. Table 4 reports the tuned parameters for each algorithm.

Panel	Algorithm	Parameter	Description	Grid	Best
Model panel (a)	LOG	Solver	Optimization problem algorithm.	lbfgs, liblinear	liblinear
		Penalty	Norm of the penalty.	l1, l2, elasticnet	l1
		C	Inverse of regularization strength.	10^{-x} for $x\epsilon[0,5]$	10^{-1}
	SVM	Kernel	Kernel type to be used in the algorithm.	rbf, sigmoid, linear	rbf
		C	Inverse of regularization strength.	10^{-x} for $x\epsilon[0,5]$	10^{-1}
	RF	Estimators	The number of trees in the forest.	[10, 20, 30,..,200]	50
		Criterion	Function to measure the quality of a split.	gini, entropy, log loss	gini
		Max depth	The maximum depth of the tree.	[2,3,4,...,30]	6
		Min samples split	Min. no. samples to split internal node.	[2,3,4,...,30]	2
		Min samples leaf	Min. no. samples required for leaf node.	[1,2,3,...,30]	1
	GB	Estimators	The number of trees in the forest.	[10, 20, 30,..,200]	60
		Learning rate	Boosting learning rate	x^{-1} for $x\epsilon[1,10]$	5^{-1}
		Max depth	The maximum depth of the tree.	[2,3,4,...,30]	5
		Min child weight	Min. sum of weight needed in child	[1,2,3,4,...,30]	1
		Subsample	Frac. of random samples for each tree	x^{-1} for $x\epsilon[1,10]$	1^{-1}
Model panel (b)	DistilBERT	Epochs	Number of epochs.	[1,2,3,...20]	6
		Learning rate	Learning rate.	10^{-x} for $x\ \epsilon\ [0,5]$	10^{-4}
		Batch size	Batch size.	4, 8, 16, 32	16
	RoBERTa	Epochs	Number of epochs.	[1,2,3,...20]	5
		Learning rate	Learning rate.	10^{-x} for $x\ \epsilon\ [0,5]$	10^{-4}
		Batch size	Batch size.	4, 8, 16, 32	16

Table 4. Hyperparameter tuning grids.

Evaluation

To ensure robust evaluation given the dataset size, we employed K-fold stratified cross validation (Kohavi, 1995). This approach reduces the effect random train-test splits have on reported evaluation metrics (Cawley & Talbot, 2010). We chose K=5, creating five equally sized folds, where each fold serves as a hold-out set, while the remaining four are utilized for model training. For performance evaluation, we computed common performance metrics including recall, precision, and the F1 score (Sokolova & Lapalme, 2009). For our final scores, we averaged the five different scores on each respective hold-out set. To compare the performance of model panel (a) and model panel (b), we used McNemar's test (Everitt, 1977), a statistical method used to compare the performance differences between two classifiers, which also reports the direction in which one classifier outperforms the other.

Results

Descriptive analysis

Table 5 presents a comprehensive summary of the descriptive statistics from the LIWC 2022 output variables, along with the results of a paired t-test conducted to test the equality of means between the variables for both SMI tweets and control tweets. We observed differences in the mean values for most of the features. However, only 49 out of the 118 variables (42%) had significant differences at the 5% level. Nevertheless, these findings suggest that linguistic characteristics may be informative in detecting SMI content. The LIWC 2022 categorizes variables into summary, linguistic dimensions, psychological processes, and dictionary variables, which we utilize to structure the presentation of the results.

The summary variables group offers a general understanding of the linguistic patterns observed in both SMI tweets and control tweets. We found that SMI tweets consist of significantly fewer words ($p<0.01$) compared to control tweets and significantly fewer words per sentence ($p<0.01$). SMI tweets also had shorter words, as the percentage of words with 7 letters or longer is significantly lower ($p<0.01$). Furthermore, we found significantly lower analytical thinking (Pennebaker et al., 2014) ($p<0.01$), which represents a metric of logical and formal thinking. This finding suggests that SMI tweets exhibit a reduced emphasis on structured, systematic thinking when compared to control tweets. Moreover, the clout metric (Kacewicz et al., 2014), which captures language indicative of leadership and status, displayed significantly higher values in SMI tweets compared to control tweets ($p<0.05$). This implies that SMIs utilize a greater sense of authority, confidence, or social standing in their online communications.

The linguistic dimensions section focuses on specific structural and grammatical usage features such as pronouns, articles, and prepositions. We found that the first personal singular, first personal plural, second personal, and impersonal pronouns are used significantly more frequently in SMI tweets compared to the control tweets ($p<0.01$). Additionally, we observed that fewer numbers are used in SMI tweets ($p<0.01$) and more prepositions are employed ($p<0.01$). Furthermore, we identified a higher usage of auxiliary verbs and adverbs in SMI tweets ($p<0.05$). Negations also occurred more frequently in SMI tweets ($p<0.01$).

The psychological processes section explores the cognitive and emotional aspects of the text in both SMI and control tweets, allowing for a better understanding of the psychological states expressed through language. This section includes variables related to cognitive mechanisms (e.g., causation, insight), affective processes (e.g., positive and negative emotions), and social processes. We found that the affiliation variable for SMI tweets is significantly higher compared to the control tweets ($p<0.01$). With respect to affect, negative tone ($p<0.01$), positive emotion ($p<0.05$) and negative emotion ($p<0.01$) appeared to be higher compared to control tweets. This finding indicates that SMI tweets tend to reflect a broader range of emotional expression. With respect to social processes, we found that more social referents are used ($p<0.01$), implying a stronger focus on social relationships and interpersonal dynamics in the language of SMI content.

The dictionary section provides a detailed breakdown of the LIWC 2022 dictionary words, and their corresponding word counts or percentages. This category allows for an in-depth examination of the usage of specific words and phrases. We found significantly fewer occurrences of dictionary words for politics ($p<0.05$) and ethnicity ($p<0.05$), and significantly more dictionary words for technology ($p<0.05$) in SMI tweets.

		SMI tweets (n=564)		Control tweets (n=564)		t-test	
Category	**Variable**	**Mean**	**Std**	**Mean**	**Std**	**p-value**	**Statistic**
Summary Variables	Word count	13.88	10.79	19.75	11.58	0.00	-8.82
	Analytical thinking	69.10	31.88	80.72	26.63	0.00	-6.64
	Clout	45.16	28.92	40.79	24.07	0.01	2.75
	Authentic	37.86	40.26	35.27	36.87	0.26	1.13
	Emotional Tone	24.40	32.40	21.07	30.17	0.07	1.79
	Words per sentence	8.11	4.72	12.37	7.99	0.00	-10.92
	Perc. words > 7 letters	12.77	9.25	19.28	11.81	0.00	-10.31
	Dictionary words	67.12	18.76	67.95	15.00	0.41	-0.82
Linguistic Dimensions	Linguistic	47.51	21.47	41.69	19.28	0.00	4.79
	Total function words	27.93	20.16	25.12	16.87	0.01	2.54
	Total pronouns	7.78	9.93	4.30	6.52	0.00	6.97
	Personal pronouns	4.17	6.41	1.94	4.24	0.00	6.89
	1st person singular	1.81	4.33	0.84	2.61	0.00	4.54
	1st person plural	0.89	3.19	0.33	1.47	0.00	3.79
	2nd person	0.87	3.09	0.39	1.80	0.00	3.24
	3rd person singular	0.18	1.70	0.10	0.86	0.32	0.99
	3rd person plural	0.33	1.51	0.26	1.97	0.48	0.70
	Impersonal pronouns	3.61	7.33	2.36	4.47	0.00	3.47
	Determiners	7.48	9.48	6.65	7.35	0.10	1.65
	Articles	3.47	5.42	3.31	4.92	0.60	0.53
	Numbers	6.58	7.41	8.87	8.63	0.00	-4.78
	Prepositions	7.48	8.15	9.69	7.31	0.00	-4.78
	Auxiliary verbs	4.22	6.44	3.48	5.52	0.04	2.07
	Adverbs	3.25	6.09	2.49	4.70	0.02	2.32
	Conjunctions	1.87	3.63	1.85	3.28	0.93	0.09
	Negations	0.91	2.86	0.54	2.05	0.01	2.51
	Common verbs	10.83	10.91	8.65	9.43	0.00	3.58
	Common adjectives	5.63	8.27	5.75	7.33	0.80	-0.26
	Quantities	3.95	7.02	3.91	5.58	0.90	0.13
Psychological Processes	Drives	2.49	5.27	2.14	3.93	0.21	1.26
	Affiliation	1.42	3.89	0.81	2.45	0.00	3.15
	Achievement	0.57	2.22	0.58	1.81	0.92	-0.11
	Power	0.50	2.09	0.77	2.65	0.05	-1.93
	Cognition	4.85	7.54	5.07	6.78	0.61	-0.51
	All-or-none	0.86	3.21	0.54	2.18	0.05	1.96
	Cognitive processes	3.95	6.62	4.51	6.43	0.14	-1.47
	Insight	0.87	2.59	1.00	2.87	0.40	-0.84
	Causation	0.52	2.27	0.66	2.07	0.30	-1.03
	Discrepancy	0.83	2.85	0.73	2.69	0.55	0.60
	Tentative	0.89	2.54	0.91	2.55	0.91	-0.12
	Certitude	0.23	1.52	0.20	1.84	0.73	0.35
	Differentiation	1.20	3.23	1.29	2.82	0.63	-0.48
	Memory	0.07	1.12	0.05	0.51	0.77	0.29
	Affect	9.25	8.43	7.65	6.59	0.00	3.57
	Positive tone	3.30	6.78	2.71	5.49	0.11	1.61
	Negative tone	5.68	6.19	4.74	4.21	0.00	2.99
	Emotion	6.13	6.79	4.58	4.35	0.00	4.54
	Positive emotion	0.85	3.77	0.36	1.77	0.01	2.81
	Negative emotion	5.16	5.86	4.08	3.88	0.00	3.65
	Anxiety	0.04	0.52	0.04	0.66	0.89	-0.14
	Anger	0.01	0.24	0.03	0.39	0.47	-0.72
	Sadness	0.00	0.08	0.04	0.39	0.05	-1.94
	Swear words	0.18	1.47	0.11	0.80	0.30	1.04
	Social processes	5.22	7.42	4.03	6.62	0.00	2.84
	Social behavior	1.62	3.63	2.00	4.01	0.09	-1.69
	Prosocial behavior	0.30	1.61	0.43	1.99	0.24	-1.17
	Politeness	0.19	1.32	0.17	1.69	0.89	0.14
	Interpersonal conflict	0.02	0.34	0.06	0.49	0.17	-1.37
	Moralization	0.10	0.85	0.13	1.29	0.73	-0.34

Table 5. LIWC features differences.

		SMI tweets (n=564)		Control tweets (n=564)		t-test	
Category	**Variable**	**Mean**	**Std**	**Mean**	**Std**	**p-value**	**Statistic**
Psychological Processes	Communication	0.83	2.97	0.87	2.69	0.81	-0.24
	Social referents	3.59	6.32	1.87	4.20	0.00	5.39
	Family	0.48	2.97	0.07	0.74	0.00	3.15
	Friends	0.13	1.14	0.02	0.28	0.02	2.27
	Female references	0.26	2.84	0.06	0.55	0.11	1.59
	Male references	0.22	1.34	0.16	1.18	0.42	0.81
Expanded Dictionary	Culture	5.12	5.68	4.63	4.46	0.11	1.61
	Politics	0.02	0.29	0.12	0.99	0.01	-2.51
	Ethnicity	0.01	0.28	0.09	0.81	0.04	-2.11
	Technology	5.09	5.64	4.42	4.18	0.02	2.28
	Lifestyle	2.48	4.65	5.33	6.53	0.00	-8.46
	Leisure	0.48	2.19	0.51	1.79	0.80	-0.26
	Home	0.20	1.02	0.01	0.20	0.00	4.27
	Work	0.94	2.89	2.97	4.65	0.00	-8.80
	Money	1.44	3.66	3.55	5.24	0.00	-7.83
	Religion	0.00	0.00	0.03	0.45	0.08	-1.75
	Physical	0.73	3.91	0.94	3.01	0.30	-1.05
	Health	0.17	1.81	0.46	1.98	0.01	-2.53
	Illness	0.11	1.60	0.20	1.09	0.27	-1.10
	Wellness	0.04	0.58	0.00	0.11	0.21	1.26
	Mental health	0.00	0.00	0.04	0.84	0.32	-1.00
	Substances	0.01	0.16	0.01	0.16	1.00	0.00
	Sexual	0.03	0.61	0.08	0.84	0.22	-1.22
	Food	0.18	1.52	0.16	1.61	0.80	0.25
	Death	0.03	0.37	0.02	0.23	0.56	0.58
	Need	0.12	0.87	0.11	0.91	0.87	0.16
	Want	0.17	1.56	0.10	1.24	0.41	0.83
	Acquire	1.10	3.18	0.73	2.58	0.03	2.16
	Lack	0.00	0.00	0.00	0.00	-	-
	Fulfilled	0.06	0.63	0.07	0.63	0.78	-0.28
	Fatigue	0.00	0.00	0.02	0.53	0.32	-1.00
	Reward	0.33	1.79	0.47	1.77	0.18	-1.35
	Risk	0.22	1.43	0.56	1.94	0.00	-3.34
	Curiosity	0.09	0.78	0.28	1.74	0.02	-2.37
	Allure	6.35	8.97	5.04	6.86	0.01	2.75
	Perception	8.98	11.35	9.12	9.44	0.82	-0.22
	Attention	0.43	1.77	1.02	3.07	0.00	-3.99
	Motion	1.77	4.70	2.02	4.88	0.38	-0.89
	Space	5.79	8.78	5.13	6.64	0.15	1.44
	Visual	0.85	2.65	1.32	4.40	0.03	-2.18
	Auditory	0.14	1.63	0.07	0.64	0.34	0.95
	Feeling	0.48	2.91	0.23	1.37	0.07	1.81
	Time	4.15	7.27	4.85	6.18	0.08	-1.73
	Past focus	2.20	4.68	1.99	4.28	0.43	0.79
	Present focus	3.48	6.21	3.04	4.93	0.18	1.33
	Future focus	1.45	4.37	1.10	3.68	0.15	1.46
	Conversation	7.60	9.25	4.99	4.94	0.00	5.90
	Netspeak	7.22	8.96	4.78	4.62	0.00	5.75
	Assent	0.17	2.26	0.06	0.69	0.28	1.09
	Nonfluencies	0.16	1.31	0.09	0.79	0.25	1.14
	Fillers	0.08	1.01	0.07	0.97	0.90	0.13
	All punctuation	66.41	104.70	51.69	30.83	0.00	3.20
	Period	11.07	12.08	9.06	8.35	0.00	3.25
	Comma	0.58	2.23	1.38	3.22	0.00	-4.85
	Question mark	16.23	95.94	3.40	17.54	0.00	3.12
	Exclamation points	1.33	7.24	1.16	5.29	0.65	0.46
	Apostrophes	1.49	4.33	1.24	3.37	0.29	1.05
	Other punctuation	35.71	24.61	35.44	22.61	0.85	0.19
	Emoji	1.59	10.64	0.47	3.81	0.02	2.35

Table 5. LIWC features differences (continued).

For the lifestyle category, the variable home was significantly higher ($p<0.01$), while the variables work ($p<0.01$) and money ($p<0.01$) were significantly lower in SMI tweets compared to control tweets. SMI tweets had significantly more references to words from the dictionary in terms of acquiring ($p<0.05$). For the motives category, SMI tweets had lower values in the risk variable ($p<0.01$) and the curiosity variable ($p<0.05$), while the allure variable was higher ($p<0.05$). We observed significantly higher values of netspeak ($p<0.01$) occurring in the SMI tweets. Moreover, the all punctuation variable revealed that SMI tweets contain significantly more punctuation ($p<0.01$), for example periods ($p<0.01$), or emojis ($p<0.05$).

Evaluation and analysis of LIWC-based classifiers

Table 6 displays the evaluation results of the 5-fold stratified cross-validation for the trained classifiers based on LIWC features from model panel (a). It reports the minimum, maximum, standard deviation and mean score across all folds. Out of these models, Random Forest (RF) and Gradient Boosting (GB) achieve the highest metric values. GB is the most accurate model, with an average precision of 73.4%, a recall of 71.5%, and an F1 score of 72.1%. In our balanced sample, a random classifier would be expected to achieve an F1 score of around 50%, as the classes are evenly distributed. The performance of the GB model, as indicated by the F1 score of 72.1%, is substantially higher than that of a random classifier. This result highlights the usefulness of linguistic features derived from LIWC for detecting SMI content.

Algorithm	Features	Metric	Min fold	Max fold	Mean	Std
LOG	LIWC	Precision	0.699	0.734	0.716	0.014
		Recall	0.661	0.760	0.704	0.037
		F1	0.684	0.744	0.709	0.022
SVM	LIWC	Precision	0.646	0.830	0.745	0.063
		Recall	0.637	0.759	0.698	0.047
		F1	0.661	0.742	0.717	0.029
RF	LIWC	Precision	0.646	0.839	0.741	0.064
		Recall	0.667	0.767	0.704	0.035
		F1	0.672	0.770	0.720	0.033
GB	LIWC	Precision	0.619	0.830	0.734	0.069
		Recall	0.664	0.776	0.715	0.051
		F1	0.645	0.773	0.721	0.043

Table 6. Results of 5-fold cross-validation for model panel (a).

In order to identify the key linguistic features valuable for detecting SMI content, we utilized the XGBoost package (Chen and Guestrin, 2016). This package allowed us to obtain feature importance scores for the 118 LIWC linguistic features by counting the number of times a feature is used to split the data across all trees. Table 7 presents the top 20 most important features for detecting SMI content, along with their average scores across all folds, as determined by XGBoost. Notably, if each of the 118 features contributed equally, the expected feature importance score would be 0.0085 (= 1/118). We found that the Clout variable has the highest feature importance. From the summary variables, we also found word count, words per sentence, percentage words > 7 letters and analytical thinking among the top 20 most important features. Furthermore, various variables from the linguistic dimensions category served as important features, such as total pronouns and personal pronouns. From the psychological processes' variables, only the social referents variable was among the most important ones. Lastly, for the dictionary category, we found variables such as conversation, technology and netspeak to be important.

When comparing the feature importance results to the earlier descriptive analysis, we observed a remarkable alignment between the two sets of findings. Many of the top 20 most important features identified through the XGBoost method were also showing significant differences between SMI and control tweets in the descriptive analysis. This consistency supports the validity of our approach and underlines the argument for the relevance of these linguistic features in detecting SMI content.

Feature	Importance	Description / Most frequently used exemplars
Clout	0.043	Language of leadership, status
Word count	0.042	Total word count
Total pronouns	0.036	I, you, that, it
Words per sentence	0.031	Average words per sentence
Personal pronouns	0.029	I, you, my, me
Conversation	0.029	yeah, oh, yes, okay
Percentage words > 7 letters	0.026	Percent words 7 letters or longer
Analytical Thinking	0.023	Metric of logical, formal thinking
Technology	0.023	car, phone, comput*, email*
Question Mark	0.022	Percentage of question mark symbols
Social referents	0.019	you, we, he, she
Linguistic	0.019	Summary variable linguistics
Home	0.017	home, house, room, bed
Perception	0.015	in, out, up, there
Netspeak	0.015	:), u, lol, haha*
Period	0.015	Percentage of period symbols
Emotion	0.015	good, love, happy, hope
Other punctuation	0.014	Other non-standard punctuation
Space	0.014	in, out, up, there
Conjunctions	0.014	and, but, so, as

Table 7. Most 20 important features (GB) and description from Boyd et al. (2022).

Evaluation of deep-learning-based models and comparison

In accordance with our research methodology, we trained the DL-based transformer models DistilBERT and RoBERTa from model panel (b) on our dataset to compare their performance to models that rely on LIWC features. We direct the readers to Table 4 for the optimal hyperparameter configurations of these models. Table 8 presents the results of the 5-fold stratified cross-validation for the trained and optimized models. Both DistilBERT and RoBERTa, which are based on raw text features, achieved higher metric values compared to the LIWC-based models. RoBERTa achieved the best performance with an average precision of 82.7%, a recall of 89.0%, and an F1 score of 85.8%.

Algorithm	Features	Metric	Min fold	Max fold	Mean	Std
DistilBERT	Raw text	Precision	0.752	0.883	0.749	0.056
		Recall	0.805	0.947	0.931	0.059
		F1	0.778	0.914	0.829	0.056
RoBERTa	Raw text	Precision	0.714	0.889	0.827	0.037
		Recall	0.866	0.956	0.890	0.038
		F1	0.808	0.864	0.858	0.021

Table 8. Results of 5-fold cross-validation for model panel (b).

To compare the performance of model panel (a) and model panel (b), we employed McNemar's test (Everitt, 1977). Our analysis revealed that the raw text-based models DistilBERT and RoBERTa significantly outperform the simpler models that rely on LIWC features (Table 9). This comparison between model panel (a) and (b) highlights the potential advantages of utilizing DL-based Transformer models in the task of detecting SMI content. While the raw text-based models DistilBERT and RoBERTa exhibit superior performance, it is worth noting that we cannot derive meaningful feature importance information like GB

classifiers. DL-based Transformer models, such as DistilBERT and RoBERTa, do not inherently offer an interpretable framework for extracting feature importance (Jawahar et al., 2019).

Algorithm	**DistilBERT**	**Result**	**RoBERTa**	**Result**
LOG	0.00***	DistilBERT > LOG	0.00***	RoBERTa > LOG
SVM	0.00***	DistilBERT > SVM	0.00***	RoBERTa > SVM
RF	0.01**	DistilBERT > RF	0.00***	RoBERTa > RF
GB	0.01**	DistilBERT > GB	0.01**	RoBERTa > GB

Table 9. McNemar's Test Results on Performance (F1) for LIWC-based vs. Raw Text Classifiers (Note: *p < 0.1, **p < 0.05, ***p < 0.01).

In summary, we can confirm that content from SMIs has distinctive linguistic characteristics as identified in LIWC differences and by our ML-based classifiers. The results are consistent with IMT and IDT theories. The LIWC-based ML models in our panels were suitable to detect SMI content with the RF and GB model providing the best results. Most important features derived from our GB algorithm were clout and word count. However, both raw text DL models DistilBERT and RoBERTa significantly outperformed the LIWC-based models, which therefore makes them most suitable for application in fraud detection systems.

Discussion

In addressing **RQ1,** as anticipated by the IMT and IDT theories, our analysis revealed significant differences in the linguistic characteristics used by SMIs compared to non-manipulative users. SMIs appeared to use significantly fewer words and fewer words per sentence, contradicting research on financial fake news and liars (e.g., Clarke et al., 2020; Zhou & Zhang, 2008). Usually, deceivers tend to use more words and a higher number of words per sentence to appear more persuasive and credible (Lausen et al., 2020). One possible explanation for our finding is the higher usage of emojis in SMI tweets, which can substitute textual expressions (Tang & Hew, 2019) and have been found in other financial subcommunities (Agrawal et al., 2022). Additionally, the increased use of netspeak found in SMI content may result in fewer words due to frequent use of abbreviations (Crystal, 2001). A high usage of netspeak was also found by Agrawal et al. (2022), who reported higher levels of informal language in certain financial subcommunities.

Our findings showed lower analytical thinking in SMI tweets compared to control tweets, suggesting a more informal and personal language in SMI tweets contrary to formal and logical (Kogan et al., 2021). We found higher levels of clout in SMI content, a metric indicating writing with confidence and leadership style. This finding is aligned with research on entrepreneur influencers that found high values of clout, utilized to convey others (Crittenden & Crittenden, 2022). SMIs used fewer references to money and numbers, similar to findings in other deceiver contexts, such as fake news (Clarke et al., 2020) and deceptive CEOs (Larcker & Zakolyukina, 2012), who use fewer words associated with value. While we identified the usage of pronouns as a significant linguistic feature in SMI content, our results contradict previous research suggesting that deceivers use fewer first-person singular pronouns and more second- and third-person pronouns (Hancock et al., 2007). Lastly, SMI tweets appeared to reflect a higher fraction of both positive and negative emotional expressions, whereas previous research found a higher fraction of positive emotional expressions and a lower fraction of negative emotional expressions in deceiver content (Siering et al., 2021).

Regarding **RQ2,** we implemented four well-known ML-based algorithms (LOG, SVM, RF, GB) utilizing LIWC features. The GB classifier achieved the highest average F1 score of 72.1% across a five-fold cross validation. Additionally, we implemented two well-known DL-based algorithms for text classification (DistilBERT & RoBERTa) that rely on the Transformer architecture. The RoBERTa model achieved the highest average F1 score of 85.8%. Overall, these findings suggest that linguistic styles are informative for detecting SMI content. Furthermore, our findings suggest that DL-based models significantly outperform traditional ML-based algorithms utilizing LIWC features for financial-domain text classification. This finding has been found on other text classification tasks such as sentiment analysis in the financial domain (Mishev et al., 2020), but was less emphasized in previous studies on stock manipulation fraud (e.g., Clarke et al. 2020).

This study contributes to the existing body of knowledge in the financial domain by providing insights into the linguistic characteristics of SMI content on social media. Our results are consistent with the IMT and IDT theories and highlight the unique linguistic characteristics of SMI content compared to non-manipulative content. Due to the increasing involvement of social media influencers in stock manipulation schemes, these insights can especially contribute to the design, development, and configuration of fraud detection systems monitoring social media data. Our results show that linguistic characteristics of social media content can be used as a foundation to accurately detect manipulative content on social media.

Our study bridges the gap between theory and practice by demonstrating the effectiveness of ML and DL techniques in detecting potential stock manipulation content. The systematic comparison of traditional ML algorithms with DL models revealed that DL models significantly outperform ML models in detecting SMI content. These findings are particularly valuable for regulators, financial institutions, and investors, who require accurate fraud detection systems to safeguard against manipulation schemes. Especially for DL models we observed highly accurate results, which demonstrates their usefulness as foundation for these systems. Such fraud detection systems can act as an early warning system for regulators, indicating which assets or social media users to investigate further or contribute to risk management frameworks of financial institutions. For individual investors, fraud detection systems monitoring social media may serve as a tool in investment decision-making.

While our study provides valuable insights into the linguistic patterns of SMI content and the effectiveness of supervised learning techniques in detecting SMI content, it is important to acknowledge the limitations and potential biases that may have affected our findings. First, our study utilizes a balanced sample of SMI content and non-manipulative content, whereas in real-world scenarios, the distribution of such content is likely to be imbalanced. While the balanced sample is suitable for highlighting differences between content for academic purposes, this may affect the performance of the models when applied to imbalanced real-world data. Second, we acknowledge that our control tweets may include some noisy labels, as we cannot fully rule out the possibility that the non-manipulative users in our dataset may also have been investigated by regulators. However, we believe this issue is minimized, due to the precautions we have taken and described earlier. Alternatively, we evaluated the possibility to establish ground truth labels by labeling through domain experts, but have considered it problematic to make assumptions on intentions of social media users (supported by Siering et al. 2019). Therefore, we relied on the assessment of the SEC as a source for our ground truth. Additionally, we made efforts to exclude bots from our control tweet sample, but we cannot rule out the possibility that we included bot-generated tweets. Lastly, our study focuses on linguistic characteristics and does not consider other formal aspects of content presentation that could aid in the detection of stock manipulation schemes. The presence of images, videos, or other multimedia content in social media posts may also prove useful in detecting stock manipulation-related content.

We propose several paths for future research that could further investigate the phenomenon of SMIs. Future research could explore the role of network structures, social connections, and interaction patterns among SMIs and their followers. Investigating the dynamics of information propagation and the influence of SMIs within their social networks could aid in the development of more effective detection and prevention strategies. Our study focused on the analysis of social media content from a single platform. Future research could investigate stock manipulation schemes across different social media platforms (such as Telegram, Discord or Reddit), examining how SMIs leverage multiple platforms to propagate their messages and manipulate stock prices. This would help develop a more holistic understanding of stock manipulation schemes and improve the effectiveness of detection mechanisms across various platforms. Future research could employ a longitudinal design to examine the evolution of stock manipulation schemes over time, investigating how the strategies employed by SMIs adapt to technological advancements and regulatory interventions.

Conclusion

By analyzing the social media content of charged SMIs from the SEC's $100 million fraud case in 2022, our study not only offers important insights into the linguistic characteristics employed by SMIs, but also underscores the potential of ML and especially DL models for SMI content detection. Our analysis revealed several linguistic characteristics commonly found in social media content of SMIs. Our study also showcased the effectiveness of recent advances in NLP based on DL to detect social media-initiated stock manipulation schemes. The results of this study have implications for various stakeholders in the financial

community, including investors, regulators and financial institutions, to better safeguard investments against stock manipulation schemes orchestrated through social media channels.

Acknowledgments

This work is part of the research project AFFIN funded by the German Federal Ministry of Education and Research (Grant no.: 01IS21045B).

References

Agrawal, P., Buz, T., & de Melo, G. (2022). WallStreetBets Beyond GameStop, YOLOs, and the Moon: The Unique Traits of Reddit's Finance Communities. *Proceedings of the Americas Conference on Information Systems (AMCIS)*

Allen, F., & Gale, D. (1992). Stock-Price Manipulation. *Review of Financial Studies*, *5*(3), 503–529. https://doi.org/10.1093/rfs/5.3.503

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication, 13*(1), 210-230.

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). The development and psychometric properties of LIWC-22. Austin, TX: University of Texas at Austin. https://www.liwc.app

Cawley, G. C., & Talbot, N. L. C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, *11*(70), 2079–2107.

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Clarke, J., Chen, H., Du, D., & Hu, Y. (2020). Fake News, Investor Attention, and Market Reaction. *Information Systems Research*, *32*(1), 35–52. https://doi.org/10.1287/isre.2019.0910

Craja, P., Kim, A., & Lessmann, S. (2020). Deep learning for detecting financial statement fraud. *Decision Support Systems*, *139*, 113421. https://doi.org/10.1016/j.dss.2020.113421

Cresci, S., Lillo, F., Regoli, D., Tardelli, S., & Tesconi, M. (2019). Cashtag Piggybacking: Uncovering Spam and Bot Activity in Stock Microblogs on Twitter. *ACM Transactions on the Web*, *13*(2), 1–27. http://arxiv.org/pdf/1804.04406

Crittenden, V., & Crittenden, W. (2022). The power of language to influence people: Mary Kay Ash the entrepreneur. *Journal of Research in Marketing and Entrepreneurship*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/JRME-05-2022-0065

Crystal, D. (2001). *Language and the Internet*. Cambridge University Press.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, *1*, 4171–4186. https://doi.org/10.18653/v1/N19-1423

Dong, W., Liao, S. Y., & Zhang, Z. (2018). Leveraging Financial Social Media Data for Corporate Fraud Detection. *Journal of Management Information Systems*, *35*(2), 461–487. https://doi.org/10.1080/07421222.2018.1451954

Dupuis, D., Smith, D. F., & Gleason, K. C. (2021). Old frauds with a new sauce: digital assets and space transition. *Journal of Financial Crime*, *30*(1), 205–220. https://doi.org/10.1108/jfc-11-2021-0242

Everitt, B. S. (1977). *The Analysis of Contingency Tables*. Springer eBooks. https://doi.org/10.1007/978-1-4899-2927-3

Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996). Knowledge discovery and data mining: towards a unifying framework. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 82–88. https://www.aaai.org/Papers/KDD/1996/KDD96-014.pdf

Felixson, K., & Pelli, A. (1999). Day end returns—stock price manipulation. *Journal of Multinational Financial Management*, *9*(2), 95–127. https://doi.org/10.1016/s1042-444x(98)00052-8

Guan, S. S. (2023). The Rise of the Finfluencer. New York University *Journal of Law and Business, Forthcoming, Santa Clara Univ. Legal Studies Research Paper* Forthcoming.

Haase, F., Rath, O., Kurka, M., & Schoder, D. (2023). Finfluencers: Opinion Makers or Opinion Followers? *ECIS 2023 Research Papers*. 432.

Hancock, J. T., Curry, L. E., Goorha, S., & Woodworth, M. H. (2007). On Lying and Being Lied To: A Linguistic Analysis of Deception in Computer-Mediated Communication. *Discourse Processes*, *45*(1), 1–23. https://doi.org/10.1080/01638530701739181

Hu, T., & Tripathi, A. K. (2016). Impact of Social Media and News Media on Financial Markets. *Thirty Seventh International Conference on Information Systems, Dublin*. https://papers.ssrn.com/abstract=2796906

Jawahar, G., Sagot, B., & Seddah, D. (2019). What Does BERT Learn about the Structure of Language? *HAL (Le Centre Pour La Communication Scientifique Directe)*. https://doi.org/10.18653/v1/p19-1356

Kacewicz, E., Pennebaker, J. W., Davis, M. M., Jeon, M., & Graesser, A. C. (2014). Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology*, *33*(2), 125–143. https://doi.org/10.1177/0261927x13502654

Kogan, S., Moskowitz, T. J., & Niessner, M. (2021). Social Media and Financial News Manipulation. *Review of Finance*. https://doi.org/10.1093/rof/rfac058

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, *2*, 1137–1143.

Kurgan, L., & Musilek, P. (2006). A survey of Knowledge Discovery and Data Mining process models. *Knowledge Engineering Review*, *21*(1), 1–24. https://doi.org/10.1017/s0269888906000737

Larcker, D. F., & Zakolyukina, A. A. (2012). Detecting Deceptive Discussions in Conference Calls. *Journal of Accounting Research*, *50*(2), 495–540. https://doi.org/10.1111/j.1475-679x.2012.00450.x

Lausen, J., Clapham, B., Siering, M., & Gomber, P. (2020). Who Is the Next "Wolf of Wall Street"? Detection of Financial Intermediary Misconduct. *Journal of the Association for Information Systems*, *21*(5), 1153–1190. https://doi.org/10.17705/1jais.00633

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach* (arXiv:1907.11692). arXiv. https://doi.org/10.48550/arXiv.1907.11692

Louwerse, M. M., Lin, D., Drescher, A., & Semin, G. R. (2010). Linguistic cues predict fraudulent events in a corporate social network. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *32*(32). https://escholarship.org/content/qt6gp464xg/qt6gp464xg.pdf?t=op2lcj

Mancini, A. L., Desiderio, A., Di Clemente, R., & Cimini, G. (2022). Self-induced consensus of Reddit users to characterise the GameStop short squeeze. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-17925-2

McCabe, C. (2022, May 13). *Twitter Stock Falls After Elon Musk Says Deal Is 'On Hold.'* The Wall Street Journal. https://www.wsj.com/articles/twitter-stock-tumbles-premarket-after-elon-musk-says-deal-is-on-hold-11652445126 (accessed on May 3rd 2023)

McCornack, S. A. (1992). Information manipulation theory. *Communication Monographs*, *59*(1), 1–16. https://doi.org/10.1080/03637759209376245

Mirtaheri, M., Abu-El-Haija, S., Morstatter, F., Ver Steeg, G., & Galstyan, A. (2021). Identifying and analyzing cryptocurrency manipulations in social media. *IEEE Transactions on Computational Social Systems, 8*(3), 607-617.

Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L. T., & Trajanov, D. (2020). Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers. *IEEE Access*, *8*, 131662–131682. https://doi.org/10.1109/ACCESS.2020.3009626

Ngai, E. W., Hu, Y., Wong, Y. J., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, *50*(3), 559–569. https://doi.org/10.1016/j.dss.2010.08.006

Öğüt, H., Doğanay, M., & Aktaş, R. (2009). Detecting stock-price manipulation in an emerging market: The case of Turkey. *Expert Systems With Applications*, *36*(9), 11944–11949. https://doi.org/10.1016/j.eswa.2009.03.065

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Dubourg, V. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.

Pflücke, F. (2022). Regulating Finfluencers. *Journal of European Consumer and Market Law*, *11(6)*, 212 ff. Available at SSRN: https://ssrn.com/abstract=4291905

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. (2014). When Small Words Foretell Academic Success: The Case of College Admissions Essays. *PLOS ONE*, *9*(12), e115844. https://doi.org/10.1371/journal.pone.0115844

Robertson, A. (2021, March 5). *John McAfee charged with securities fraud for 'pump and dump' cryptocurrency scheme*. The Verge. https://www.theverge.com/2021/3/5/22315494/john-mcafee-fraud-securities-scheme-charges-cryptocurrency (accessed on May 3rd 2023)

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2020). *DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter* (arXiv:1910.01108). arXiv. https://doi.org/10.48550/arXiv.1910.01108

SEC. (2015, July 29). *Investor Alert: Fraudulent Stock Promotions*. U.S. Securities and Exchange Commission. https://www.investor.gov/introduction-investing/general-resources/news-alerts/alerts-bulletins/investor-alerts/investor-35 (accessed on May 3rd 2023)

SEC. (2022, December 14). *SEC Charges Eight Social Media Influencers in $100 Million Stock Manipulation Scheme Promoted on Discord and Twitter* [Press release]. U.S. Securities and Exchange Commission. https://www.sec.gov/news/press-release/2022-221 (accessed on May 3rd 2023)

Siering, M. (2019). The economics of stock touting during Internet-based pump and dump campaigns. *Information Systems Journal*, *29*(2), 456–483. https://doi.org/10.1111/isj.12216

Siering, M., Clapham, B., Engel, O., & Gomber, P. (2017). A Taxonomy of Financial Market Manipulations: Establishing Trust and Market Integrity in the Financialized Economy through Automated Fraud Detection. *Journal of Information Technology*, *32*(3), 251–269. https://doi.org/10.1057/s41265-016-0029-z

Siering, M., Koch, J., & Deokar, A. V. (2016). Detecting Fraudulent Behavior on Crowdfunding Platforms: The Role of Linguistic and Content-Based Cues in Static and Dynamic Contexts. *Journal of Management Information Systems*, *33*(2), 421–455. https://doi.org/10.1080/07421222.2016.1205930

Siering, M., Muntermann, J., & Grčar, M. (2021). Design Principles for Robust Fraud Detection: The Case of Stock Market Manipulations. *Journal of the Association for Information Systems*, *22*(1), 156–178. https://doi.org/10.17705/1jais.00657

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, *45*, 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Strauss, N., & Smith, C. (2019). Buying on rumors: how financial news flows affect the share price of Tesla. *Corporate Communications: An International Journal*, *24*(4), 593–607. https://doi.org/10.1108/ccij-09-2018-0091

Tang, Y., & Hew, K. F. (2019). Emoticon, Emoji, and Sticker Use in Computer-Mediated Communication: A Review of Theories and Research Findings. *International Journal of Communication*, *13*, 27. http://hub.hku.hk/bitstream/10722/275809/1/Content.pdf

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In *The social psychology of intergroup relations* (pp. 33–37). Brooks/Cole.

Tardelli, S., Avvenuti, M., Tesconi, M., & Cresci, S. (2022). Detecting inorganic financial campaigns on Twitter. *Information Systems*, *103*, 101769. https://doi.org/10.1016/j.is.2021.101769

Throckmorton, C. S., Mayew, W. J., Venkatachalam, M., & Collins, L. M. (2015). Financial fraud detection using vocal, linguistic and financial cues. *Decision Support Systems*, *74*, 78–87. https://doi.org/10.1016/j.dss.2015.04.006

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All you Need. *Advances in Neural Information Processing Systems*, *30*. https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

West, J., & Bhattacharya, M. (2016). Some Experimental Issues in Financial Fraud Mining. *Procedia Computer Science*, *80*, 1734–1744. https://doi.org/10.1016/j.procs.2016.05.515

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., … Rush, A. (2020). Transformers: State-of-the-Art Natural Language Processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45. https://doi.org/10.18653/v1/2020.emnlp-demos.6

Yaffe-Bellany, D. (2022, May 27). How Influencers Hype Crypto, Without Disclosing Their Financial Ties. *The New York Times*. https://www.nytimes.com/2022/05/27/technology/crypto-influencers.html (accessed on May 3rd 2023)

Zhou, L., & Zhang, D. (2008). Following linguistic footprints: Automatic deception detection in online communication. *Communications of the ACM*, *51*(9), 119–122. https://doi.org/10.1145/1378727.1389972