Journal of the Association for Information Systems (2020) **21**(5), 1153-1190
doi: 10.17705/1jais.00633

RESEARCH ARTICLE

ISSN 1536-9323

Who Is the Next "Wolf of Wall Street"? Detection of Financial Intermediary Misconduct

Jens Lausen[1], Benjamin Clapham[2], Michael Siering[3], Peter Gomber[4]
[1]Goethe University Frankfurt, Germany, lausen@wiwi.uni-frankfurt.de
[2]Goethe University Frankfurt, Germany, clapham@wiwi.uni-frankfurt.de
[3]Goethe University Frankfurt, Germany, siering@wiwi.uni-frankfurt.de
[4]Goethe University Frankfurt, Germany, gomber@wiwi.uni-frankfurt.de

Abstract

Financial intermediaries are essential for investors' participation in financial markets. Because of their position within the financial system, intermediaries who commit misconduct not only harm investors but also undermine trust in the financial system, which ultimately has a significant negative impact on the economy as a whole. Building upon information manipulation theory and warranting theory and making use of self-disclosed data with different levels of external verification, we propose different classifiers to automatically detect financial intermediary misconduct. In particular, we focus on self-disclosed information by financial intermediaries on the business network LinkedIn. We match user profiles with regulator-disclosed information and use these data for classifier training and evaluation. We find that self-disclosed information provides valuable input for detecting financial intermediary misconduct. In terms of external verification, our classifiers achieve the best predictive performance when also taking regulator-confirmed information into account. These results are supported by an economic evaluation. Our findings are highly relevant for both investors and regulators seeking to identify financial intermediary misconduct and thus contribute to the societal challenge of building and ensuring trust in the financial system.

Keywords: Financial Misconduct, Fraud Detection, Financial Intermediaries, Self-Disclosed Information, Information Verification, Machine Learning, Predictive Supervision

Gautam Pant was the accepting senior editor. This research article was submitted on March 14, 2018 and went through four revisions.

1 Introduction

Financial intermediaries are essential for investors because they exhibit a strong influence not only on financial performance but also on wealth and life planning. Intermediaries such as investment advisors screen the market and suggest investment opportunities, while other actors, such as brokers, provide market access for trading financial instruments, enabling investors to participate in financial markets (Allen & Santomero, 1997). With increased usage of the internet and electronic communication, personal interactions between investors and financial intermediaries have significantly diminished, potentially impeding the trust building process (Ba & Pavlou, 2002; Castells, 2010). In addition, the global financial crisis and widely publicized financial market manipulations have challenged investors' confidence in financial intermediaries and the financial system as a whole (Palazzo & Rethel, 2008). In fact, the issue of misconduct and fraud by intermediaries was raised by the film industry in the 2013 movie *The Wolf of Wall Street*, which is based on the memoirs of the former stockbroker Jordan Belfort. Belfort defrauded 1,513 clients and was responsible for investor losses of approximately USD 200 million (Bloomberg, 2018). Fraud and misconduct reduce investors' willingness to rely on intermediaries for participation in financial markets and thus negatively affect market-based allocation decisions and the economy as a whole. Consequently, regulatory interventions and new instruments are needed to increase trust and ensure the proper functioning of the financial system. Information systems and analytics play a crucial role within this context because they can help identify misconduct and thus increase trust in financial markets by preventing the next *Wolf of Wall Street*.

Previous studies have proposed various approaches to identifying different kinds of financial market manipulations (Ngai et al., 2011). However, these studies have neglected the identification of financial intermediaries who are likely to commit misconduct. Currently, social media networks provide an important source of information regarding potential and ongoing business contacts and relationships and are thus increasingly relevant for selecting financial intermediaries (Bazarova & Choi, 2014; Krasnova et al., 2010). Therefore, many professionals, including financial intermediaries, make use of these networks for advertising and to establish contacts. Personal information disclosed within social media networks can be confirmed by external sources, leading to different levels of information reliability. However, self-disclosed information that is subjected to different levels of external verification, has not yet been considered as a means of identifying intermediary misconduct.

Based on *information manipulation theory* (McCornack, 1992), we argue that the information disclosure of intermediaries who are likely to commit misconduct differs from that of reliable market participants. Moreover, drawing on *warranting theory* (Walther et al., 2009), we propose that external verification of self-disclosed information provides additional value for the identification of misconduct. Following this rationale, we address the following research question: *Can self-disclosed information that is subjected to different levels of external verification be used to detect financial intermediaries who are likely to commit misconduct*?

In this paper, we identify different feature sets that enable investors and regulatory/supervisory authorities to distinguish financial intermediaries who have committed financial misconduct from reliable ones. We compose a comprehensive dataset of information that is self-disclosed by financial intermediaries on the professional social media network LinkedIn. Additionally, we extract information regarding misconduct from BrokerCheck, an open access database operated by the Financial Industry Regulatory Authority (FINRA),[1] and match this information with profiles on LinkedIn.

We evaluate different classification models that automatically detect financial intermediaries who are likely to commit misconduct by making use of several feature sets that include different levels of external verification. Furthermore, we examine the economic relevance of our classifiers by means of an economic evaluation based on payments of victim compensation and fines. We find that self-disclosed information provided by financial intermediaries is valuable for detecting financial intermediary misconduct. Specifically, classifiers that also take externally verified information into account achieve a high level of classification performance and their application offers considerable economic value to society.

Our study has important implications for research. Based on information manipulation theory as well as warranting theory, we outline the relevance of self-disclosed information and different levels of external verification of such information for the detection of intermediary misconduct. Confirming information manipulation theory, we show that there is a difference between the self-disclosed information of honest versus dishonest financial intermediaries that can be used to identify intermediary misconduct. In line with warranting theory, information that is externally verified significantly increases classification performance.

Moreover, our results are also highly relevant from a societal and economic point of view, as they enable building classifiers for the automated identification of financial intermediary misconduct. Such classifiers may be used by investors to screen intermediaries in advance, thereby reducing the likelihood of incurring losses through misconduct. An automated classification system is also helpful for regulatory authorities to establish fair and efficient markets. Regulatory/supervisory authorities have limited resources to oversee the large number of financial intermediaries executing an ever-increasing number of client transactions. Therefore, such models support authorities' engagement in predictive supervision by allowing them to allocate their resources more efficiently to identify and closely monitor intermediaries that are more likely to commit

[1] The website BrokerCheck by FINRA is available via https://brokercheck.finra.org/

misconduct. Using data analytics and machine learning techniques, we contribute to the societal and economic challenge of how to build or rebuild trust in financial markets via the identification of intermediaries who engage in misconduct.

The paper proceeds as follows. In Section 2, we provide background information on financial market misconduct and its detection. Then, we develop our research hypotheses based on the theories underlying our feature selection. Section 3 presents our dataset and research methodology, especially the different classification models applied to detect financial misconduct that use unverified and externally verified self-disclosed information. Subsequently, in Section 4, we present, evaluate, and discuss the results of our empirical study, and then conclude with Section 5.

2 Research Background and Hypotheses

2.1 Misconduct by Financial Intermediaries

Misconduct in financial markets directly harms investors and deteriorates market participants' trust in the financial system. Households and investors that are less willing to participate in financial markets often fail to achieve returns sufficient for retirement plans and other financial goals. The reluctance to use financial markets also increases companies' cost of capital because capital becomes scarce, which ultimately reduces economic growth. Consequently, building and preserving trust in financial markets represents a major societal challenge.

Financial misconduct is widely regarded as being both common and costly (Dyck, Morse, & Zingales, 2010). Retail investors suffer damages amounting to billions of dollars each year. The Council of Economic Advisors (2015) estimates that the aggregate annual cost of conflicted advice in individual retirement accounts (i.e., investment advice where conflicts of interest are present because of high commissions for intermediaries) amounts to USD 17 billion.

There are various types of misconduct committed by brokers or investment advisors. One type of misconduct is providing false and misleading information to customers. This type of misconduct can be characterized as a principal-agent problem, where the broker benefits at the expense of the client or the market (Cumming, Johan, & Li, 2011). One example is a broker or investment advisor breaching the "suitability rule,"[2] meaning that transactions or investments in securities are not in accordance with the client's investment profile. Further examples include brokers or investment advisors charging exaggerated fees or failing to obtain the best price for a client in securities transactions.

Other types of misconduct affecting financial markets include *front running*, *scalping*, and *churning*. Front running refers to brokers making use of their private information about a client's order by buying or selling a security in advance of the client's trade, allowing them to profit from a price movement that may be caused by the client's (potentially large) trade (Cataldo & Killough, 2003). Churning describes the excessive buying and selling of securities on a client's account without the consent of the client and disregarding the client's interests in order to generate higher commissions for the intermediary (Cumming & Johan, 2008). Scalping refers to the practice of investment advisors purchasing a security before recommending it to a client without disclosing the benefit that they may derive from a potentially higher price should the customer follow their recommendation (Hazen, 2010). A more detailed overview of financial market manipulations performed by intermediaries is presented by Siering et al. (2017).

Brokers and investment advisors in the US are subject to a comprehensive system of regulations. FINRA, the responsible competent authority, mandates the disclosure of material facts about every broker and investment advisor, including any allegations or instances of wrongdoing (Lazaro, 2014). Using public and nonpublic regulatory data provided by FINRA, several studies have already proposed approaches to detect misconduct based on past intermediary misconduct (Egan, Matvos, & Seru, 2019; Qureshi & Sokobin, 2015). However, McCann, Qin, and Yan (2017) point out that the publicly available information provided via FINRA's website BrokerCheck is not sufficient to identify brokers who are likely to commit misconduct and does not help investors protect themselves. Therefore, this paper presents different classifiers to detect misconduct based on self-disclosed information of financial intermediaries subjected to different levels of external verification.

2.2 Automated Detection of Misconduct and Fraud in Financial Markets

Previous research has shown that data mining techniques are useful and efficient for identifying fraudulent activities in financial markets because manual detection is time consuming, expensive, and impractical, given the large amount of data to be analyzed (West & Bhattacharya, 2016). Initial studies on fraud detection in financial markets mainly rely on logistic regressions (Lee, Ingram, & Howard, 1999;

[2] The suitability rule is reflected in FINRA Rule 2111

Persons, 1995) and neural networks (Fanning & Cogger, 1998). Later studies focus on different data mining techniques to detect fraud. Bolton and Hand (2002) as well as Ngai et al. (2011) provide a comprehensive overview of research on data mining techniques for automated fraud detection in financial markets.

Data mining techniques have been extensively applied to the detection of credit card fraud (e.g., Bhattacharyya et al., 2011) and accounting fraud (e.g., Wang, 2010). Based on a real-life dataset of credit card transactions, Bhattacharyya et al. (2011) were able to identify fraudulent transactions using random forests and support vector machines. With respect to accounting fraud, Kirkos, Spathis, and Manolopoulos (2007) compare different data mining techniques based on structured, quantitative variables to detect fraudulent financial statements. They show that decision trees, neural networks, and Bayesian belief networks can correctly classify (non)fraudulent financial statements. Beyond structured data, researchers have also applied text mining methods to analyze linguistic cues in unstructured parts of regulatory or financial disclosures. Humpherys et al. (2011) apply naive Bayes and decision trees, taking linguistic variables from the management discussion and analysis section into account to distinguish between fraudulent and nonfraudulent financial statements. Similarly, Glancy and Yadav (2011) propose a quantitative model using text mining methods to detect fraudulent financial statements. In addition to structured and textual data from financial statements, Dong, Liao, and Zhang (2018) focus on user-generated content from financial social media platforms and show that this type of unstructured data adds incremental value to the detection of corporate fraud.

While the application of data mining techniques to detect credit card fraud and accounting fraud has been analyzed in detail, research on automatic detection of securities fraud is scarce (Ngai et al., 2011). This is particularly true for misconduct committed by financial intermediaries, who are an essential part of every securities transaction. While previous research focuses on identifying single incidents of fraudulent behavior, for example by analyzing market data or financial statements, it disregards the detection of individual intermediaries who are likely to commit misconduct. Consequently, this paper aims to close this research gap by developing classification models to detect financial intermediaries who are likely to commit misconduct.

2.3 Theoretical Background

Millions of users routinely self-disclose personal information by participating in social media networks (Bazarova & Choi, 2014). Jourard (1971, p. 2) defines self-disclosure as "the act of revealing personal information to others." Users of social media networks primarily self-disclose information to attract attention (Hollenbaugh & Ferris, 2014) and to maintain and develop relationships (Krasnova et al., 2010). Financial intermediaries and other professionals also disclose profile information on business networks to reach or interact with potential customers.

For the purpose of automatically identifying financial intermediary misconduct, we base our feature selection on two important theoretical streams in the context of self-disclosure in social media networks and fraud detection. First, we rely on information manipulation theory (McCornack, 1992) to explain why individuals who commit misconduct communicate differently, compared to honest individuals. Second, warranting theory (Walther et al., 2009) provides the theoretical basis for explaining why the level of external verification can influence the utility of different feature sets.

According to information manipulation theory, deceivers violate four key communication principles (McCornack, 1992). First, deceivers exaggerate or understate the quantity of information in order to conceal or misrepresent information. Second, deceivers tend to alter the quality of information or simply lie to disguise facts. Third, deceivers mislead receivers by providing information that is out of context. Fourth, deceivers may purposely communicate information in an ambiguous manner in order to confuse the receiver. Information manipulation theory has been empirically tested in the context of financial reporting fraud. In this context, researchers demonstrated that writers of misleading financial statements actually use techniques posited in information manipulation theory to deceive (Glancy & Yadav, 2011; Humpherys et al., 2011). Also, previous research has shown that classifiers that account for feature sets based on information manipulation theory can identify fraudulent projects in the context of crowdfunding campaigns (Siering, Koch, & Deokar, 2016).

Even if deceivers try to make their profiles look similar to those of truth-tellers, as suggested by interpersonal deception theory (Buller & Burgoon, 1996), their fabricated profiles will necessarily be less detailed and precise than authentic ones. However, interpersonal deception theory does not fit our context because it builds on repeated communication exchanges between the sender and receiver (Buller & Burgoon, 1996) whereas the LinkedIn profiles analyzed in this study represent static information as a "monolog" of the intermediary. In conclusion, we assume that financial intermediaries' self-disclosed information on the business network LinkedIn differs between intermediaries who are likely to commit misconduct and those who are not likely to commit misconduct.

Warranting theory (Walther et al., 2009) proposes that individuals' self-disclosed information is more valuable if it cannot be easily manipulated by the individuals themselves. This theory postulates that information that is harder to manipulate is more plausible or trustworthy than information that is easier to manipulate. Self-disclosed information on social media network profiles is subject to different levels of external verification. Other social media network users can confirm information provided by the individual profiles. Furthermore, self-disclosed information can be counterchecked with data provided by known and reliable external sources, which provides an even stronger source of verification than that provided by possibly unknown third parties. In the case of financial intermediaries, publicly available data provided by the regulatory authority FINRA's website BrokerCheck can be used as a reliable external source that is difficult to manipulate.

2.4 Research Hypotheses

In order to answer our research question—*Can self-disclosed information that is subjected to different levels of external verification be used to detect financial intermediaries who are likely to commit misconduct?*—we develop a set of research hypotheses, each representing a different level of external verification. According to the theoretical foundations of information manipulation theory, honest and dishonest individuals communicate differently, producing anomalies that can be detected by classification mechanisms in order to identify misconduct. Therefore, if classifiers based on related features can detect misconduct better than pure chance, then the self-disclosed information of financial intermediaries can serve as a valuable source of information that can be used to detect intermediaries who engage in misconduct. Information manipulation theory states that deceivers tend to conceal or misrepresent information and thus communicate and provide information differently, as compared to honest individuals (McCornack, 1992). Consequently, we anticipate that unverified, self-disclosed information disclosed in business networks such as LinkedIn is valuable for the detection of financial intermediary misconduct and thus hypothesize:

H1: Self-disclosed information, which has not been verified by third parties, is valuable for the detection of financial intermediaries who are likely to commit misconduct.

According to warranting theory, self-disclosed personal information incorporates higher credibility if the individual cannot easily manipulate it (Walther et al., 2009). In social media networks, other users can confirm information disclosed by social media network participants. This confirmation may occur, for example, through other users' endorsements of self-disclosed information such as skills. As proposed by information manipulation theory, deceivers tend to provide dubious or even false information in order to mislead their counterparties (McCornack, 1992). The possibility of confirming self-disclosed information seems to be a useful means to distinguish between trustworthy financial intermediaries and those that commit misconduct. Thus, we anticipate that self-disclosed information verified by other users is more reliable than unverified information, and hypothesize:

H2: Classifiers accounting for both unverified self-disclosed information and self-disclosed information that is verified by other network users perform better than classifiers that take only unverified self-disclosed information into account.

Beyond social media network users, regulatory authorities can also verify intermediaries' self-disclosed information. Consequently, self-disclosed information can be counterchecked with data published by the regulator. As regulatory data represents a reliable and neutral source, regulatory confirmations represent an even stronger source of external verification than confirmations by potentially anonymous or unknown users of a social network. Thus, we hypothesize:

H3: Classifiers that additionally account for self-disclosed information verified by regulatory authorities perform better than classifiers based on self-disclosed and user-confirmed information only.

3 Research Methodology

3.1 Data Mining Process

In order to investigate our research hypotheses and to develop different classifiers to detect financial intermediaries who commit misconduct, we adapt the knowledge discovery from databases (KDD) process outlined by Fayyad, Piatetsky-Shapiro, and Smyth (1996). This process model, which is the most cited model in the field of data mining and knowledge discovery, is well suited for academic research settings and data mining tasks that require substantial data preprocessing (Kurgan & Musilek, 2006). We create the target dataset by means of data extraction from business profiles on LinkedIn, which we match to regulatory data from FINRA's website BrokerCheck. Then, we clean and preprocess the data and subsequently select appropriate data mining and machine learning techniques to evaluate the resulting classifiers both statistically and economically. The entire data mining process is depicted in Figure 1.

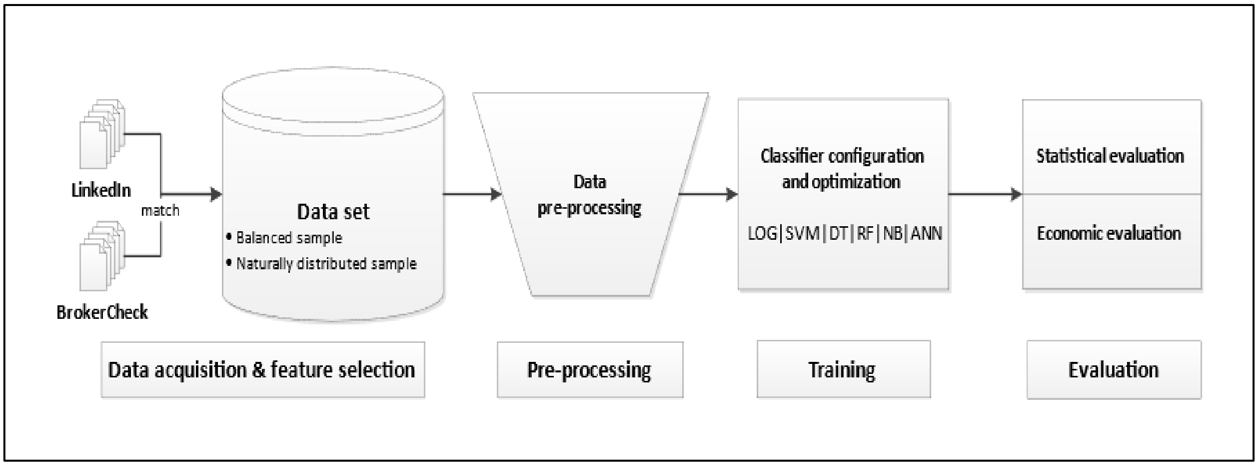

Figure 1. Data Mining Process

3.2 Data Acquisition

In order to train and evaluate different classifiers to detect financial intermediaries who are likely to commit misconduct, we used BrokerCheck as our starting point, which offers a complete record of all brokers registered in the US. We randomly drew two samples: a balanced sample for training the different classifiers and for providing an initial evaluation, as well as a naturally distributed sample for an additional evaluation based on real-world class distribution. This was necessary since we observed that, historically, only 6.83% of the intermediaries listed on BrokerCheck have actually engaged in misconduct.

The balanced sample was composed of 400 brokers with a history of misconduct and 400 brokers without a history of misconduct and was used for classifier training. We used random undersampling for the no-misconduct class to undersample the majority class at random until it has the same number of observations as the minority class (Chawla, 2009; Japkowicz, 2000). We used a balanced dataset to train the classifiers since unbalanced data for training often leads to poor classification results, e.g., by biasing the decision to only one class, as this would minimize the overall error. Several studies (e.g., Chawla, 2009; Dupret & Koda; Jain & Nag, 1997) show that training decision models on balanced samples leads to better classification results since models require sufficient exposure to the infrequent class to reach their full potential.

We additionally collected a second, distinct sample for testing the optimized models with naturally distributed data. This testing sample was collected randomly and represents the natural distribution of intermediaries with and without a history of misconduct. This enabled us to evaluate whether the trained models based on the balanced sample can deal with naturally distributed data. For the testing sample, we collected another 2,051 brokers—141 with a history of misconduct and 1,910 without.

Both samples were collected randomly to ensure the representativeness of the collected data. Because not all brokers have a profile on LinkedIn, we needed to scan more brokers on BrokerCheck than we included in the final dataset of 2,851 brokers that self-disclose information on their LinkedIn profile. To identify all 2,851 matched LinkedIn profiles, we inspected 4,729 registered brokers on BrokerCheck in total (1,319 brokers for the equally balanced sample and 3,410 for the naturally distributed sample). Consequently, we determined that 60.29% of the inspected financial intermediaries had a LinkedIn profile, which could unambiguously be assigned.

Since brokers with LinkedIn profiles might have different characteristics than brokers without a LinkedIn profile, it was necessary to rule out a potential selection bias. Therefore, we compared the information provided on BrokerCheck for both groups. Most importantly, our dependent variable (broker has committed misconduct or not) is almost identically distributed in both groups: While 6.87% of the intermediaries in our naturally distributed sample with a LinkedIn profile were identified as having committed misconduct, this is also true for 6.77% of the brokers who did not self-disclose information on LinkedIn. Moreover, the other characteristics provided by BrokerCheck are also highly comparable for brokers with and without LinkedIn profiles (see Table 1). Consequently, in particular regarding our dependent variable, there is no selection bias caused by merging the data with LinkedIn profiles. Nevertheless, brokers who self-disclosed information on LinkedIn have shorter average mean employment durations (83.08 vs. 97.00 months) and more state licenses (14.68 vs. 12.60), compared to brokers without LinkedIn profiles. Although these variables reveal similar distributions for both groups, these differences may weaken the generalizability of our results for brokers without LinkedIn profiles.

Table 1. Comparison of Brokers With and Without LinkedIn Profiles

	Brokers with LinkedIn profiles *N* = 2,851				Brokers without LinkedIn profiles *N* = 1,878			
Feature*	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**
Investment advisor (dummy variable)	0.00	1.00	0.55	0.50	0.00	1.00	0.49	0.50
Average employment duration (months)	0.50	598.00	83.08	66.11	1.00	442.00	97.00	77.78
Number of employment positions	1.00	35.00	3.65	2.67	1.00	29.00	3.72	2.86
Number of exams passed	1.00	12.00	4.14	1.45	0.00	16.00	3.42	1.57
Number of state licenses	0.00	60.00	14.68	16.70	0.00	59.00	12.60	15.54

Note: * For details on the features, please see Section 3.3.

Each broker in both final samples (i.e., balanced sample and naturally distributed sample) was registered on both the regulatory authority's website BrokerCheck and on the business network LinkedIn. Consequently, our classification models are based on a dataset composed of both autonomously self-disclosed information on LinkedIn and publicly reported regulator-confirmed information on BrokerCheck. This information makes our dataset unique and particularly useful for analyzing the information provided by financial intermediaries as well as different levels of external verification.

BrokerCheck contains information about the background and experience of brokers and investment advisors and discloses information about misconduct that resulted in regulatory actions, arbitrations, and complaints. In our dataset, the group of financial intermediaries who have a history of misconduct consists of brokers (who are also potentially registered as investment advisors) and have *customer disputes* and *regulatory actions* on their records. These disclosures of misconduct relate to actions taken in the role of a broker or investment advisor that damaged individual investors or society as a whole. *Customer disputes* are mainly based on the misbehavior of brokers, such as a misrepresentation of material facts, unsuitable recommendations of financial products, and securities fraud such as churning or front running. Examples of *regulatory actions* include unauthorized trading and insider trading. We excluded pending decisions and lawsuits and thus only considered those disclosures marked with a final status for which victim compensation or a fine had been paid. This ensured that we only included cases in which intermediaries admitted wrongdoing and were willing to pay victim compensation or where intermediaries were convicted of wrongdoing and ordered to pay a fine or victim compensation. Therefore, in the following, the term "misconduct" refers to a customer dispute or a regulatory event that is final, settled, or resolved with a judgment against the intermediary. Thereby, we do not differentiate between different types of misconduct because any intermediary misconduct is harmful to investors and weaken investors' trust in financial intermediaries and markets. Moreover, distributions and median values of incurred damages in Figure A2 in the Appendix show that the severity of different types of misconduct is highly comparable. Therefore, there is no need to differentiate between different types of misconduct for the purposes of our study. As supplementary information, we also report further summary statistics about brokers convicted of misconduct and different types of misconduct in Table A1 and Figure A1 in the Appendix. We define misbehaving intermediaries as those that have a history of one or more misconduct cases according to the criteria outlined above, whereas the group of financial intermediaries not committing misconduct consists of brokers with no record of misconduct at all. In order to account for the severity of different misconduct cases, we perform an economic evaluation in Section 4.3 that accounts for the amount of the compensation or fine levied against the intermediary.

In addition to the information collected from BrokerCheck, we manually collected self-disclosed information from matched profiles on LinkedIn. LinkedIn is the world's largest business-related social media networking website on which individuals can self-disclose personal information including, working experience, education, skills, and other relevant work-related information. LinkedIn profiles are matched with registered brokers on BrokerCheck by name, employment history, and location. We only added brokers to our dataset if matched LinkedIn profiles were distinct. If common names led to multiple possible profile matches, we further considered name suffixes, nicknames, previous jobs, or the unique FINRA identification number to match brokers to the correct profile. To control for fake profiles, we followed common techniques used in social network analysis (Adikari & Kaushik, 2014) and only considered profiles of intermediaries that include logical and reasonable information.

Therefore, in a first step, we qualitatively assessed whether the information provided in the different sections was free of self-contradiction. Specifically, we verified whether there was a logical flow concerning education and job experience, whether the disclosed skills were suitable for a broker or investment advisor, and whether the stated interests reflected current and previous employers, universities, and groups related to financial services topics. In a second step, we excluded those suspicious profiles whose number of connections to legitimate users was significantly below the average number of connections of profiles in our sample. In total, we judged only three profiles as suspicious based on the qualitative assessment.[3] Since these profiles all had fewer than five connections (three, one, and zero), they were excluded because the number of connections was significantly lower than the average number of 289.5 connections in our naturally distributed sample. However, our results remain robust even if we add the three potentially fake profiles to our dataset.

3.3 Feature Selection

We extracted a large number of features collected from publicly available information on BrokerCheck and LinkedIn. Figure 2 schematically depicts the presentation of information on both BrokerCheck and LinkedIn. As described in Section 3.2, in order to train classifiers to detect financial intermediaries who are likely to commit misconduct, we separated brokers into two groups: brokers with a history of misconduct and brokers with no record of misconduct. As the dependent variable, we use a binary variable of 1 for brokers with a history of misconduct and 0 for brokers without any misconduct record.

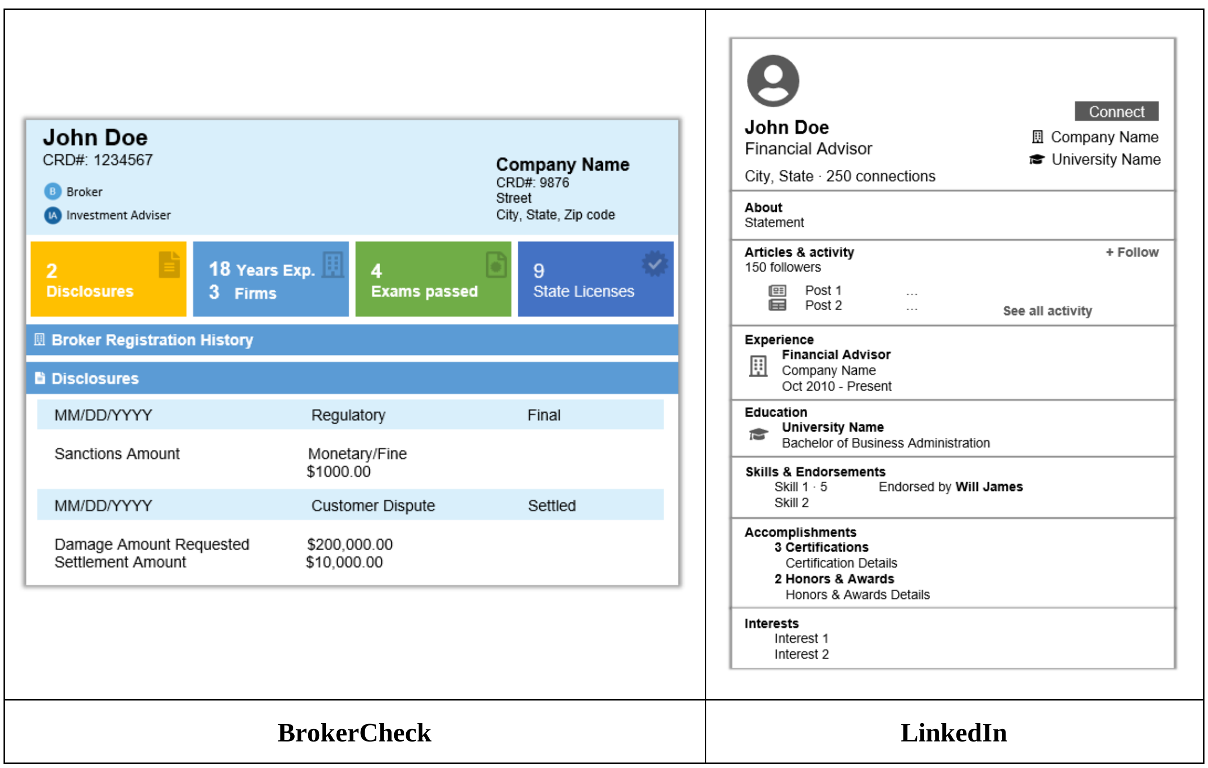

Figure 2. Publicly Available Information Provided on BrokerCheck and LinkedIn

[3] Since we explicitly searched for real-world individuals working for specific employers and since we use strict profile matching criteria based on the information disclosed on BrokerCheck, we ruled out fake profiles containing entirely made up information. Thus, we only found a small number of profiles that do not clearly satisfy the criteria of the qualitative assessment.

Table 2. Features Used for Classification Based on Self-Disclosed Structured Data and Linguistic Cues Derived from LinkedIn

Category	Feature	Description
personal information	li_male	Variable equaling 1 if broker is male, 0 if female
	li_picture	Variable equaling 1 if broker has a profile picture, 0 otherwise
	li_interests	Total number of self-disclosed interests
	li_location	Variable equaling 1 if location is urban, 0 otherwise
network activity	li_connections	Number of connections
	li_follower	Number of followers
	li_posts	Total number of posts
	li_rec_gi	Number of recommendations given on LinkedIn
professional information	li_job_cat	Classification of the self-disclosed career level into the following categories: advisor/analyst, senior advisor/associate, vice president, president/director/owner
	li_firm_cat	Classification of self-disclosed employer into the following categories: asset manager, bank, large bank, insurance, independent
	li_jobs	Number of self-disclosed employment positions; multiple positions are counted as separate jobs
	li_empl_details	Variable equaling 1 if durations of employment positions are self-disclosed, 0 otherwise
	li_avg_empl_dur	Average duration of employment positions in months calculated based on self-disclosed information
	li_cur_empl_details	Variable equaling 1 if duration of current employment is self-disclosed, 0 otherwise
	li_cur_empl_dur	Duration of current self-disclosed employment
	li_uni	Classification of the self-disclosed education level into the following categories: bachelor's degree, master's degree or higher, other university or college degrees
	li_uni_related	Variable equaling 1 if a self-disclosed university degree is job-related, 0 otherwise
	li_cert	Total number of self-disclosed certificates
	li_awards	Total number of self-disclosed awards
	li_skill	Number of self-disclosed skills
profile summary	li_sum	Variable equaling 1 if broker uses a profile summary disclosing a statement about herself, 0 otherwise
	li_sum_words	Number of words in the self-disclosed profile summary
	li_sum_neg_words	Share of negative words in the self-disclosed profile summary
	li_sum_pos_words	Share of positive words in the self-disclosed profile summary
	li_sum_str_words	Share of strong words in the self-disclosed profile summary
	li_sum_compl_words	Share of complex words in the self-disclosed profile summary
	li_sum_emtl_words	Share of emotional words in the self-disclosed profile summary
	li_sum_uncert_words	Share of words signaling uncertainty in the self-disclosed profile summary
	li_sum_modal_words	Share of modal words in the self-disclosed profile summary
	li_sum_wps	Number of words per sentence in the self-disclosed profile summary
	li_sum_fog	Fog index
	li_sum_sen	Sentiment derived from the self-disclosed profile summary

Table 2 contains all features and their descriptions based on self-disclosed structured data and linguistic cues derived from LinkedIn. The features are divided into four different categories: *personal information*, *social network activity*, *professional information*, and *profile summary*. Personal information includes information that describes the profile owner; social network activity concerns information on how active an intermediary is on the social network; professional information is information presented about work experience, previous employment positions, skills, and education; and profile summary refers to eye-catching information intended to give the profile viewer a striking first impression of the profile owner. We extract linguistic cues via textual analysis from the profile summary, as described in Section 3.4. All features in Table 2 represent self-disclosed information without external verification.

Based on the four communication principles of information manipulation theory, we expect self-disclosed information of intermediaries who are likely to commit misconduct to be different from self-disclosed information of intermediaries who are unlikely to commit misconduct. In particular, the first principle, understatement or exaggeration of the quantity of information provided, is reflected in all of our features representing quantitative information (e.g., number of posts, number of interests, or length of the profile summary) and is therefore represented in each category shown in Table 2. The second principle anticipates altered information or lies in the presentation of professional information and is typically represented in descriptions of past employments, education, or the profile summary. The third principle, covering the relevance of self-disclosed information, is likely to be reflected in the profile summary as well as in the overall quantity of information provided, e.g., regarding interests, skills, certificates, etc. The fourth principle, addressing ambiguity of information, is represented in the uncertainty and/or complexity expressed in the profile summary.

Table 3 describes different levels of externally confirmed information based on information manipulation theory as well as warranting theory. *User-confirmed information* refers to self-disclosed information on LinkedIn profiles that is confirmed by other users. *Regulator-confirmed information* refers to self-disclosed information that is confirmed by regulatory information published on BrokerCheck. Both categories represent information that cannot be easily manipulated. In particular, user recommendations and the number of endorsements per skill may relate to intermediary misconduct because we would anticipate that intermediaries who are likely to commit misconduct have significantly fewer recommendations than intermediaries who are unlikely to commit misconduct. Also, while intermediaries who are likely to commit misconduct might try to polish their profiles by advertising a variety of skills, as suggested by information manipulation theory, other users on LinkedIn, e.g., we would expect customers or colleagues of the intermediary to only endorse these skills if the intermediary does good work and is unlikely to commit misconduct. Thus, we expect that there should be a difference in the number of endorsements per skill for brokers who are likely to commit misconduct versus those who are unlikely to do so.

Table 3. Features Used for Classification Based on User- and Regulator-Confirmed Data

Category	Feature	Description
User-confirmed information	li_rec_ob	Number of obtained recommendations on LinkedIn
	li_end_skill	Number of endorsements per skill calculated from self-disclosed skills and their endorsements on LinkedIn
Regulator-confirmed information	bc_ia	Variable equaling 1 if a broker is also registered as an investment advisor on BrokerCkeck, and 0 otherwise
	bc_avg_empl_dur	Average duration of employment positions in months calculated based on regulatory disclosed information on BrokerCheck
	bc_jobs	Number of employment positions according to BrokerCheck
	bc_exams	Number of passed exams according to BrokerCheck
	bc_licences	Number of state licenses according to BrokerCheck
	bc_li_exp_dev	Deviation of work experience between LinkedIn and BrokerCheck in months
	bc_li_jobs_dev	Deviation of number of employment positions between LinkedIn and BrokerCheck

Finally, since financial intermediaries cannot manipulate the information provided on BrokerCheck, regulator-confirmed information is more reliable than user-confirmed external validation. We thus presume that discrepancies between the LinkedIn profile and information published on BrokerCheck are valuable for identifying misconduct. In particular, we anticipate that deviations between regulatory information on BrokerCheck and self-disclosed information on LinkedIn will be useful for the detection of intermediary misconduct. First, several brokers with misconduct in our sample conceal frequent job changes on LinkedIn—in particular brief employment periods and employment related to a misconduct event. In our sample, we identified brokers that reported only some of their jobs (e.g., four out of eight jobs, or 5 out of 11 jobs), thus revealing a deviation between their full employment history on BrokerCheck and their self-disclosed history on LinkedIn. Second, several brokers did not disclose their employment history before the point at which a misconduct event occurred. For example, two brokers in our sample only disclosed their most recent job (out of three) because they committed misconduct at their second job, again revealing a deviation between the information they provided on LinkedIn and that reported by BrokerCheck. Third, some brokers misrepresented their work experience on LinkedIn and reported more work experience on LinkedIn than reported on BrokerCheck to appear more experienced.

3.4 Data Preprocessing and Textual Analysis

Preprocessing is necessary for nonnumeric features so that these features can be used by machine learning algorithms. In particular, self-disclosed information regarding the firm where the intermediary is employed, the current position, and the location provided on LinkedIn has to be categorized. We classified all firms and job titles into four categories each (see Table 2). Specifically, we categorized self-provided job titles in terms of career level according to standard career levels in the financial industry (Eccles & Crane, 1987). Firms were categorized based on their primary business model (e.g., bank or insurance company) and banks were further split into large and small institutions based on the total assets reported in their annual filings.[4] We defined the location of the broker or investment advisor as urban if the city or metropolitan area provided on LinkedIn had more than 200,000 inhabitants and rural otherwise. To make categorial features processable for our machine learning techniques, we used one-hot encoding (also called dummy encoding), which is a standard approach for nominal variables (Wooldridge, 2009).

In order to analyze the profile summaries that brokers and investment advisors provided to describe themselves on LinkedIn, we performed common text preprocessing steps and generate quantitative linguistic features. First, we removed parts of the text that did not contain relevant information, such as email addresses, website URLs, numbers, single-character words, and state abbreviations. In addition, we removed dots that did not represent the end of a sentence (e.g., in company suffixes such as Inc. or Ltd., common abbreviations such as Mr., No., or Jr., and those following middle initials). Second, we transform the cleaned text into lower-case letters and split the text into individual words.

We used the Harvard IV-4 dictionary to calculate common textual analysis measures such as share of positive, negative, strong, and emotional words. Although there are specific dictionaries tailored to financial contexts (Loughran & McDonald, 2011), we relied on the more general Harvard IV-4 dictionary because brokers and investment advisors introduce themselves in profile summaries using general rather than financial language. In addition, we followed Zhou et al. (2004) to determine the share of uncertainty and modal words to measure uncertainty in texts. Based on the number of positive and negative words, we also determined the sentiment of the profile summaries (see Equation (1)). To analyze the complexity of the profile summary, we calculated the average number of words per sentence using the Stanford CoreNLP toolkit (Manning et al., 2014). We also included the share of complex words (i.e., a word with more than two syllables) and the fog index (Li, 2008) as depicted in Equation (2) as readability measures.

$$\text{sentiment} = \frac{positive\ words - negative\ words}{positive\ words + negative\ words} \tag{1}$$

$$\text{fog index} = 0.4\ (Words\ per\ sentence + percentage\ of\ complex\ words) \tag{2}$$

[4] Total assets are based on the annual financial statements as of 2017. The critical threshold for large banks amounts to USD 800 billion, which separates large and small banks at the observed gap in the data.

For most of the self-disclosed information gathered from LinkedIn, users deliberately decide whether to provide information on a specific category. For example, users can disclose their number of connections, skills, or interests or they may choose to disclose none of this information. In the latter case, the variables measuring information disclosure were set to 0. Also, if a broker had no written profile summary, all variables based on textual analysis of the profile summary were set to 0. Missing values in a narrower sense only exist if employment durations were not self-disclosed on LinkedIn, making the deviation with regulator-confirmed information on BrokerCheck not measurable (less than 6% of the observations). We replaced these missing values with 0 and included a dummy control variable to check whether employment details were disclosed on LinkedIn (*li_empl_details*).

Since many machine learning techniques require standardized data because they would otherwise estimate a larger effect for variables on a larger scale, we standardized our numerical features with zero mean and unit variance and used a K-nearest-neighbor (K = 50) approach based on all features to drop outliers with distances above the 99th percentile in our training data to avoid biases in our models (James et al., 2017).

3.5 Machine Learning Techniques Applied

We relied on different machine learning techniques in order to develop classifiers to detect financial intermediaries who are likely to commit misconduct. Specifically, we used logistic regression (LOG) as a baseline, and apply support vector machine (SVM), decision tree (DT), random forest (RF), naive Bayes (NB), and artificial neural networks (ANN) [5] as machine learning techniques. These techniques have generated promising results in different data mining applications (e.g., Dong et al., 2018; Humpherys et al., 2011; Kirkos et al., 2007). For technical details regarding the different machine learning techniques, we refer to the literature on this topic (e.g., Duda, Hart, & Stork, 2012; Han & Kamber, 2006; James et al., 2017; Vapnik, 1998). Comprehensive overviews and detailed discussions of different machine learning methods for financial fraud detection are provided by, e.g., Ngai et al. (2011), Bhattacharyya et al. (2011), and West & Bhattacharya (2016).

In order to ensure robust and generalizable models, we applied a bagging classifier approach, which is widely used for classification problems, in order to avoid overfitted models (Breiman, 1996). Specifically, for each machine learning technique, we trained multiple models using a random bootstrap sample of 80% of our data for each single model and performed classification using a majority vote of all classifiers. Since RF classifiers already represent a specific kind of bagging classifier (also using a random subset of features for each single model in the forest), we did not use an additional bagging classifier for RF. We also did not apply bagging for our ANN models because their performance is better when using the whole training dataset versus using the bagging classifier approach.

3.6 Classifier Configuration and Hyperparameter Tuning

To analyze our research hypotheses, which predict that self-disclosed, user-confirmed, and regulator-confirmed information is valuable for detecting financial intermediary misconduct, we created different classifiers based on different levels of external verification. Table 4 provides an overview of all composed classifiers used for our empirical analysis. Each classifier configuration represents a different level of verification. Classifier A is the baseline and uses only features based on self-disclosed information, whereas Classifiers B, C, and D additionally use different sets of features based on user- and regulator-confirmed information. For each classifier, we applied different machine learning techniques as described in Section 3.5. For the sake of completeness, we also included one classifier based on regulator-confirmed information only and one classifier based on regulator-confirmed and user-confirmed information. The configuration of these additional classifiers can be found in the Appendix (see Table D1).

Table 4. Classifier Configuration

Classifier	Self-disclosed information	User-confirmed information	Regulator-confirmed information
A	X		
B	X	x	
C	X		X
D	X	x	X

[5] We use feed-forward neural networks.

We trained each machine learning technique for all classifiers on the balanced sample and tune hyperparameters to optimize the F1 score using a grid search. As described in Section 3.5, we used bagging classifiers and thus further optimized the number of trained classifiers for each machine learning technique. An overview of the tuned hyperparameters, the respective parameter grids, and the configuration for the best model for each machine learning technique can be found in Table B1 in the Appendix.

3.7 Evaluation Methodology

3.7.1 Statistical Evaluation

For the training and optimization of our models using the balanced sample, we use tenfold stratified cross-validation in order to avoid overfitting of the models. This technique has been proven to be the best method for model selection in case of real-world datasets (Kohavi, 1995). Then, we evaluated the classification performance of the resulting classifiers using the naturally distributed sample. In each case, we calculated a confusion matrix and computed the common performance metrics accuracy, recall, precision, specificity, and the F1 score (Sokolova & Lapalme, 2009). In order to evaluate the performance between the different classifiers, we used McNemar's test (Everitt, 1977), which compares the performance of two different classifiers. Since McNemar's test is a two-sided test, we also reported the direction in which one classifier outperformed the other.

In addition to assessing one specific configuration of a classifier, we also evaluated our models when considering different classification thresholds that need to be reached to classify an observation as positive. For this purpose, we used two different common graphical representations of the classification thresholds: the precision-recall curve and the receiver operating characteristic (ROC) curve. The precision-recall curve plots the relationship between precision (y-axis) and recall (x-axis) for all possible classification thresholds and visualizes the interdependence of these two measures when the classification threshold changes. Consequently, it is particularly informative for imbalanced datasets (Saito & Rehmsmeier, 2015). The ROC curve simultaneously displays the two classification errors: the Type 1 error (x-axis, false-positive rate = 1-specificity) and the Type 2 error (y-axis, recall = 1-Type 2 error) for all possible classification thresholds, while the area under the curve (AUC) summarizes the overall performance of a classifier over all possible classification thresholds (James et al., 2017).

3.7.2 Economic Evaluation

Beyond the above-mentioned machine learning metrics, we also performed an economic evaluation of the classifiers proposed in this paper. Domain-specific evaluations are important for assessing the value of classifiers created for specific classification problems and allow for additional statistical analysis (Groth, Siering, & Gomber, 2014). To assess the economic gain achievable by a misconduct detection mechanism, we designed an evaluation methodology that accounts for interaction between investors and financial intermediaries. Specifically, we derived the economic value of an automated classifier by computing the investor's potential damage that could be avoided by using the classifiers. Since we used the classification results of the naturally distributed testing sample representing randomly collected real-world data, the economic evaluation is representative.

Four different cases have to be considered when using the classifiers: (1) If a financial intermediary is classified correctly and subsequently commits misconduct (true positive, TP), an economic loss in the amount of the investor's damage is prevented, which can thus be considered an economic gain. To approximate an investor's damage, we rely on the compensation payment (cp) paid by the financial intermediary. We thereby account for the severity of different misconduct events. Nevertheless, in this case, the investor must select a different intermediary to execute her or his trade or investment, which leads to additional search costs (sc). (2) If an intermediary is classified incorrectly and actually commits misconduct (false negative, FN), the investor incurs a damage equal to the compensation payment. (3) If the financial intermediary is incorrectly classified as an intermediary who is likely to commit misconduct (false positive, FP), the investor will unnecessarily select a new intermediary and must bear additional search costs. (4) If the intermediary is correctly classified as someone who is unlikely to commit misconduct (true negative, TN), the investor will continue working with this intermediary and will bear no additional costs. Based on these considerations, we calculate the economic gain resulting from the classification (c) of each intermediary (i) as outlined in Equation (3).

$$economic\ gain_i = \begin{cases} cp_i - sc & if\ c\ \in \{TP\} \\ -cp_i & if\ c\ \in \{FN\} \\ -sc & if\ c\ \in \{FP\} \\ 0 & if\ c\ \in \{TN\} \end{cases} \tag{3}$$

Specifically, the economic gain is calculated per broker and case and is based on the compensation payment (*cp*) levied against the broker, or the average compensation payment if a broker has a history of several misconduct events. Search costs as defined above refer to the cost of finding another suitable intermediary and include corresponding opportunity costs (e.g., resulting from nonexecuted trades). Search costs differ among investors depending on the effort necessary to find a new intermediary to work with and the individual loss incurred because of lost opportunity costs. Therefore, search costs are difficult to quantify. Nevertheless, compared to potential damages an investor might suffer, for example, through losses incurred in retirement plans because of unsuitable investment advice or false information, search costs are negligible (Egan, 2019). Thus, we assume search costs within our economic evaluation to be zero. However, we conduct a sensitivity analysis with different levels of search costs and varying classification thresholds to ensure the robustness of our results.

To analyze the economic value of the different classifiers, we compared their average economic gain, averaging across the economic gains resulting from the classification of each intermediary in the naturally distributed sample separately for each combination of classifier and machine learning technique. While the proposed evaluation reveals the economic value from the investors' perspective, it also corresponds to the regulator's objective function, which is to protect investors by ensuring fair and efficient markets.

4 Empirical Study

4.1 Descriptive Statistics

Table 5 provides the descriptive statistics for all features of the balanced dataset and the results of the Wilcoxon Rank-Sum (WRS) test for equality of means between values of the features for financial intermediaries with and without misconduct histories. For most features representing self-disclosed information on LinkedIn, we observe differences in means; however, for more than one third of the features, differences between mean values for financial intermediaries with and without misconduct are significant. For example, intermediaries with misconduct tend to have significantly longer profile summaries, although the content tends to be significantly more difficult to read (based on fog index, words per sentence, share of complex words). This indicates a more ambiguous content, which, in line with information manipulation theory, suggests that dishonest individuals tend to polish their profiles to mislead receivers. Thus, this indicates that self-disclosed information on LinkedIn is potentially useful to identify misconduct among financial intermediaries.

As theorized by warranting theory, manipulation becomes more difficult when information is externally validated and thus one might assume that intermediaries who are unlikely to commit misconduct receive more external validation than intermediaries who are likely to commit misconduct. For features representing user-confirmed information (a weaker form of verification), the descriptive statistics support this assumption. In particular, the number of skill endorsements is significantly higher for intermediaries without misconduct, compared to intermediaries who have committed misconduct. Moreover, the number of recommendations is also higher for intermediaries with no history of misconduct, although the difference is not significant.

For regulator-confirmed information, the WRS test shows highly significant differences for intermediaries with and without misconduct for all features. This is especially true for features that account for deviations between self-disclosed information and information provided by BrokerCheck. These results provide an initial indication that features based on user- and regulator-confirmed information are especially valuable for detecting financial intermediaries who commit misconduct.

Table C1 in the Appendix provides the descriptive statistics for the naturally distributed sample. We observe similar differences between the misconduct and no-misconduct classes. Moreover, we conducted a WRS test to investigate whether the training and testing samples are comparable with respect to the features used to detect misconduct. At the 5% level, there is no significant difference between intermediaries with misconduct histories and those without in the training and testing sample. Those without misconduct in the testing sample show significant but small differences concerning the number of self-disclosed jobs on LinkedIn, indicating that the training and testing samples are comparable in terms of the features used to detect intermediary misconduct.

4.2 Classifier Evaluation

4.2.1 Cross-Validation Results Based on the Balanced Sample

As described in Section 3.2, we trained and optimized our classifiers for each machine learning technique based on the balanced sample. Table 6 presents the ten-fold stratified cross-validation results of the trained and optimized models. The parameter configurations of the best classifiers for each machine learning technique are reported in Table B1 in the Appendix. The evaluation metrics show meaningful values for all classifiers and machine learning techniques. Specifically, DT and RF yield the highest values for most of the metrics and the majority of classifiers, closely followed by LOG and ANN.

Table 5. Descriptive statistics for the Balanced Sample and Wilcoxon Rank-Sum Test for Equality of Means

		Misconduct (*N* = 400)				No misconduct (*N* = 400)				WRS test
Feature		Min	Max	Mean	SD	Min	Max	Mean	SD	*P*
Self-disclosed										
personal information	li_male	0.0	1.0	0.9	0.3	0.0	1.0	0.7	0.5	0.00***
	li_ picture	0.0	1.0	0.5	0.5	0.0	1.0	0.5	0.5	0.43
	li_ Interests	0.0	174.0	10.2	15.3	0.0	215.0	12.3	19.5	0.04**
	li_location	0.0	1.0	0.7	0.5	0.0	1.0	0.8	0.4	0.01***
network activity	li_ connections	0	500	240	179	0	500	286	189	0.00***
	li_ follower	0	6389	171	595	0	6031	153	491	0.65
	li_ posts	0.0	50.0	8.3	17.3	0.0	50.0	7.1	15.5	0.55
	li_ rec_gi	0.0	14.0	0.3	1.0	0.0	4.0	0.3	0.8	0.34
professional information	li_job_adv	0.0	1.0	0.3	0.5	0.0	1.0	0.3	0.5	0.85
	li_job_vp	0.0	1.0	0.3	0.5	0.0	1.0	0.3	0.4	0.11
	li_job_pres	0.0	1.0	0.3	0.4	0.0	1.0	0.2	0.4	0.22
	li_job_sen	0.0	1.0	0.1	0.2	0.0	1.0	0.1	0.3	0.04**
	li_company_larbank	0.0	1.0	0.3	0.4	0.0	1.0	0.2	0.4	0.46
	li_company_bank	0.0	1.0	0.1	0.2	0.0	1.0	0.1	0.3	0.39
	li_company_inde	0.0	1.0	0.6	0.5	0.0	1.0	0.5	0.5	0.00***
	li_company_insur	0.0	1.0	0.1	0.2	0.0	1.0	0.1	0.3	0.11
	li_company_am	0.0	1.0	0.0	0.2	0.0	1.0	0.1	0.3	0.22
	li_jobs	0.0	11.0	2.5	1.8	0.0	13.0	2.6	1.8	0.13
	li_empl_details	0.0	1.0	1.0	0.2	0.0	1.0	0.9	0.3	0.33
	li_avg_empl_dur	2.7	766.2	150.5	119.9	3.0	572.1	115.8	99.1	0.00***
	li_cur_empl_details	0.0	1.0	0.9	0.3	0.0	1.0	0.9	0.3	0.90
	li_cur_empl_dur	0.0	766.0	149.4	132.5	0.0	573.0	122.0	112.6	0.15
	li_uni_ba	0.0	1.0	0.6	0.5	0.0	1.0	0.7	0.5	0.20
	li_uni_ma	0.0	1.0	0.2	0.4	0.0	1.0	0.2	0.4	0.10*
	li_uni	0.0	1.0	0.8	0.4	0.0	1.0	0.8	0.4	0.50
	li_uni_related	0.0	1.0	0.5	0.5	0.0	1.0	0.6	0.5	0.46
	li_ cert	0.0	8.0	0.9	1.5	0.0	10.0	0.7	1.4	0.11
	li_ awards	0.0	8.0	0.2	0.9	0.0	10.0	0.1	0.7	0.43
	li_ skill	0.0	50.0	8.1	11.0	0.0	50.0	10.9	12.1	0.00***
profile summary	li_sum	0.0	1.0	0.6	0.5	0.0	1.0	0.5	0.5	0.00***
	li_sum_words	0.0	314.0	56.7	78.5	0.0	319.0	43.6	68.6	0.00***
	li_ sum_neg_words	0.0%	11.1%	0.2%	0.7%	0.0%	6.3%	0.2%	0.7%	0.95
	li_ sum_pos_words	0.0%	25.0%	1.9%	2.9%	0.0%	19.0%	1.7%	2.9%	0.05*
	li_ sum_str_words	0.0%	50.0%	3.3%	4.9%	0.0%	69.2%	2.9%	5.5%	0.01**
	li_ sum_compl_words	0.0%	75.0%	16.5%	15.1%	0.0%	61.5%	13.4%	15.7%	0.01***
	li_ sum_emtl_words	0.0%	3.2%	0.1%	0.4%	0.0%	4.3%	0.1%	0.4%	0.49
	li_ sum_uncert_words	0.0%	3.9%	0.2%	0.6%	0.0%	8.2%	0.2%	0.7%	0.72
	li_ sum_modal_words	0.0%	9.1%	0.4%	0.9%	0.0%	8.5%	0.3%	0.8%	0.20
	li_ sum_wps	0.0	60.0	10.9	9.9	0.0	82.0	8.3	10.5	0.00***
	li_ sum_fog	0.0	31.6	11.0	9.2	0.0	43.0	8.7	9.7	0.00***
	li_ sum_sen	-1.0	1.0	0.4	0.5	-1.0	1.0	0.3	0.5	0.02**

User confirmed										
	li_ rec_ob	0.0	7.0	0.1	0.6	0.0	11.0	0.3	1.2	0.31
	li_end_skill	0.0	28.9	1.5	3.5	0.0	43.4	2.3	4.4	0.00***
Regulator confirmed										
	bc_ia	0.0	1.0	0.8	0.4	0.0	1.0	0.5	0.5	0.00***
	bc_avg_empl_dur	6.1	598.0	110.9	94.9	3.0	415.0	86.0	71.6	0.00***
	bc_jobs	1.0	35.0	4.3	3.2	1.0	13.0	3.4	2.3	0.00***
	bc_exams	1.0	12.0	4.5	1.6	1.0	11.0	4.0	1.4	0.00***
	bc_licenses	0.0	55.0	17.7	13.5	0.0	55.0	12.8	16.5	0.00***
	bc_li_exp_dev	0.0	554.0	76.8	102.2	0.0	410.0	77.7	81.6	0.00***
	bc_li_jobs_dev	0.0	30.0	2.5	2.8	0.0	12.0	1.7	1.8	0.00***

Note: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Table 6. Cross-Validation Results Based on the Balanced Sample Scores for the Different Evaluation Metrics Are Reported in %)

	Classifier A					**Classifier B**				
Cues	**Self-disclosed information**					**Self-disclosed + user-confirmed information**				
Tech.	**Acc.**	**Rec.**	**Prec.**	**Spec.**	**F1**	**Acc.**	**Rec.**	**Prec.**	**Spec.**	**F1**
LOG	61.11	65.75	60.65	56.48	62.96	61.99	67.68	61.17	56.31	64.08
SVM	63.26	63.28	63.26	63.21	63.26	61.49	63.22	61.49	59.75	62.21
DT	63.01	64.74	62.86	61.27	63.60	63.01	66.00	62.53	60.01	64.07
RF	64.27	66.75	64.00	61.78	65.17	63.13	65.91	62.52	60.35	64.07
NB	62.24	74.81	59.97	49.62	66.50	61.45	74.29	59.32	48.56	65.86
ANN	62.12	64.90	61.85	59.34	63.18	61.99	63.38	61.92	60.61	62.52
	Classifier C					**Classifier D**				
Cues	**Self-disclosed + regulator-confirmed information**					**Self-disclosed + user- + regulator-confirmed information**				
Tech.	**Acc.**	**Rec.**	**Prec.**	**Spec.**	**F1**	**Acc.**	**Rec.**	**Prec.**	**Spec.**	**F1**
LOG	70.45	76.26	68.43	64.65	72.06	70.45	76.52	68.30	64.39	72.12
SVM	70.58	73.18	69.71	68.01	70.58	70.33	74.44	69.00	66.25	71.37
DT	74.12	78.24	72.60	70.01	75.21	73.23	76.21	72.16	70.27	74.00
RF	73.48	78.54	71.53	68.43	74.80	73.74	77.02	72.46	70.45	74.60
NB	66.45	71.73	64.33	61.30	67.61	67.24	73.07	65.02	61.56	68.58
ANN	70.96	74.94	69.29	67.01	71.95	71.59	76.21	69.79	67.01	72.74

Comparing the Classifiers A through D, our results offer initial indications for the evaluation of our research hypotheses: First, Classifier A achieves meaningful scores, indicating that compared to random guessing (which would achieve an accuracy score of 50% for balanced data), classifiers based on self-disclosed information can add value to the detection of intermediary misconduct (H1).

Second, while Classifier B achieves slightly lower scores than Classifier A for most of the machine learning techniques and since the results of comparing Classifier D with Classifier C are mixed, we cannot yet determine that user-confirmed information is valuable for the detection (H2). Third, the comparison of classifiers including regulator-confirmed information (C and D) with Classifier A reveals that Classifiers C and D achieve higher scores for almost all evaluation metrics, suggesting that regulator-confirmed information may add value to the classification models (H3). Nevertheless, these findings represent only initial indications based on the training results from the ten-fold stratified cross-validation; the hypotheses need to be further analyzed based on the evaluation of the naturally distributed hold out sample.

Table 7. Classifier Evaluation for Classifiers Using Self-Disclosed Information Only and McNemar's Test Results on Classifier Performance (in %, Naturally Distributed Sample)

	Classifier A						
Cues	**Self-disclosed information**					**McNemar's test**	
Techn.	**Acc.**	**Rec.**	**Prec.**	**Spec.**	**F1**	**A vs. Naive**	
LOG	64.16	71.63	12.69	63.61	21.56	0.00***	A > Naive
SVM	69.58	62.41	13.35	70.10	22.00	0.00***	A > Naive
DT	67.38	73.76	14.13	66.91	23.72	0.00***	A > Naive
RF	64.02	74.47	13.01	63.25	22.15	0.00***	A > Naive
NB	58.41	72.34	11.14	57.38	19.30	0.00***	A > Naive
ANN	67.67	72.34	14.05	67.33	23.53	0.00***	A > Naive

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

4.2.2 Classifier Evaluation and Analysis Based on the Naturally Distributed Sample

H1: Self-disclosed information. Table 7 presents the results of using the trained machine learning models to account for self-disclosed personal information in the absence of external verification (Classifier A) for classifying the naturally distributed sample.

All machine learning techniques achieve meaningful accuracy, recall, and specificity, showing that the majority of true misconduct cases (recall of up to 74.47% for RF) and true non-misconduct cases (specificity up to 70.10% for SVM) are classified correctly. Still, precision and thus the F1 score yield lower values (maximum precision score is 14.13% for DT). This implies that in terms of the share of all predicted misconduct cases, only 14.13% (every seventh case) represent true misconduct cases. Nevertheless, this precision score is in line with other classifiers that have been applied to highly imbalanced datasets (Tan et al., 2015; Zhang & Mani, 2003) and has been regularly observed in the context of datasets including very few observations of the class to be predicted (Menzies et al., 2007).

Moreover, from the perspective of regulators and/or supervisors, even though cases must be subsequently inspected manually, using the classifier as a decision support tool would be useful because it improves the hit ratio from every 15th to every 7th case, compared to randomly selecting a subset of all financial intermediaries (given the historical unequal distribution of only 6.83% of true misconduct cases). Since regulatory/supervisory resources are limited, using the proposed classifiers would free up the capacity to manually inspect more intermediaries. This is discussed by Zhang et. al., (2004) who develop a learning algorithm for fraud detection in transaction data. From the perspective of investors, besides predicting true misconduct cases correctly (measured by recall), it is even more important that reliable brokers are predicted correctly (for our classifiers, this holds for 96% to 98% of the predictions according to the negative predictive value, which equals the share of true non-misconduct cases within the share of predicted non-misconduct cases). Yet, recall and the correct detection of misconduct cases become increasingly important in the context of rising search costs. We further elaborate on this topic in our economic evaluation in Section 4.3.

The value of self-disclosed information for detecting intermediaries who are likely to commit misconduct is also supported by McNemar's test. For all machine learning techniques, Classifier A significantly outperforms a naive classification algorithm that randomly classifies financial intermediaries as having committed misconduct or not based on the historical misconduct ratio of 6.83%. Consequently, given the high percentage of correctly classified misconduct cases, as well as the high percentage of correctly classified non-misconduct cases, we conclude that self-disclosed personal information can be used to detect financial intermediaries who are likely to commit misconduct, supporting H1.

H2: User-confirmed information. Table 8 presents the results of Classifier B, which additionally takes self-disclosed information verified by other users into account. While recall is slightly higher for most of the machine learning techniques, as compared to Classifier A, accuracy, precision, and specificity are lower. The results of McNemar's test even show that Classifier A significantly outperforms Classifier B for five of the six machine learning techniques. Only NB yields better overall performance. Consequently, the results do not provide support for H2.

In order to analyze potential reasons for this result, Table 9 shows detailed summary statistics for features representing user-confirmed information on LinkedIn and provides evidence for why features based on user-confirmed information do not improve the classification. Looking at the number of obtained recommendations (li_rec_ob), only 154 out of the 2,051 brokers in the naturally distributed sample received recommendations from other users, while intermediaries without misconduct received more recommendations. Nevertheless, for the large majority of observations, this feature does not provide any information for the classifiers (about 90% of the observations).

For the feature li_end_skill representing the number of endorsements per skill on LinkedIn, there are many more observations that have values larger than zero (890 out of 2,051). Also, the percentiles show differences between misconduct versus non-misconduct cases that confirm the results of the WRS test in Table 5 and Table C1. Nevertheless, for more than 50% of the observations, this feature does not provide significant information for the classifiers. Thus, depending on the machine learning technique applied and the respective hyperparameter configuration of the model, both features based on user-confirmed information may not deliver sufficient information gain to improve the models. Moreover, including this type of information may lead to models that perform even worse than those that only consider self-disclosed information.

H3: Regulator-confirmed information. Compared to user-confirmed information (i.e., recommendations and endorsements on LinkedIn), which depends on the motivation of other users to provide content (Crowston & Fagnot, 2018), the provision of regulator-confirmed information is mandatory. Therefore, regulator-confirmed information is available for every broker and provides an even stronger confirmation of self-disclosed information than user confirmations.

Table 8. Classifier Evaluation for Classifiers Using Self-Disclosed and User-Confirmed Information and McNemar's Test Results on Classifier Performance (in %, Naturally Distributed Sample)

	Classifier B						
Cues	**Self-disclosed + user-confirmed information**					**McNemar's test**	
Techn.	**Acc.**	**Rec.**	**Prec.**	**Spec.**	**F1**	**B vs. A**	
LOG	60.12	73.76	11.75	59.11	20.27	0.00***	A > B
SVM	61.53	77.30	12.59	60.37	21.65	0.00***	A > B
DT	63.29	73.76	12.68	62.51	21.64	0.00***	A > B
RF	61.87	75.18	12.43	60.89	21.33	0.00***	A > B
NB	60.02	70.92	11.38	59.21	19.61	0.00***	B > A
ANN	63.73	73.76	12.82	62.98	21.85	0.00***	A > B
Note: *p < 0.1, **p < 0.05, ***p < 0.01.							

Table 9. Detailed Summary Statistics for Features Based on User-Confirmed Information

Number of obtained recommendations on LinkedIn (li_rec_ob)										
Misconduct	**Count**	**Mean**	**SD**	**Min**	**50%**	**75%**	**90%**	**95%**	**99%**	**Max**
0	1910	0.22	1.68	0	0	0	0	1	5	60
1	141	0.03	0.21	0	0	0	0	0	1	2
Note: Of the overall 2,051 observations, 154 observations have a value greater than zero.										
Proportion of endorsements to skills on LinkedIn (li_end_skill)										
Misconduct	**Count**	**Mean**	**SD**	**Min**	**50%**	**75%**	**90%**	**95%**	**99%**	**Max**
0	1910	2.57	4.35	0.00	0.00	3.99	8.00	11.30	19.98	40.00
1	141	1.34	2.90	0.00	0.00	0.56	5.92	8.00	11.89	15.24
Note: Of the overall 2,051 observations, 890 observations have a value greater than zero.										

Table 10 provides the results generated by applying the trained models on the naturally distributed sample for classifiers when self-disclosed information verified by regulatory authorities is also taken into account. Compared to Classifiers A and B, Classifiers C and D yield higher scores for all machine learning metrics, indicating that self-disclosed information combined with regulator-confirmed information adds value to the detection of intermediary misconduct. Classifier D using RF shows the highest overall performance for all machine learning metrics with an accuracy of 77.96%, correctly predicting 75.18% of true misconduct cases (recall) and 78.17% of true non-misconduct cases (specificity).

Further, Classifier D based on RF yields the highest precision and F1 score, leading to a higher proportion of true misconduct cases identified in all predicted misconduct cases than all other classifiers. Classifier D based on RF is closely followed by Classifier D using DT, and Classifier D based on SVM performs best in terms of identifying true misconduct cases (recall of 77.30%). Nevertheless, for SVM, accuracy, precision, and specificity are lower in comparison to the other machine learning techniques, meaning that we determine Classifier D based on RF to be the best-performing classifier.

Table 11 provides the results of McNemar's test comparing the performance of the different classifiers. The results show that, compared to Classifiers A and B, Classifiers C and D significantly add value, thus supporting H3. In line with warranting theory, regulatory confirmations provide a stronger signal because they are harder to manipulate than user confirmations and thus add additional value to the classification. Furthermore, when comparing Classifiers C and D, we obtain improved classification results for classifiers that include user-confirmed information (Classifier D) for four of the six machine learning techniques. This suggests that user-confirmed information can be valuable in combination with regulator-confirmed information, especially for tree-based models and neural networks. For these models, the apparently low availability of user-confirmed information in combination with regulator-confirmed information provides enough information gain to be incorporated into the models, therefore improving the performance of the classifiers.

Table 10. Classifier Evaluation for Classifiers Using Self-Disclosed as Well as User and Regulator-Confirmed Information (in %, Naturally Distributed Sample)

	Classifier C					**Classifier D**				
Cues	**Self-disclosed + regulator-confirmed**					**Self-disclosed + user- + regulator-confirmed**				
Tech.	**Acc.**	**Rec.**	**Prec.**	**Spec.**	**F1**	**Acc.**	**Rec.**	**Prec.**	**Spec.**	**F1**
LOG	69.62	76.60	15.47	69.11	25.74	70.75	75.89	15.90	70.37	26.29
SVM	72.31	74.47	16.48	72.15	26.99	69.62	77.30	15.57	69.06	25.92
DT	70.16	82.98	16.60	69.21	27.66	75.96	73.05	18.46	76.18	29.47
RF	75.87	75.18	18.73	75.92	29.99	77.96	75.18	20.27	78.17	31.93
NB	69.82	71.63	14.85	69.69	24.60	69.04	71.63	14.51	68.85	24.13
ANN	69.28	75.18	15.12	68.85	25.18	72.60	74.47	16.64	72.46	27.20

Table 11. McNemar's Test Results on Classifier Performance for Classifiers Using Self-Disclosed as Well as User and Regulator-Confirmed Information Compared to Classifiers A and B as Benchmarks (Naturally Distributed Sample)

	Classifier	**C**		**D**		**D**		**D**	
	Benchmark	**A**		**A**		**B**		**C**	
Tech.									
LOG		0.00***	C > A	0.00***	D > A	0.00***	D > B	0.00***	D > C
SVM		0.01**	C > A	0.96	D > A	0.00***	D > B	0.00***	C > D
DT		0.01**	C > A	0.00***	D > A	0.00***	D > B	0.00***	D > C
RF		0.00***	C > A	0.00***	D > A	0.00***	D > B	0.00***	D > C
NB		0.00***	C > A	0.00***	D > A	0.00***	D > B	0.00***	C > D
ANN		0.12	C > A	0.00***	D > A	0.00***	D > B	0.00***	D > C

Note: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

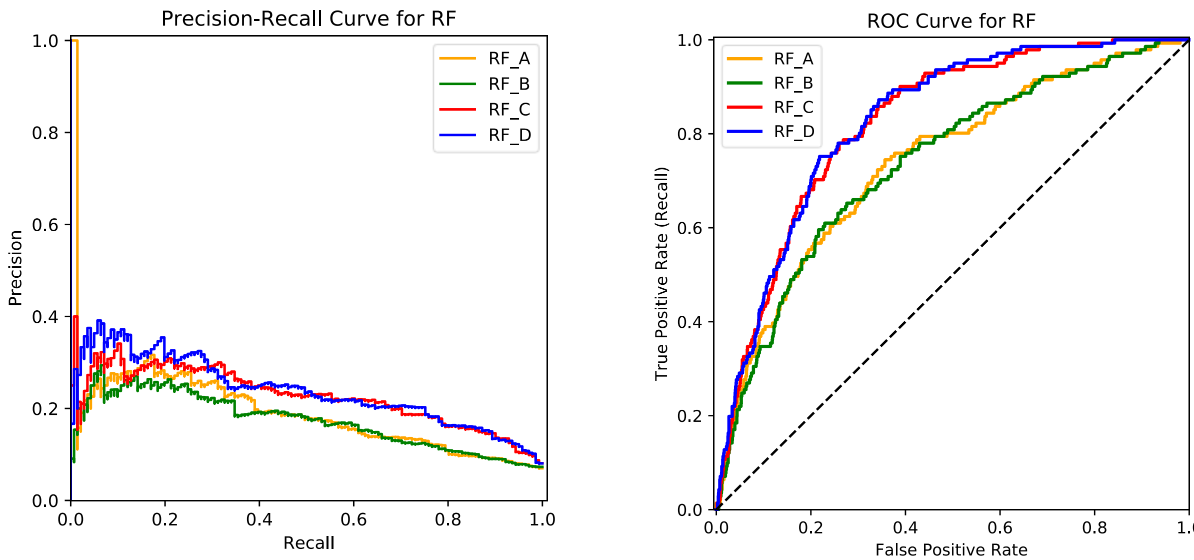

Figure 3. Precision-Recall Curve and ROC Curve Comparing Classifiers Using Random Forests as Machine Learning Technique

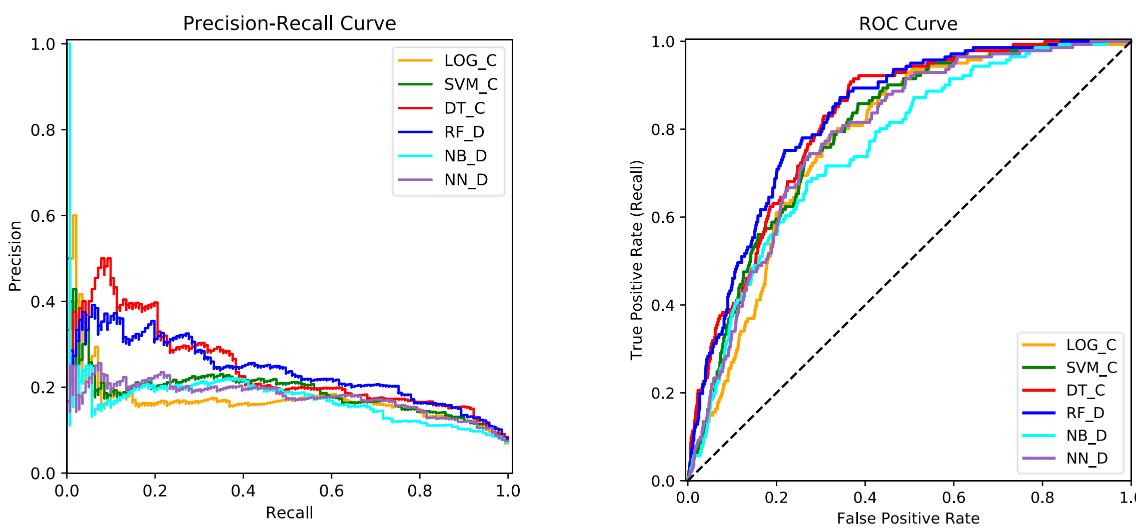

Figure 4: Precision-Recall Curve and ROC Curve Comparing the Best Classifiers for Each Machine Learning Technique According to AUC Score

Table 12. AUC Scores for All Classifiers and Machine Learning Techniques

	Classifier			
Techn.	**A**	**B**	**C**	**D**
LOG	70.79%	70.43%	77.58%	77.44%
SVM	71.57%	71.70%	79.42%	79.04%
DT	74.25%	73.45%	81.97%	81.33%
RF	73.99%	73.73%	82.41%	82.83%
NB	68.85%	69.16%	75.62%	75.76%
ANN	71.87%	71.16%	77.39%	78.47%

The results for the additional classifiers E and F, which consider regulator-confirmed information alone and user- and regulator-confirmed information combined (without self-disclosed information), respectively, are presented in Tables D2-D4 in the Appendix. Applying the models to the naturally distributed sample, Classifiers E and F also significantly outperform a naïve classification algorithm, thus supporting that regulator-confirmed information is valuable. Nevertheless, Classifier D significantly outperforms Classifiers E and F for, respectively, five and four machine learning techniques (See Table D5 in the Appendix), thus confirming the value of self-disclosed information for detecting financial intermediary misconduct (H1).

While the results for Classifiers A to D reported in Tables 7, 8, and 10 each represent one specific parameter configuration with optimized classification thresholds for each machine learning technique, Figure 3, as well as Figures D1 and D2 in the Appendix, present the precision-recall curve together with the ROC curve, illustrating the performance of our classifiers for different classification thresholds, as described in Section 3.7. Looking at the precision-recall curves, all machine learning techniques illustrate similar patterns for Classifiers A through D. While precision declines with increasing recall, the precision for higher recall scores (above 50%) ranges between 10% and 20% for all machine learning techniques. Considering the ROC curves, we find similar patterns regarding the relation of recall and the false positive rate for all classifiers.

Comparing the different classifiers based on both curves, we do not observe any clear differences between Classifiers A and B or between Classifiers C and D. In contrast, we identify meaningful differences between Classifiers A and C, A and D, B and C, and B and D. This provides further evidence that classifiers including regulator-confirmed information improve classification performance (H3), whereas we do not find any clear evidence that user-confirmed information is valuable (H2). These results are supported by the AUC scores for the ROC curves presented in Table 12.

Although Classifiers C and D exhibit comparable AUC scores, Classifier D, based on RF, yields the overall best performance with a score of 82.83%. Figure 4 compares the best classifiers for each machine learning technique according to AUC score. The charts show that Classifier D using RF dominantly outperforms all other classifiers. This holds for both the precision-recall curve as well as the ROC curve and provides further support that Classifier D using RF yields the overall best performance. This result again supports that for classifiers using regulator-confirmed as well as user-confirmed information, user-confirmed information adds value to the classification.

4.3 Economic Evaluation Based on the Naturally Distributed Sample

For our economic evaluation, as outlined in Section 3.7, we rely on the classifier results for the naturally distributed sample. Table 13 shows the average economic gain for each classifier. For brevity, we only report the values of RF, since it is best-performing machine learning technique.

As Table 13 indicates, all classifiers add economic value since the average economic gain (based on actual compensations and search costs of zero) is positive and significantly different from zero. Although there is a risk of misclassifying financial intermediaries, the resulting economic gains from using the classifiers more than compensate for losses caused by incorrect classification. Classifier A yields an average economic gain of USD 4,525.45. Consequently, classifiers based on self-disclosed information of financial intermediaries are economically valuable (H1). In accordance with the classifier evaluation, Classifier B yields a slightly lower average economic gain of USD 4,280.75. Therefore, classifiers that also consider user-confirmed information do not outperform classifiers that only take unverified self-disclosed information into account, which does not support H2. Classifiers C and D lead to higher economic gains compared to Classifier A, yielding USD 4,695.47 and USD 4,711.58, respectively. Thus, Classifier D achieves the highest economic value. Thus, our results indicate that classifiers detecting misconduct based on self-disclosed information in combination with user- and regulator-confirmed information are economically valuable, supporting H3. Nevertheless, Classifiers C and D do not significantly outperform Classifiers A and B from an economic point of view.

Table 13. Economic Evaluation of the Best-Performing Machine Learning Techniques per Classifier (Naturally Distributed Sample)

Technique	Classifier	Average economic gain (in USD)	WRS test vs. Naive	WRS test vs. A	WRS test vs. B
RF	A	4,525.45	0.00***	-	-
	B	4,280.75	0.00***	0.90	-
	C	4,695.47	0.00***	0.91	0.99
	D	4,711.58	0.00***	1.00	0.90
Note: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.					

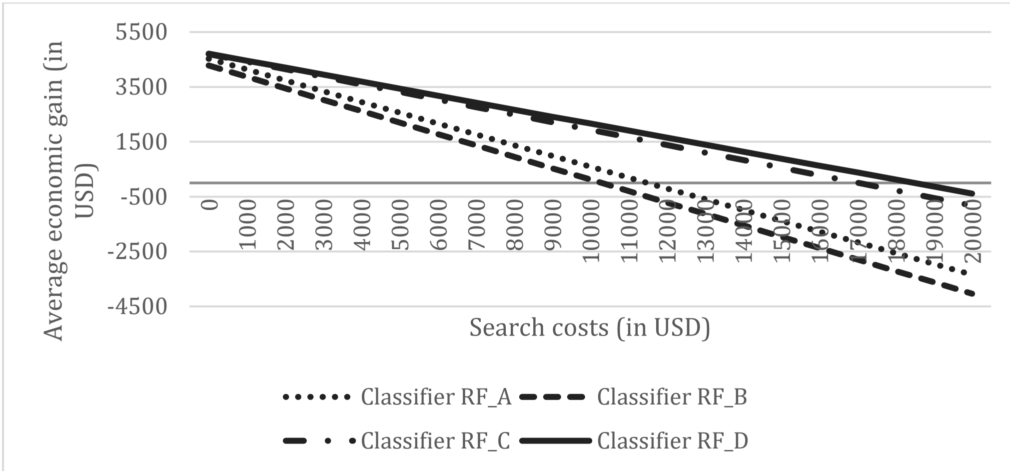

Figure 5. Sensitivity Analysis of Different Levels of Search Costs for Classifiers A to D

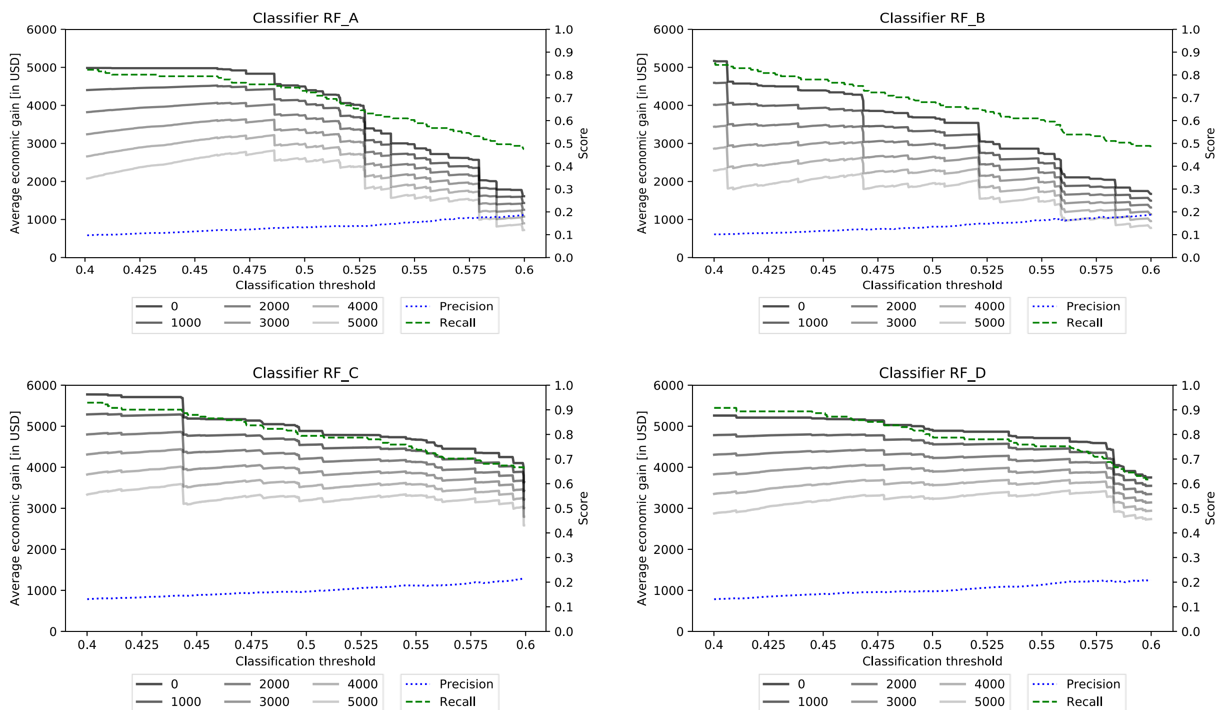

Figure 6. Sensitivity Analysis of Economic Gain for Varying Search Costs and Classification Thresholds

In addition to the economic evaluation of the optimized configuration of the classifiers, we perform a sensitivity analysis to evaluate how varying levels of search costs and different classification thresholds influence the economic gain of the proposed classifiers. Starting with varying search costs, Figure 5 shows how the economic value of the different classifiers (again based on RF) declines with rising search costs when keeping the classification threshold constant. Search costs represent the costs incurred by investors when searching for a new intermediary, which becomes necessary whenever the model classifies a broker as likely to commit misconduct. Classifier D leads to the highest average economic gain across all levels of search costs. Moreover, Classifier D even adds economic value for search costs of up to USD 18,477, which is far above realistic costs (Egan, 2019; Hortaçsu & Syverson, 2004). For realistic levels of search costs,[6] all classifiers add value.

The classification threshold also impacts the economic gain of the classifiers due to the number of FN and FP, which lead to economic losses in terms of

[6] For example, Egan (2019) reports search costs of USD 150 per USD 10,000 investment for the median investor.

compensation payments and search costs. Therefore, the optimal classification threshold leading to the maximum economic gain varies for different levels of search costs. Figure 6 shows the results of our sensitivity analysis regarding varying classification thresholds and search costs. The left y-axis shows the average economic gain in USD for the same classifier with varying classification thresholds (x-axis) and search costs (zero to USD 5,000). The right y-axis shows the value of the evaluation scores (precision or recall) for the respective classification threshold. We vary the thresholds between 0.40 and 0.60 as this leads to meaningful recall and precision. While high recall and the avoidance of FN is the primary goal when compensation payments are high and search costs are low, precision and the avoidance of FP becomes economically more important in the context of higher search costs.

The results show that the positive economic gain of our classifiers is robust across a wide range of different classification thresholds and varying levels of search costs. However, the optimal classification threshold, which leads to the highest possible economic gain yielded by a specific classifier, depends on the amount of search cost being considered. Specifically, the higher the assumed search costs, the higher the optimal classification threshold. This can be explained by the fact that a higher classification threshold increases precision (sacrificing recall at the same time) and thus reduces the number of FP, which should particularly be avoided in the context of high search costs. Nevertheless, for higher classification thresholds, the gain in precision does not compensate for the loss in recall and thus causes lower economic gains. This is especially true for Classifiers A and B and must be considered when choosing the desired classification threshold. While we observe more jagged lines for Classifiers A and B, the lines for Classifiers C and D become smoother. The existence of jagged lines and the smoothing effect observed from Classifiers A to D can be explained by the tree-structure of random forests: more features lead to more splits, and therefore generate more leaf nodes, representing a more granular assignment of class probabilities. This decreases the effect of marginal changes in classification thresholds on economic gains. Consequently, regarding the economic gain of the proposed classifiers, the sensitivity analysis indicates that investors and other model users should customize the classifiers used based on their own individual search costs in real-world applications. In this context, classifiers with less jagged lines (Classifier D) are more practical since they allow classification thresholds to be chosen on a more continuous basis.

In summary, our economic evaluation shows that the developed classifiers provide economic value. Although the economic value of the classifiers is not significantly different, we find indications that a classifier based on externally verified information should be favored, compared to a classifier only taking unverified self-disclosed information into account.

4.4 Discussion

Based on information manipulation theory, we analyze whether self-disclosed information is valuable for detecting financial intermediaries who are likely to commit misconduct. Referring to information disclosure, our results show significant differences in means between financial intermediaries with and without misconduct histories, which thus supports information manipulation theory. Further, this self-disclosed information can be used to detect and address intermediary misconduct in financial markets. The approaches proposed in this paper achieve promising classification performance. The application of our proposed classifiers could yield considerable economic gains for society by preventing intermediary misconduct and thus strengthening trust in the financial system. In particular, the results provide evidence that self-disclosed information is valuable for detecting financial intermediary misconduct (H1).

Moreover, confirming warranting theory, our results show that self-disclosed information in combination with different levels of external verification is valuable to classify intermediaries that do versus those that do not commit misconduct. This is particularly true for self-disclosed information confirmed by regulatory authorities (H3) because regulatory verification is hard to manipulate. Information verified by other users, however, does not significantly increase classification performance (H2). While verifications by reliable third parties such as regulators provide the most value in our classification, verifications by third parties such as other users on LinkedIn may lead to moderate increases in classification performance in certain classifier configurations. Potentially, the performance of classifiers including user-confirmed information could be improved if social networks were to offer additional verification tools or incentivize users to provide more mutual verifications.

For all classifiers, our results show that RF is the most promising machine learning technique for this particular classification problem. Nevertheless, all other machine learning techniques also exhibit promising results for the detection of financial intermediary misconduct.

The results of the economic evaluation support that self-disclosed information of financial intermediaries significantly adds value to the detection of intermediary misconduct, compared to naïve detection approaches. Here, all classifiers show significant positive economic gains. When evaluating different levels of search costs, the classifiers that use self-disclosed information in combination with user- and regulator-confirmed information add value for search costs of up to USD 18,477, which represent costs far above the realistic level of costs associated with identifying a new broker or investment advisor even taking opportunity costs into account. Moreover, the sensitivity analysis conducted in the course of our economic evaluation shows how investors can customize the classifiers by setting the classification threshold based on their individual level of search costs. Thus, to answer our research question, classifiers that use self-disclosed information in combination with different levels of external verification are valuable for detecting financial intermediary misconduct.

In terms of research methodology, we used a balanced dataset for training to manage the problem of unequal class distribution of intermediaries with and without misconduct. Our results highlight that the proposed classifiers outperform a naive classification algorithm, as described in Section 4.2.2. In light of the historical unbalanced class distribution (only 6.83% of the intermediaries actually commit misconduct), a simple classification model would achieve a high level of accuracy by classifying each intermediary as unlikely to commit misconduct. However, recall of such a simple classifier would be zero. Because investors incur immense losses in cases of misconduct, it is essential to identify as many intermediaries that are likely to commit misconduct as practically possible, even if thereby some reliable intermediaries are incorrectly classified and investigated. In the field of automated misconduct detection, recall is, therefore, more important than precision and accuracy. From this perspective, and compared to the naïve approach, our results demonstrate high levels of recall for the proposed classifiers.

Furthermore, a useful classifier for investors should also maximize the negative predictive value (true non-misconduct cases within the share of predicted non-misconduct cases) so that an investor who searches for a new intermediary can rely on the predictions of non-misconduct cases. This is even more important than achieving high precision scores because of the above-mentioned losses for investors when classifying true misconduct cases as non-misconduct cases. All of our classifiers fulfill these requirements and show high scores for negative predictive value (between 96% and 98%).

From an ethical point of view, our proposed classifiers do not suffer from biases against certain groups or minorities, as has been recently witnessed for machine learning algorithms applied to criminal sentencing in the US (Angwin, 2016) because we do not include features related to poverty, joblessness, or social marginalization. Moreover, the group of financial intermediaries targeted by our proposed machine learning algorithm is quite homogeneous in terms of education, job situation, and social environment. Furthermore, if a financial intermediary is falsely classified as misbehaving, this classification enables the regulator/supervisor to investigate the intermediary more closely, but classification does not directly lead to negative consequences or penalties.

We are aware of certain limitations of our study. To determine financial intermediaries who committed misconduct, we rely on disclosures provided by the regulatory authority FINRA's website BrokerCheck. However, intermediaries who committed misconduct in the past but whose actions remain unobserved by regulators and customers do not have a disclosure record for these unobserved cases. Nevertheless, as it is obligatory to disclose actions and consequences related to misconduct, BrokerCheck is the most comprehensive source for misconduct disclosures. Additionally, our proposed classifiers can easily be adapted by training on an updated dataset in case of a previously trustworthy intermediary being accused of misconduct.

Moreover, financial intermediaries might strategically change their behavior regarding self-disclosure of personal information on business networks in order to avoid being detected by our proposed mechanism once they are aware that such classifiers are in place. This issue typically exists in various applications of fraud detection. However, since our classifiers cover a wide range of features with different levels of external verification applying countermeasures or imitating trustworthy intermediaries is difficult. More importantly, Classifiers B, C, and D make use of features that incorporate whether self-disclosed information is confirmed or even deviates from externally confirmed information. Thus, if misbehaving intermediaries were to attempt to polish their profiles to avoid being detected, Classifiers B, C, and D would nevertheless be capable of identifying intermediary misconduct based on information that is externally verified by other users on LinkedIn or by the regulator, which is hard or even impossible to manipulate (in line with warranting theory). Consequently, our classifiers should also work in the long run. However, the proposed classifiers should regularly be retrained once they are put in place to cope with potential changes in the way people disclose personal information on business networks.

Finally, we are aware that our classifiers and features are based on data that is available on the business network LinkedIn. There are also other business networks such as Maimai, the largest professional social network in China, and Xing, a European competitor of LinkedIn, so one might argue that the results of this study might differ when taking other platforms into account. Nevertheless, as users provide very similar information on all such networks, the proposed classifiers can be easily applied to other similar platforms.

5 Conclusion

Financial intermediaries are essential for investors to participate in financial markets and exhibit a large influence on investors' financial performance, wealth, and life planning. Consequently, intermediaries play a crucial role in the financial system. Investors' trust in these intermediaries is a fundamental prerequisite to ensuring fair and efficient financial markets and capital provision by investors to corporations. Trust in intermediaries has become particularly important because of increased reliance on electronic communication and less personal interaction between investors and intermediaries, which may impede the trust building process. Therefore, misconduct by intermediaries needs to be detected and scandals like the *Wolf of Wall Street* must be avoided to protect investors from losses and to preserve trust in the financial system.

This paper contributes to the literature on financial misconduct and offers new insights to the scarce literature on automated detection of financial intermediary misconduct. Based on self-disclosed information provided on intermediaries' profiles on the business network LinkedIn, we can detect intermediaries who are likely to commit misconduct. The best performing classifier, combining self-disclosed information with externally verified information, is able to detect misconduct among financial intermediaries with a recall of 77.02% and an accuracy of 73.74% for the balanced training sample and a recall of 75.18% and an accuracy of 77.96% and for the naturally distributed validation sample in which intermediaries with a history of misconduct represent the minority class at 6.87%.

We also contribute to the literature on automated misconduct and fraud detection in general by highlighting the value of self-disclosed information in combination with different levels of external verification. We show that self-disclosed information differs between trustworthy financial intermediaries and those who have committed misconduct. Therefore, our results confirm information manipulation theory and provide evidence that self-disclosed information is useful for classifying whether individuals commit misconduct or not. Supporting warranting theory, we show that self-disclosed information that is verified by a third party and thus harder to manipulate, provides additional value for detecting misconduct. Thereby, we show that verifications by reliable third parties such as regulators provide most value for the classification.

From a practical perspective, our results are relevant for investors and regulators alike. The economic evaluation of the classifiers confirms a significant economic value for investors given realistic levels of search costs. Furthermore, our sensitivity analysis shows investors how to customize the classifiers and optimize the classification results according to their individual level of search costs. By using the classifiers, investors will be less likely to be damaged by intermediary misconduct and can thereby avoid severe losses. Moreover, our classifier that is based on self-disclosed information alone provides sufficient classification accuracy and economic value to be useful for investors in countries where no regulatory data regarding financial intermediaries is publicly available.

Furthermore, the proposed approach allows regulators/supervisors to engage in predictive supervision. Thereby, they can efficiently allocate resources to review those intermediaries more closely that are classified by the system as likely to commit misconduct. Consequently, predictive supervision based on our approach enables authorities to detect potential misconduct earlier and may therefore help prevent the next *Wolf of Wall Street*. As such, the proposed classifiers can facilitate investor protection against financial intermediary misconduct, which would increase trust in the financial system and would therefore be valuable for the society as a whole.

The analysis of self-disclosed information using different levels of external verification can also be valuable for fraud detection in other fields, thus providing future research opportunities. For example, self-disclosed information on business network profiles could be valuable for corporate compliance departments, who could use this information to supplement their own, verified data. Moreover, the significance of self-disclosed information on social media networks such as Facebook that disclose private rather than job-related information could also be investigated in this vein. However, the substantial difference between information provided to friends only and information that is provided publicly would have to be accounted for in such an analysis. Future research might also investigate the possibility of developing classifiers to detect specific types of misconduct committed by intermediaries or to identify intermediaries implicated in multiple misconduct events. Our results show that analytics and machine learning techniques combined with the massive and ever-increasing amount of self-disclosed information available in social networks provide powerful tools for finding solutions to important societal challenges.

References

Adikari, S., & Kaushik, D. (2014). Identifying Fake Profiles in LinkedIn. *Proceedings of the 19th Pacific Asia Conference on Information Systems*.

Allen, F., & Santomero, A. M. (1997). The theory of financial intermediation. *Journal of Banking & Finance, 21*(11-12), 1461-1485.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias: There’s software used across the country to predict future criminals. And it’s biased against blacks. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Ba, S., & Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: price premiums and buyer behavior. *MIS Quarterly, 26*(3), 243-268.

Bazarova, N. N., & Choi, Y. H. (2014). Self-Disclosure in Social Media: Extending the Functional Approach to Disclosure Motivations and Characteristics on Social Network Sites. *Journal of Communication, 64*(4), 635-357.

Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems, 50*(3), 602-613.

Bloomberg. (2018). “Wolf of Wall Street” Jordan Belfort isn’t paying his debts, U.S. says. https://www.bloomberg.com/news/articles/2018-05-16/-wolf-of-wall-street-belfort-isn-t-paying-his-debts-u-s-says

Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science, 17*(3), 235-249.

Breiman, L. (1996). Bagging predictors. M*achine Learning, 24*(2), 123-140.

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory, 6*(3), 203-242.

Castells, M. (2010). *The rise of the network society* (2nd ed.). Wiley-Blackwell.

Cataldo, A. J., & Killough, L. N. (2003). Market makers’ methods of stock manipulation. management *Accounting Quarterly, 4*(4), 10-13.

Chawla, N. V. (2009). Data mining for imbalanced datasets: An overview. In O. Maimon & L. Rokach (Eds.), *Data mining and knowledge discovery handbook* (pp. 875-886) Springer.

Council of Economic Advisors. (2015). *The effects of conflicted investment advice on retirements savings.* https://permanent.access.gpo.gov/gpo55500/cea_coi_report_final.pdf

Crowston, K., & Fagnot, I. (2018). Stages of motivation for contributing user-generated content: A theory and empirical test. *International Journal of Human-Computer Studies*, *109*, 89-101.

Cumming, D., & Johan, S. (2008). Global market surveillance. *American Law and Economics Review*, *10*(2), 454-506.

Cumming, D., Johan, S., & Li, D. (2011). Exchange trading rules and stock market liquidity. *Journal of Financial Economics*, *99*(3), 651-671.

Dong, W., Liao, S., & Zhang, Z. (2018). Leveraging financial social media data for corporate fraud detection. *Journal of Management Information Systems, 35*(2), 461-487.

Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern classification* (2nd ed.). Wiley-Interscience.

Dupret, G., & Koda, M. (2001). Bootstrap re-sampling for unbalanced data in supervised learning. *European Journal of Operational Research, 134*(1), 141-156.

Dyck, A., Morse, A., & Zingales, L. (2010). Who blows the whistle on corporate fraud? *The Journal of Finance, 65*(6), 2213-2253.

Eccles, R. G., & Crane, D. B. (1987). Managing through networks in investment banking. *California Management Review*, *30*(1), 176-195.

Egan, M. (2019). Brokers versus retail investors: Conflicting interests and dominated products. *The Journal of Finance, 74*(3), 1217-1260.

Egan, M., Matvos, G., & Seru, A. (2019). The market for financial adviser misconduct. *Journal of Political Economy*, *127*(1), 233-295.

Everitt, B. S. (1977). *The analysis of contingency tables.* Chapman and Hall.

Fanning, K. M., & Cogger, K. O. (1998). Neural network detection of management fraud using published financial data. *International Journal of Intelligent Systems in Accounting, Finance & Management, 7*(1), 21-41.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine, 17*(3), 37-54.

Glancy, F. H., & Yadav, S. B. (2011). A computational model for financial reporting fraud detection. *Decision Support Systems, 50*(3), 595-601.

Groth, S. S., Siering, M., & Gomber, P. (2014). How to enable automated trading engines to cope with news-related liquidity shocks? Extracting signals from unstructured data. *Decision Support Systems, 62*, 32-42.

Han, J., & Kamber, M. (2006). *Data mining: Concepts and techniques* (2nd ed.). Elsevier.

Hazen, T. L. (2010). Are existing stock broker standards sufficient? Principles, rules, and fiduciary duties. *Columbia Business Law Review, 2010*(3), 710-762.

Hollenbaugh, E. E., & Ferris, A. L. (2014). Facebook self-disclosure: Examining the role of traits, social cohesion, and motives. *Computers in Human Behavior*, *30*, 50-58.

Hortaçsu, A., & Syverson, C. (2004). Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds. *The Quarterly Journal of Economics*, *119*(2), 403-456.

Humpherys, S. L., Moffitt, K. C., Burns, M. B., Burgoon, J. K., & Felix, W. F. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems, 50*(3), 585-594.

Jain, B. A., & Nag, B. N. (1997). Performance evaluation of neural network decision models. *Journal of Management Information Systems, 14*(2), 201-216.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017*). An introduction to statistical learning: With applications in R* (Corrected at 8th printing). Springer.

Japkowicz, N. (2000). The class imbalance problem: Significance and strategies. In *Proceedings of the International Conference on Artificial Intelligence*.

Jourard, S. M. (1971). *Self-disclosure: An experimental analysis of the transparent self.* Wiley-Interscience.

Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications, 32*(4), 995-1003.

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*.

Krasnova, H., Spiekermann, S., Koroleva, K., & Hildebrand, T. (2010). Online social networks: Why we disclose. Journal of Information *Technology, 25*(2), 109-125.

Kurgan, L. A., & Musilek, P. (2006). A survey of Knowledge Discovery and Data Mining process models. *The Knowledge Engineering Review*, *21*(1), 1-24.

Lazaro, C. (2014). Has expungement broken Brokercheck. *Journal of Business and Securities Law*, *14*(2), 125-150.

Lee, T. A., Ingram, R. W., & Howard, T. P. (1999). The difference between earnings and operating cash flow as an indicator of financial reporting fraud. *Contemporary Accounting Research, 16*(4), 749-786.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics, 45*(2), 221-247.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance, 66*(1), 35-65.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*.

McCann, C., Qin, C., & Yan, M. (2017). How widespread and predictable is stock broker misconduct? *The Journal of Investing*, *26*(2), 6-25.

McCornack, S. A. (1992). Information manipulation theory. *Communication Monographs, 59*(1), 1-16.

Menzies, T., Dekhtyar, A., Distefano, J., & Greenwald, J. (2007). Problems with precision: A Response to "Comments on 'Data Mining Static Code Attributes to Learn Defect Predictors.'" *IEEE Transactions on Software Engineering*, *33*(9), 637-640.

Ngai, E.W.T., Hu, Y., Wong, Y. H., 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.

Palazzo, G., & Rethel, L. (2008). Conflicts of interest in financial intermediation. *Journal of Business Ethics, 81*(1), 193-207.

Persons, O. S. (1995). Using financial statement data to identify factors associated with fraudulent financial reporting. *Journal of Applied Business Research, 11*(3), 38-46.

Qureshi, H. & Sokobin, J. (2015). Do investors have valuable information about brokers? (Working paper, FINRA Office of the Chief Economist).

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS One, 10*(3), e0118432.

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.

Siering, M., Koch, J.-A., & 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.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427-437.

Tan, M., Tan, L., Dara, S., & Mayeux, C. (2015). Online defect prediction for imbalanced data. *Proceedings of the ACM International Conference on Mobile Software Engineering and Systems*.

Vapnik, V. N. (1998). *Statistical learning theory*. Wiley.

Walther, J. B., van der Heide, B., Hamel, L. M., & Shulman, H. C. (2009). Self-generated versus other-generated statements and impressions in computer-mediated communication. *Communication Research, 36*(2), 229-253.

Wang, S. (2010). A comprehensive survey of data mining-based accounting-fraud detection research. *Proceedings of the International Conference on Intelligent Computation Technology and Automation*.

West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. *Computers & Security, 57*, 47-66.

Wooldridge, J. M. (2009). *Introductory econometrics: A modern approach* (4th ed., International Student Edition). Cengage Learning.

Zhang, J., Bloedorn, E., Rosen, L., & Venese, D. (2004). Learning rules from highly unbalanced data sets. *Proceedings of the Fourth IEEE International Conference on Data Mining*.

Zhang, J., & Mani, I. (2003). *kNN approach to unbalanced data distributions: A case study involving information extraction.* Paper presented at the ICML Workshop on Learning from Imbalanced Datasets II. https://www.site.uottawa.ca/~nat/Workshop2003/jzhang.pdf

Zhou, L., Burgoon, J. K., Nunamaker, J. F., & Twitchell, D. (2004). Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications. *Group Decision and Negotiation, 13*(1), 81-106.

Appendix A: Descriptive Statistics for Brokers with Misconduct Cases

Table A1. Descriptive Statistics for Brokers with Misconduct Cases

N = 541									
Feature	**Sum**	**Mean**	**SD**	**Min**	**25%**	**50%**	**75%**	**95%**	**Max**
Disclosures	1,257	2.32	1.97	1.00	1.00	2.00	3.00	6.00	21.00
Customer disputes	983	1.82	1.66	0.00	1.00	1.00	2.00	5.00	21.00
Regulatory actions	109	0.20	0.57	0.00	0.00	0.00	0.00	1.00	5.00
Non job-related disclosures*	165	0.30	0.84	1.00	1.00	1.00	2.00	3.00	9.00
Final customer disputes and regulatory actions**	804	1.49	1.06	0.00	0.00	0.00	0.00	2.00	8.00
Average damage amount requested***	-	349	1,597	0.00	5.00	32.63	197	1,000	20,000
Total settlement amount***	156,622	290	1,884	0.00	7.50	27.00	112	700	38,554
Total amount of fines***	931	1.72	21.11	0.00	0.00	0.00	0.00	3.00	482

Note: *Other mandatory disclosures that are not directly linked to a broker's professional activity (e.g., regarding default in the broker's personal financial situation or criminal tasks like assault or theft).

**Disclosures with a final status as described in Section 3.2.

*** In USD 1,000; based on final customer disputes and regulatory actions, respectively.

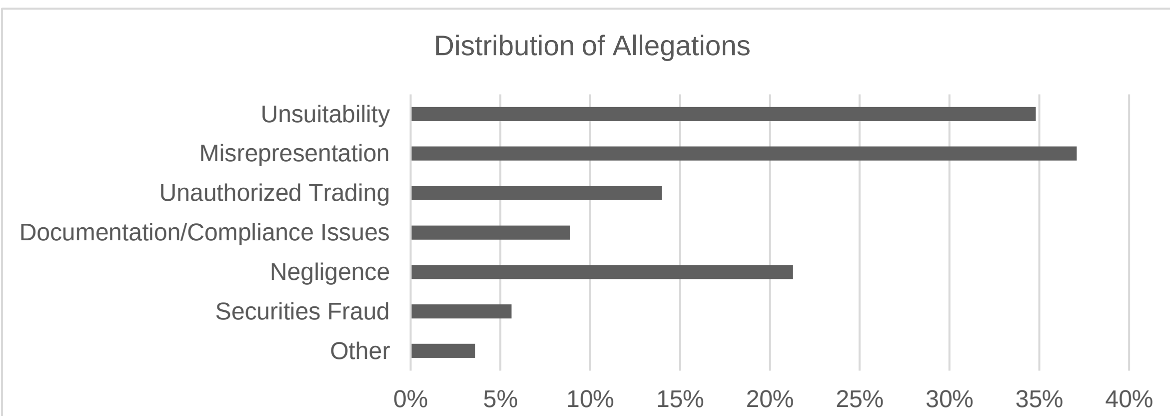

Figure A1: Distribution of allegations among all final customer disputes and regulatory actions

Figure A1 shows the distribution of allegations (types of misconduct) among all observed misconduct cases in our dataset. One misconduct case can have multiple allegations, e.g., unsuitability and misrepresentation, therefore the overall sum is not equal to 100%. Allegations are categorized by the most prevalent misconduct categories on BrokerCheck. The categories cover the following respective allegations: (1) unsuitability: investment advice unsuitable to the customer's preferences; (2) misrepresentation: active misrepresentation or disguise of facts regarding the nature, risks, or fees of a financial product, unauthorized trading; (3) trading without clients' permission; (4) documentation/compliance issues: practicing without a license, failure to document undertaken businesses properly, failure to complete mandatory reporting; (5) negligence: failure to execute orders/liquidate assets, failure to maintain/supervise portfolio properly, failure to follow customer's instructions properly; (6) securities fraud: gambling with customers' assets, excessive trading, churning, front running, scalping.

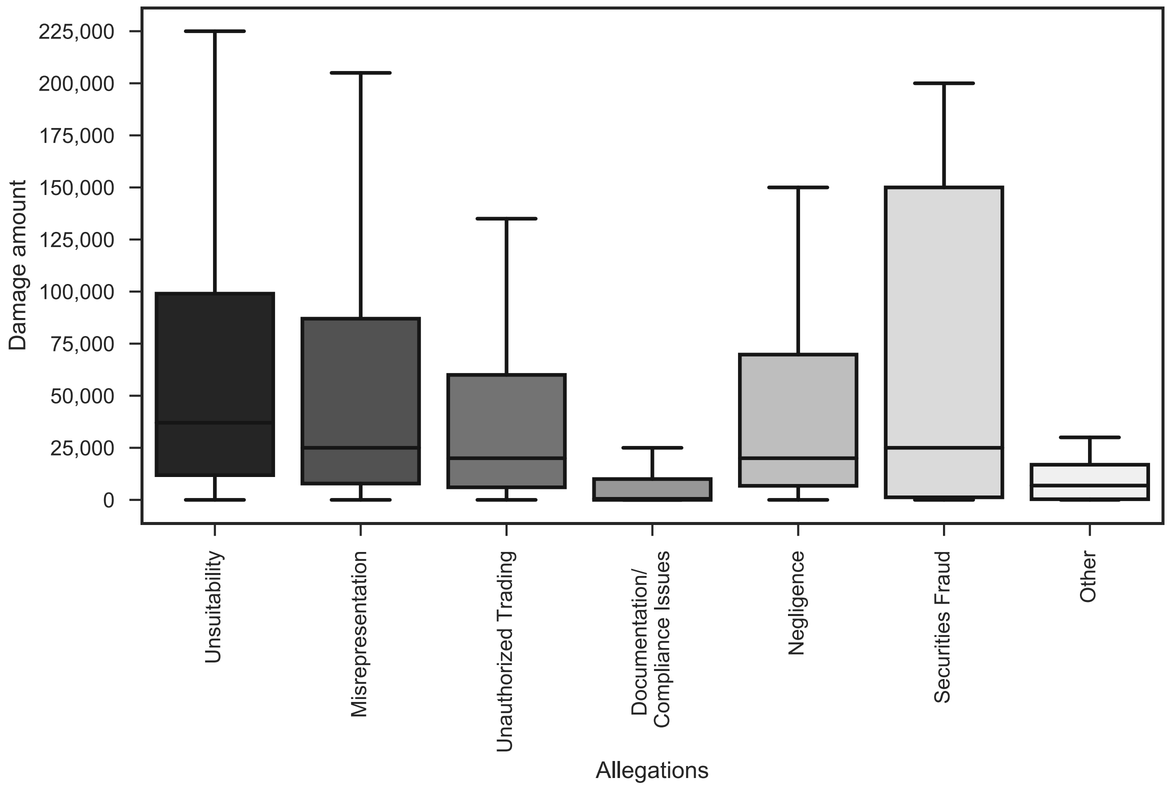

Figure A2: Distributions of Damage Amounts per Case by Allegation Category

Figure A2 provides boxplots to reveal the distributions of damage amounts (in USD) per case (settlement amount or fine) by allegation category (type of misconduct) among all observed misconduct cases in our dataset. Allegations are categorized by the most prevalent misconduct categories on BrokerCheck. The categories cover the following respective allegations: (1) unsuitability: investment advice unsuitable to the customer's preferences; (2) misrepresentation: active misrepresentation or disguise of facts regarding the nature, risks, or fees of a financial product; (3) unauthorized trading: trading without permission of the client; (4) documentation/compliance issues: practicing without a license, failure to document undertaken businesses properly, failure to complete mandatory reporting; (5) negligence: failure to execute orders/liquidate assets, failure to maintain/supervise portfolio properly, failure to follow customer's instructions properly; (6) securities fraud: gambling with customers' assets, excessive trading, churning, front running, scalping.

Appendix B: Hyperparameter Tuning

Table B1. Tuned Parameters, Parameter Grid, and Configuration for the Best Classifiers (According to the AUC Score) for Each Applied Machine Learning Technique

Techn.	Parameter	Description	Grid	Best
LOG	Solver	Algorithm for optimization	saga, liblinear	liblinear
	Penalty	Norm used for regularization	l1, l2	l1
	C	Inverse of regularization strength	$10^x, x \in [-5, 6]$	10^{-1}
	#Estimators	Number of estimators for bagging	[10, 20, 30, …, 500]	200
SVM	Kernel	Used kernel type	linear, poly, rbf, sigmoid	linear
	C	Penalty parameter	$10^x, x \in [-5, 6]$	10^{-1}
	Shrinking	Usage of shrinking heuristic	True, False	True
	#Estimators	Number of estimators for bagging	[10, 20, 30, …, 500]	20
DT*	Criterion	Function to measure quality of fit	gini, entropy	gini
	Min samples split	Min. samples required for a split	[2, 100]	12
	Max. depth	Maximum depth of a tree	[10, 100]	70
	Min. samples leaf	Min. samples required for leaf nodes	[1, 20]	1
	#Estimators	Number of estimators for bagging	[10, 20, 30, …, 500]	200
RF*	Criterion	Function to measure quality of fit	gini, entropy	entropy
	Min. samples split	Min. samples required for a split	[2, 100]	12
	Max. depth	Maximum depth of a tree	[1, 100]	71
	Min samples leaf	Min. samples required for leaf nodes	[1, 20]	1
	#Estimators	Number of trees in the forest	[10, 20, 30, …, 500]	200
NB	Variance smoothing	Portion of largest variance of all features added to variance for calculation stability	$10^x, x \in [-11, -1]$	10^{-2}
	#Estimators	Number of estimators for bagging	[10, 20, 30, …, 500]	80
ANN	Activation function	Activation function for hidden layer	tanh, relu	tanh
	#Hidden layers	Number of hidden layers	1, 2, 4, 8	2
	#Nodes in layer	Number of nodes in each layer	4, 8, 16, 32, 64, 128, 256	8
	Learning rate	Initial learning rate	$10^x, x \in [-5, -3]$	10^{-3}
	l2 regularizer	Penalty parameter	$[10^x, 0], x \in [-3, -1]$	0
	Solver	Solver for weight optimization	adam, sgd	adam
	Epochs	Number of epochs	50, 100, 200	100
	Batch size	Size of batches	20, 50, 100	100

Note: *We apply pruning for decision tree and random forest models to prevent overfitting and to increase computational efficiency (Duda et al., 2012; Han & Kamber, 2006).

Appendix C: Descriptive Statistics for the Naturally Distributed Sample

Table C1. Descriptive statistics for the Naturally Distributed Sample and the Wilcoxon Rank-Sum Test for Equality of Means

	Full dataset N = 2,051				Misconduct N = 141 (6.87%)				No Misconduct N = 1,910 (93.13%)				WRS test
Feature	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**P**
Self-disclosed													
li_male	0.0	1.0	0.7	0.5	0.0	1.0	0.9	0.3	0.0	1.0	0.7	0.5	0.00***
li_ picture	0.0	1.0	0.5	0.5	0.0	1.0	0.5	0.5	0.0	1.0	0.5	0.5	0.73
li_ Interests	0.0	603.0	13.6	26.4	0.0	108.0	10.4	14.5	0.0	603.0	13.9	27.1	0.05*
li_location	0.0	1.0	0.7	0.4	0.0	1.0	0.7	0.5	0.0	1.0	0.8	0.4	0.50
li_ connections	0	500	290	188	1	500	255	193	0	500	292	188	0.05**
li_ follower	0	10509	170	506	0	2105	181	373	0	10509	170	515	0.08*
li_ posts	0.0	50.0	7.2	16.0	0.0	50.0	9.5	17.8	0.0	50.0	7.0	15.8	0.13
li_ rec_gi	0.0	13.0	0.3	1.0	0.0	3.0	0.1	0.5	0.0	13.0	0.3	1.0	0.30
li_job_adv	0.0	1.0	0.4	0.5	0.0	1.0	0.3	0.5	0.0	1.0	0.4	0.5	0.49
li_job_vp	0.0	1.0	0.3	0.5	0.0	1.0	0.3	0.5	0.0	1.0	0.3	0.5	0.71
li_job_pres	0.0	1.0	0.2	0.4	0.0	1.0	0.3	0.4	0.0	1.0	0.2	0.4	0.15
li_job_sen	0.0	1.0	0.1	0.3	0.0	1.0	0.1	0.2	0.0	1.0	0.1	0.3	0.28
li_company_larbank	0.0	1.0	0.3	0.4	0.0	1.0	0.3	0.4	0.0	1.0	0.3	0.4	0.72
li_company_bank	0.0	1.0	0.1	0.3	0.0	1.0	0.1	0.2	0.0	1.0	0.1	0.3	0.55
li_company_inde	0.0	1.0	0.4	0.5	0.0	1.0	0.6	0.5	0.0	1.0	0.4	0.5	0.00***
li_company_insur	0.0	1.0	0.1	0.3	0.0	1.0	0.1	0.2	0.0	1.0	0.1	0.3	0.18
li_company_am	0.0	1.0	0.1	0.3	0.0	1.0	0.0	0.2	0.0	1.0	0.1	0.3	0.27
li_jobs	0.0	12.0	2.9	2.0	0.0	8.0	2.3	1.6	0.0	12.0	2.9	2.0	0.00***
li_empl_details	0.0	1.0	0.9	0.2	0.0	1.0	1.0	0.2	0.0	1.0	0.9	0.2	0.45
li_avg_empl_dur	3.0	658.0	105.5	87.6	10.0	658.0	134.7	98.2	3.0	561.0	103.4	86.4	0.00***
li_cur_empl_details	0.0	1.0	0.9	0.3	0.0	1.0	0.9	0.3	0.0	1.0	0.9	0.3	0.92
li_cur_empl_dur	0.0	658.0	115.1	105.9	0.0	658.0	140.4	121.1	0.0	561.0	113.1	104.4	0.21
li_uni_ba	0.0	1.0	0.7	0.5	0.0	1.0	0.6	0.5	0.0	1.0	0.7	0.5	0.55
li_uni_ma	0.0	1.0	0.2	0.4	0.0	1.0	0.2	0.4	0.0	1.0	0.2	0.4	0.23
li_uni	0.0	1.0	0.8	0.4	0.0	1.0	0.8	0.4	0.0	1.0	0.8	0.4	0.78
li_uni_related	0.0	1.0	0.6	0.5	0.0	1.0	0.5	0.5	0.0	1.0	0.6	0.5	0.03**
li_ cert	0.0	12.0	0.7	1.4	0.0	8.0	0.9	1.6	0.0	12.0	0.7	1.4	0.71
li_ awards	0.0	15.0	0.2	1.0	0.0	9.0	0.2	1.1	0.0	15.0	0.2	1.0	0.64
li_ skill	0.0	50.0	10.5	11.9	0.0	50.0	8.2	11.0	0.0	50.0	10.7	12.0	0.01**
li_sum	0.0	1.0	0.5	0.5	0.0	1.0	0.7	0.5	0.0	1.0	0.5	0.5	0.00***
li_sum_words	0.0	333.0	47.5	73.8	0.0	299.0	59.3	75.4	0.0	333.0	46.7	73.6	0.00***
li_ sum_neg_words	0.0%	11.1%	0.2%	0.8%	0.0%	8.0%	0.2%	0.8%	0.0%	11.1%	0.2%	0.8%	0.55
li_ sum_pos_words	0.0%	100.0%	1.7%	3.5%	0.0%	12.0%	1.7%	2.5%	0.0%	100.0%	1.7%	3.6%	0.19
li_ sum_str_words	0.0%	33.3%	2.8%	4.5%	0.0%	14.3%	2.9%	3.3%	0.0%	33.3%	2.8%	4.5%	0.04**
li_ sum_compl_words	0.0%	100.0%	14.2%	16.7%	0.0%	66.7%	17.7%	14.5%	0.0%	100.0%	14.0%	16.9%	0.00***

li_ sum_emtl_words	0.0%	11.1%	0.1%	0.5%	0.0%	3.4%	0.1%	0.6%	0.0%	11.1%	0.1%	0.5%	0.74
li_ sum_uncert_words	0.0%	14.3%	0.3%	0.8%	0.0%	14.3%	0.6%	1.6%	0.0%	10.0%	0.2%	0.7%	0.01***
li_ sum_modal_words	0.0%	10.5%	0.3%	0.8%	0.0%	6.9%	0.6%	1.2%	0.0%	10.5%	0.3%	0.8%	0.00***
li_ sum_wps	0.0	100.0	8.8	10.8	0.0	42.0	12.8	10.9	0.0	100.0	8.5	10.7	0.00***
li_ sum_fog	0.0	48.4	9.2	10.1	0.0	30.3	12.2	9.0	0.0	48.4	9.0	10.1	0.00***
li_ sum_sen	-1.0	1.0	0.3	0.5	-1.0	1.0	0.4	0.5	-1.0	1.0	0.3	0.5	0.02**
User confirmed													
li_ rec_ob	0.0	60.0	0.2	1.6	0.0	2.0	0.0	0.2	0.0	60.0	0.2	1.7	0.25
li_end_skill	0.0	40.0	2.5	4.3	0.0	15.2	1.3	2.9	0.0	40.0	2.6	4.4	0.00***
Regulator confirmed													
bc_ia	0.0	1.0	0.5	0.5	0.0	1.0	0.9	0.4	0.0	1.0	0.5	0.5	0.00***
bc_avg_empl_dur	0.5	588.0	83.4	70.2	9.0	444.0	104.7	87.4	0.5	588.0	81.9	68.5	0.00***
bc_jobs	1.0	22.0	3.5	2.6	1.0	16.0	4.2	2.7	1.0	22.0	3.5	2.5	0.00***
bc_exams	1.0	9.0	4.1	1.4	2.0	8.0	4.5	1.5	1.0	9.0	4.1	1.4	0.00***
bc_licenses	0.0	60.0	14.5	17.2	0.0	60.0	18.7	13.2	0.0	55.0	14.2	17.4	0.00***
bc_li_exp_dev	0.0	475.0	79.5	85.7	0.0	402.0	72.9	92.1	0.0	475.0	80.0	85.2	0.00***
bc_li_jobs_dev	0.0	19.0	2.0	2.1	0.0	13.0	2.5	2.5	0.0	19.0	2.0	2.1	0.01**

Note: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Appendix D. Classifier Evaluation

D1. Graphical Analysis

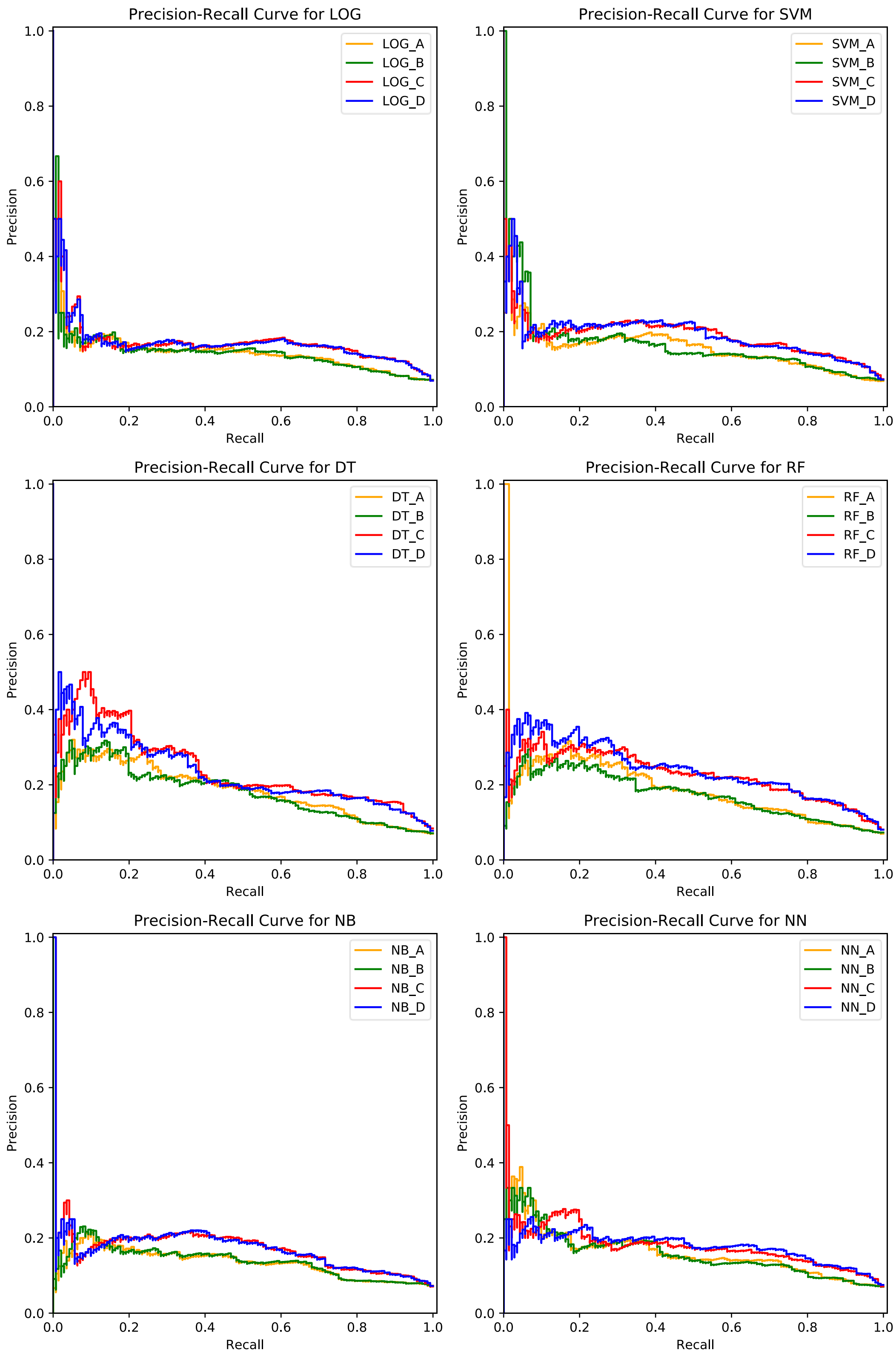

Figure D1: Precision-Recall Curve for All Classifiers and Machine Learning Techniques

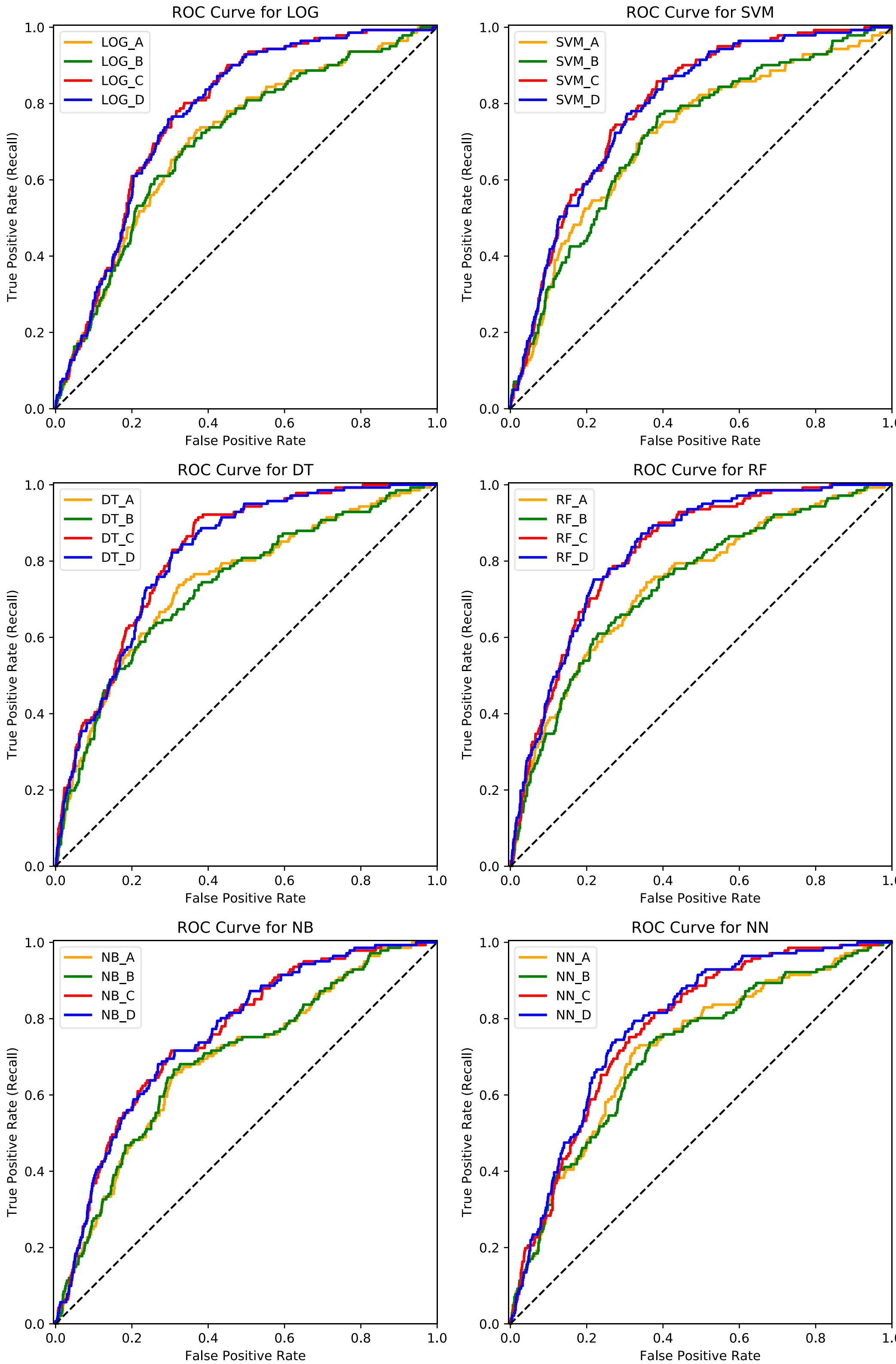

Figure D2. ROC Curves for All Classifiers and Machine Learning Techniques

D2. Classifier Evaluation for the Additional Classifiers

Table D1. Additional Classifiers

Classifier	Self-disclosed information	User-confirmed information	Regulator-confirmed information
E			x
F		x	x

Table D2: Classifier Evaluation for Classifiers Using User and Regulatory Confirmed Information Only in %, Training Results Based on the Balanced Sample)

	Classifier E					Classifier F				
Cues	**Regulator-confirmed information**					**User- and regulator-confirmed information**				
Tech.	**Acc.**	**Rec.**	**Prec.**	**Spec.**	**F1**	**Acc.**	**Rec.**	**Prec.**	**Spec.**	**F1**
LOG	68.94	76.66	66.34	61.31	71.00	69.07	76.46	66.55	61.72	71.05
SVM	66.92	82.70	62.71	51.39	71.24	67.42	82.95	63.17	52.14	71.64
DT	72.60	75.77	71.16	69.48	73.27	73.86	78.07	71.93	69.74	74.75
RF	74.49	78.19	73.07	70.83	75.36	74.24	77.48	73.06	71.03	75.01
NB	68.29	78.24	64.63	58.78	70.62	68.29	78.24	64.61	58.78	70.62
ANN	70.83	75.37	69.13	66.33	71.91	70.96	75.88	69.07	66.08	72.14

Table D3: Classifier Evaluation for Classifiers Using Regulator-Confirmed Information Only and McNemar's Test Results on Classifier Performance (in %, Naturally Distributed Sample

Classifier E							
Cues	**Regulator-confirmed information**					**McNemar's test**	
Tech.	**Acc.**	**Rec.**	**Prec.**	**Spec.**	**F1**	**E vs. Naive**	
LOG	66.21	75.89	13.97	65.50	23.59	0.00***	E > Naive
SVM	68.11	76.60	14.81	67.49	24.83	0.00***	E > Naive
DT	70.45	80.85	16.45	69.69	27.34	0.00***	E > Naive
RF	74.70	73.76	17.75	74.76	28.61	0.00***	E > Naive
NB	67.77	71.63	13.99	67.49	23.41	0.00***	E > Naive
ANN	71.14	71.63	15.47	71.10	25.44	0.00***	E > Naive

Table D4: Classifier Evaluation for Classifiers Using User and Regulator-Confirmed Information Only and McNemar's Test Results on Classifier Performance (in %, Naturally Distributed Sample)

Classifier F							
Cues	**User- and regulator-confirmed information**					**McNemar's test**	
Tech.	**Acc.**	**Rec.**	**Prec.**	**Spec.**	**F1**	**F vs. Naïve**	
LOG	64.46	82.27	14.15	63.14	24.14	0.00***	F > Naive
SVM	70.21	73.76	15.34	69.95	25.40	0.00***	F > Naive
DT	77.82	73.05	19.81	78.17	31.16	0.00***	F > Naive
RF	76.16	73.76	18.71	76.34	29.84	0.00***	F > Naive
NB	66.21	73.76	13.68	65.65	23.09	0.00***	F > Naive
ANN	70.70	73.76	15.57	70.47	25.71	0.00***	F > Naive

Table D5: McNemar's Test Results on Classifier Performance for Classifiers Using Self-Disclosed as well as User- and Regulator-Confirmed Information Compared to Classifiers E And F as Benchmarks Naturally Distributed Sample)

	Classifier	**D**		**D**		**F**	
	Benchmark	**E**		**F**		**E**	
Tech.							
LOG		0.00***	D > E	0.00***	D > F	0.00***	E > F
SVM		0.07*	D > E	0.48	F > D	0.00***	F > E
DT		0.00***	D > E	0.02**	F > D	0.00***	F > E
RF		0.00***	D > E	0.02**	D > F	0.00***	F > E
NB		0.23	D > E	0.01**	D > F	0.00***	E > F
ANN		0.08*	D > E	0.02**	D > F	0.37	E > F
Note: *p < 0.1, **p < 0.05, ***p < 0.01.							

About the Authors

Jens Lausen (lausen@wiwi.uni-frankfurt.de) is a research assistant and doctoral candidate at the Chair of e-Finance at Goethe University Frankfurt and in the research area "Data Science for Financial Services" of the efl—the Data Science Institute. Within his research, he focuses on empirical research in the areas of market microstructure, regulatory impact analysis, and decision support systems in finance and financial regulation. His work has been presented at various international conferences of the information systems and finance research communities and has been published in different outlets, such as the *Journal of the Association for Information Systems*.

Benjamin Clapham (clapham@wiwi.uni-frankfurt.de) is a postdoctoral research associate at Goethe University Frankfurt. His research focuses on technological and regulatory developments in financial markets. In particular, his research interests focus on market microstructure, algorithmic trading, and financial market manipulations. His work has been presented at various international conferences and has been published in outlets such as *Journal of the Association for Information Systems, Journal of Information Technology, and Information Systems Frontiers.*

Michael Siering (siering@wiwi.uni-frankfurt.de) is a postdoctoral research associate at Goethe University Frankfurt and works as a project manager in the financial services industry. His research focuses on decision support systems in electronic markets, with a focus on the analysis of user-generated content. His work has been published in outlets such as *Journal of Management Information Systems, Journal of the Association for Information Systems, Information Systems Journal*, *Journal of Information Technology,* and *Decision Support Systems.*

Peter Gomber (gomber@wiwi.uni-frankfurt.de) is a professor at Goethe University Frankfurt (Chair of e-Finance). He is co-chairman and member of the board of the efl—the Data Science Institute, an industry-academic partnership between Frankfurt and Darmstadt Universities and leading industry partners. Prof. Gomber's academic work focuses on fintech, information systems in financial markets, market microstructure theory and regulatory impact on financial markets. He published several articles on the above topics in leading international journals such as *Journal of Management Information Systems*, *Journal of the Association for Information Systems*, *Journal of Information Technology*, and *Decision Support Systems*. He was awarded with the Reuters Innovation Award, the University Award of the Deutsches Aktieninstitut, the IBM Shared University Research Grant and multiple Best Paper awards of international conferences. Prof. Gomber is a member of the Exchange Council of the Frankfurt Stock Exchange, deputy chairman of the supervisory board of b-next AG and a member of the supervisory board of Clearstream Banking. Before joining University of Frankfurt, Prof. Gomber worked as the head of Market Development Cash Markets and Xetra Research at Deutsche Börse AG.