Who Is the Next "Wolf of Wall Street"? Detection of Financial Intermediary Misconduct[*, **]

Forthcoming in the Journal of the Association for Information Systems (JAIS)

Jens Lausen[a], Benjamin Clapham[b], Michael Siering[c], and Peter Gomber[d]

[a]Goethe University Frankfurt, Germany, lausen@wiwi.uni-frankfurt.de

[b]Goethe University Frankfurt, Germany, clapham@wiwi.uni-frankfurt.de

[c]Goethe University Frankfurt, Germany, siering@wiwi.uni-frankfurt.de

[d]Goethe University Frankfurt, Germany, gomber@wiwi.uni-frankfurt.de

Abstract Financial intermediaries are essential for investors' participation in financial markets. Due to their position within the financial system, intermediaries committing 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 as well as warranting theory and making use of self-disclosed data with varying levels of external verification, we propose different classifiers that automatically detect financial intermediaries committing misconduct. Therefore, we focus on self-disclosed information by financial intermediaries on the business network LinkedIn. We match user profiles with regulatory-disclosed information and use this data for classifier training and evaluation. We find that self-disclosed information provides valuable input to detect financial intermediary misconduct. Regarding external verification, our classifiers achieve the best predictive performance when additionally taking regulatory-confirmed information into account. These results are supported by an economic evaluation. Our findings are highly relevant for both investors and regulators in order to identify financial intermediaries committing 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

[*] Please cite as: Lausen, J., Clapham, B., Siering, M., Gomber, P. (2020). Who Is the Next "Wolf of Wall Street"? Detection of Financial Intermediary Misconduct. Journal of the Association for Information Systems, Forthcoming.

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

1 Introduction

Financial intermediaries are essential for investors since they exhibit 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. Others, for example brokers, provide market access for trading financial instruments. Thus, brokers and investment advisors enable investors to participate in financial markets (Allen & Santomero, 1997). With increased usage of the Internet and electronic communication, personal interaction between investors and financial intermediaries has been reduced, which impedes the process of building trust (Ba & Pavlou, 2002; Castells, 2010). At the same time, the global financial crisis and widely noticed financial market manipulations have challenged investors' confidence towards financial intermediaries and their trust in the financial system (Palazzo & Rethel, 2008). Even the film industry took up the issue of misconduct and fraud by intermediaries 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 the willingness to rely on intermediaries, to participate in financial markets and thus negatively affect market-based allocation decisions and therefore the economy as a whole. Consequently, regulatory interventions and new instruments are needed to increase trust and to ensure the proper functioning of the financial system. Information systems and analytics play a crucial role within this context because they can help to identify misconduct and therefore increase trust in financial markets by preventing the next "Wolf of Wall Street".

Previous studies have proposed various approaches to identify different kinds of financial market manipulations (Ngai, Hu, Wong, Chen, & Sun, 2011). Nevertheless, these studies have

neglected the identification of financial intermediaries who commit misconduct. Today, social networks provide an important source of information regarding potential and ongoing business contacts as well as relationships. Thus, they are increasingly relevant for selecting financial intermediaries (Bazarova & Choi, 2014; Krasnova, Spiekermann, Koroleva, & Hildebrand, 2010). Therefore, many professionals, among them financial intermediaries, make use of these networks for advertising and setting-up contacts. Social networks offer the possibility to disclose information about oneself, while other members and external sources can confirm this information leading to different levels of reliability. However, such self-disclosed information with varying levels of external verification has not been considered yet to identify misconduct.

Based on *information manipulation theory* (McCornack, 1992), we argue that information disclosure of intermediaries committing misconduct differs from information disclosure of reliable market participants. Moreover, drawing on *warranting theory* (Walther, van der Heide, Hamel, & Shulman, 2009), we propose that external verification of self-disclosed information provides additional value to identify misconduct. Following this rationale, we address the research question *whether self-disclosed information with varying levels of external verification can be used to detect financial intermediaries committing misconduct*. We determine different feature sets that enable investors and regulatory/supervisory authorities to distinguish financial intermediaries that have committed financial misconduct from reliable ones. We compose a comprehensive data set of information that is self-disclosed by financial intermediaries on the professional social 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 committing misconduct by making use of different feature sets with varying levels of external verification. Furthermore, we examine the economic relevance of our classifiers by means of

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

an economic evaluation based on the financial compensations requested and paid. We find that self-disclosed information provided by financial intermediaries is valuable to detect financial intermediary misconduct. Specifically, classifiers additionally taking externally verified information into account achieve a promising classification performance and their application leads to considerable economic value for the 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 misconduct. Confirming *information manipulation theory*, we show that self-disclosed information of honest and dishonest financial intermediaries differs. This can be used to detect intermediaries committing misconduct. Supporting *warranting theory*, features that are externally verified significantly increase the classification performance. Moreover, our results are also highly relevant from a societal and economic point of view as they enable to build classifiers for the automated identification of financial intermediary misconduct. Investors can use the classifiers to check intermediaries in advance and, thereby, to reduce the likelihood of incurring losses due to 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 that execute a rising number of transactions for their clients. Therefore, such models support authorities to engage in predictive supervision by allocating their resources more efficiently to identify and closely monitor those intermediaries which are more likely to commit 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 committing 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 data set and the 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. Section 5 concludes.

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 do not achieve sufficient returns, for example for retirement plans. 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 common and costly at the same time (Dyck, Morse, & Zingales, 2010). Particularly, retail investors suffer damages in billions of dollars each year. The Council of Economic Advisors (2015) estimates that the aggregate annual costs of conflicted advice in individual retirement accounts (i.e., investment advice where conflicts of interest due to high commissions for intermediaries are present) amount to USD 17 billion.

Misconduct committed by brokers or investment advisors appears in different types. One type of misconduct is the improper dealing with customers by providing false and misleading information. 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 are brokers or investment advisors charging exaggerated fees or failing to obtain the best price for a client in a securities transaction.

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

Other types of misconduct affecting financial markets are 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. Thereby, they profit from the price movement due to the client’s (potentially large) trade that might move the market (Cataldo & Killough, 2003). “Churning” characterizes the excessive buying and selling of securities on a client’s account without consent of the client and disregarding the client’s interest in order to generate higher commissions for the intermediary (Cumming & Johan, 2008). “Scalping” is the practice of investment advisors to purchase a security before recommending the same security to a client without disclosing their position to benefit from a potentially higher price if the customer follows the recommendation (Hazen, 2010). A more detailed overview of financial market manipulations that are conducted by intermediaries is presented by Siering, Clapham, Engel, and Gomber (2017).

Brokers and investment advisors in the US are subject to a comprehensive system of regulation. FINRA, the responsible competent authority, mandates the disclosure of material facts about every broker and investment advisor, among them any allegations and wrongdoings (Lazaro, 2014). Using public and non-public 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 committing misconduct and is not helpful for investors to protect themselves. Therefore, this paper presents different classifiers to detect misconduct based on self-disclosed information of financial intermediaries with varying 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 to identify fraudulent activities in financial markets, where manual detection is time consuming, expensive, and impractical due to the large amount of data to be analyzed (West & Bhattacharya,

2016). First studies on fraud detection in financial markets mostly rely on logistic regressions (Lee, Ingram, & Howard, 1999; 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 concerning 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, Jha, Tharakunnel, & Westland, 2011) and accounting fraud (e.g., Wang, 2010). Based on a real-life data set of credit card transactions, Bhattacharyya et al. (2011) are able to identify fraudulent transactions using random forests and support vector machines. Regarding 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 are able to correctly classify (non-)fraudulent financial statements. Besides structured data, researchers also apply text mining methods to analyze linguistic cues in unstructured parts of regulatory or financial disclosures. Humpherys, Moffitt, Burns, Burgoon, and Felix (2011) apply Naïve Bayes and decision trees taking linguistic variables from the management discussion and analysis section into account to distinguish between fraudulent and non-fraudulent 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) take user-generated content from financial social media platforms into account and show that this type of unstructured data adds incremental value for 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. Previous research focuses on identifying single incidents of fraudulent behavior, for example by analyzing market data or

financial statements, but disregards detecting individual intermediaries committing misconduct. Consequently, this paper aims at closing this research gap by developing classification models to detect financial intermediaries committing misconduct.

2.3 Theoretical Background

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

In order to automatically identify financial intermediary misconduct, we base our feature selection on two important theoretical streams in the context of self-disclosure in social networks and fraud detection. First, we rely on *information manipulation theory* (McCornack, 1992) to motivate why individuals committing misconduct communicate differently compared to honest individuals. Second, *warranting theory* (Walther et al., 2009) provides the theoretical underpinning 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 information quality or simply lie to disguise facts. Third, the relevance of information can be out of context to mislead the receiver. The fourth principle states that the manner in which information is communicated can be chosen to be ambiguous with the objective to confuse the receiver. *Information manipulation theory* has been empirically tested in the context of financial reporting fraud. In this context, researchers were able to show that writers of misleading financial statements actually use techniques posited in *information manipulation theory* to deceive (Glancy & Yadav, 2011; Humpherys et al., 2011). Also, classifiers that take feature sets based on *information manipulation theory* into account are able to 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), they still deviate from these profiles since deceivers cannot report as detailed and precise on information they did not experience. However, *interpersonal deception theory* does not fit to the setup at hand because it builds on the repeated communication exchange between sender and receiver (Buller & Burgoon, 1996) whereas the LinkedIn profiles analyzed in this study are static information representing a "monolog" of the intermediary. Therefore, we assume that financial intermediaries' self-disclosed information on the business network LinkedIn differs between intermediaries who committed misconduct and intermediaries who have not committed 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. The theory postulates that information, which is harder to manipulate, is more plausible or trustworthy than information that is easier to manipulate. In case of self-disclosed information in social media profiles, different levels of external verification can be observed. Here, other users can confirm information provided by the individual profiles. Furthermore, self-disclosed information can be counter-checked with data provided by known and reliable external sources, which is an even stronger verification than 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 which is hard to manipulate.

2.4 Research Hypotheses

In order to answer our research question *whether self-disclosed information with varying levels of external verification can be used to detect financial intermediaries committing 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, which leads to anomalies that help to detect

misconduct by appropriate detection mechanisms. Therefore, self-disclosed information of financial intermediaries can be valuable to detect intermediaries committing misconduct if classifiers based on related features can detect misconduct better than pure chance. *Information manipulation theory* states that deceivers tend to conceal or misrepresent information and thus communicate and provide information differently compared to honest individuals (McCornack, 1992). Consequently, unverified, self-disclosed information in business networks such as LinkedIn can be assumed to be valuable to detect financial intermediaries committing misconduct. Therefore, we derive our first hypothesis as follows:

H1: Self-disclosed information, which is not verified by third parties, is valuable to detect financial intermediaries committing misconduct.

According to *warranting theory*, self-disclosed personal information incorporates higher credibility if the individual cannot easily manipulate it on her own (Walther et al., 2009). In social networks, other users can confirm information disclosed by social network participants. This confirmation may occur, for example, by 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 to confirm self-disclosed information appears to be a valuable distinction for trustworthy financial intermediaries and intermediaries committing misconduct. Thus, self-disclosed information verified by other users can be regarded as more reliable than unverified information, which leads us to our second research hypothesis:

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

Besides users of the social network, also regulatory authorities can verify intermediaries' self-disclosed information. This means, self-disclosed information can be counter-checked with data published by the regulator. As regulatory data represents a reliable and neutral source,

regulatory confirmations are an even stronger external verification than confirmations by anonymous or unknown users of a social network. Consequently, we investigate the following third hypothesis:

H3: Classifiers additionally taking into account 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 committing 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 pre-processing (Kurgan & Musilek, 2006). We create the target data set by means of data extraction from business profiles on LinkedIn, which are matched to regulatory data from FINRA's website BrokerCheck. In a next step, we clean and pre-process the data. Subsequently, we select appropriate data mining and machine learning techniques and evaluate the resulting classifiers both statistically and economically. The whole data mining process is depicted in Figure 1.

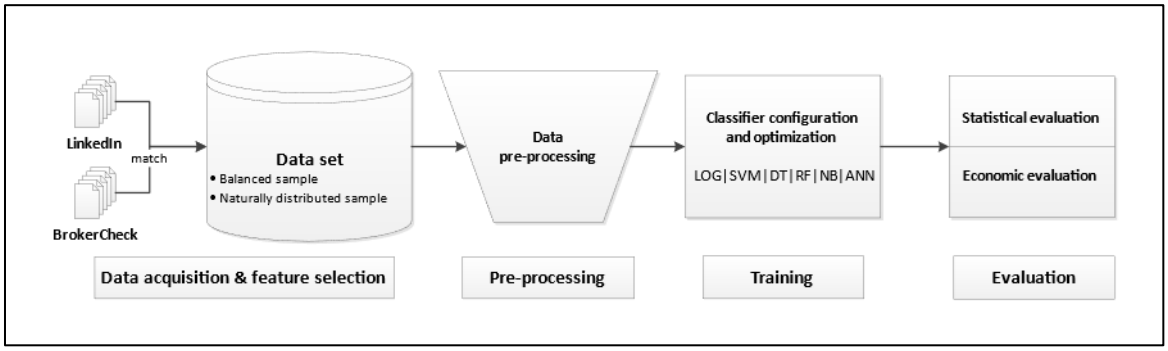


Figure 1: *Data mining process*

3.2 Data Acquisition

In order to train and evaluate different classifiers to detect financial intermediaries committing misconduct, we take BrokerCheck as our starting point, which represents a complete record of all brokers registered in the US. We randomly draw 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 the real-world class distribution. This is necessary since we observe that historically only 6.83 percent of the intermediaries listed on BrokerCheck have actually performed misconduct.

The balanced sample is composed of 400 brokers who already committed misconduct and 400 brokers without misconduct cases and is used for classifier training. Thereby, we use random under-sampling for the no-misconduct class to under-sample the majority class at random until it has the same number of observations as the minority class (Chawla, 2009; Japkowicz, 2000). We use a balanced data set for training 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 collect a second, distinct sample for testing the optimized models with naturally-distributed data. This testing sample is collected randomly and represents the natural distribution of intermediaries having committed and not having committed misconduct. This enables us to evaluate whether the trained models based on the balanced sample are able to deal with naturally-distributed data. For the testing sample, we collect another 2,051 brokers, thereof 141 with misconduct and 1,910 without.

Both samples are collected randomly to ensure representativeness of the collected data. Because not all brokers have a profile on LinkedIn, we need to scan more brokers on BrokerCheck than we include in the final data set of 2,851 brokers that self-disclose information

on their LinkedIn profile. To identify all 2,851 matched LinkedIn profiles, we inspect 4,729 registered brokers on BrokerCheck in total (thereof 1,319 brokers for the equally balanced sample and 3,410 for the naturally-distributed sample). Consequently, 60.29 percent of the inspected financial intermediaries have a LinkedIn profile, which can unambiguously be assigned.

Since brokers with LinkedIn profiles might have different characteristics than brokers without a LinkedIn profile, we need to rule out a potential selection bias. Therefore, we compare the information provided on BrokerCheck for both groups. Most importantly, our dependent variable (broker has committed misconduct or not) is almost identically distributed within both groups: While 6.87 percent of the intermediaries in our naturally-distributed sample for whom a LinkedIn profile can be identified committed misconduct, this holds for 6.77 percent of the brokers who do not self-disclose information on LinkedIn. Moreover, also the other characteristics provided by BrokerCheck are highly comparable for brokers with and without LinkedIn profile (see Table 1). Consequently, with respect to our independent variable, there is no selection bias due to merging the data with LinkedIn profiles. Nevertheless, brokers who self-disclose information on LinkedIn have on average a shorter mean employment duration (83.08 vs. 97.00 months) and more state licenses (14.68 vs. 12.60) compared to brokers without LinkedIn profile. Although these variables show similar distributions for both groups, these differences may weaken the generalizability of our results to brokers without LinkedIn profiles.

Table 1: Comparison of brokers with and without LinkedIn profile

	Brokers with LinkedIn profile N = 2,851				**Brokers without LinkedIn profile 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 employments	1.00	35.00	3.65	2.67	1.00	29.00	3.72	2.86
Number of exams	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

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

Each broker in both final samples (i.e., balanced sample and naturally-distributed sample) is simultaneously registered on the regulatory authority's website BrokerCheck and on the business network LinkedIn. Consequently, our classification models are based on a data set composed of both, autonomously self-disclosed information on LinkedIn and publicly reported regulatory confirmed information on BrokerCheck. This information makes our data set unique and particularly useful to analyze the information provision by financial intermediaries as well as varying levels of verification.

BrokerCheck contains information about the background and experience of brokers and investment advisors and discloses information about misconduct resulting in regulatory actions, arbitrations and complaints. In our data set, the group of financial intermediaries who already committed misconduct consists of brokers (who may also be registered as investment advisors) and exhibits disclosed *customer disputes* and *regulatory actions*. These disclosures of misconduct relate to actions in the role of a broker or investment advisor that damage individual investors or the society as a whole. *Customer disputes* are mainly based on misbehavior of brokers like misrepresentation of material facts, unsuitable recommendations of financial products, and securities fraud such as churning or front running. Examples for *regulatory actions* are unauthorized trading and insider trading. In order to exclude pending decisions and lawsuits, we only consider those disclosures with a final status and where a compensation or a fine has been paid. This guarantees that we only include cases where intermediaries admitted wrongdoing and were willing to pay a compensation or where intermediaries were convicted. Therefore, in the following, the term "misconduct" refers to a customer dispute or a regulatory event which is final, settled, or where a judgment occurred. Thereby, we do not differentiate between different types of misconduct because any misconduct by intermediaries is harmful for investors and weakens investors' trust in financial intermediaries and markets. Moreover, distributions and median values of occurred damages in Figure 8 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 within our study. As additional information, we report further summary statistics about brokers with misconduct cases and different

types of misconduct in Table 14 and Figure 7 in the appendix. For our analysis, misbehaving intermediaries are those that have one or more misconduct cases according to the criteria outlined above, while the group of financial intermediaries who did not commit misconduct consists of brokers with no misconduct at all. In order to account for the severity of different misconduct cases, we perform an economic evaluation in Section 4.3 that takes the amount of compensation payment, which the intermediary either has to pay to the customer or to the regulator, into account.

In addition to the information collected from BrokerCheck, we hand-collect self-disclosed information from matched profiles on LinkedIn. LinkedIn is the world's largest business-related social networking website on which individuals can self-disclose personal information including, among others, working experience, education, and skills. LinkedIn profiles are matched with registered brokers on BrokerCheck according to names, employments, and location. We only add brokers to our data set where matched LinkedIn profiles are distinct. In case of common names leading to multiple possible profile matches, we further consider name suffixes, nicknames, previous employments, or the unique FINRA identification number to match the correct profile. In order to control for fake profiles, we follow common techniques in the social network analysis (Adikari & Kaushik, 2014) and only consider those profiles of intermediaries that include logical and reasonable information. Therefore, we qualitatively assess in a first step whether the information provided in the different sections is free of self-contradictions. Specifically, we verify whether there is a logical flow concerning education and job experience, whether the disclosed skills are suitable for a broker or investment advisor, and whether the stated interests reflect current and previous employers, universities, and groups related to financial services topics. In a second step, we exclude those suspicious profiles whose number of connections to legitimate users is significantly below the average number of connections of

profiles in our sample. In total, only three profiles are suspicious based on the qualitative assessment.[3] Since all these profiles have less than five connections (three, one, and zero), they are excluded as the number of connections is significantly below the average number of 289.5 connections in our naturally-distributed sample. However, our results remain robust if we add the three potential fake profiles to our data set.

3.3 Feature Selection

We extract a large number of features collected from publicly available information on BrokerCheck and LinkedIn. Figure 2 schematically depicts the presentation of information on both websites BrokerCheck and LinkedIn.

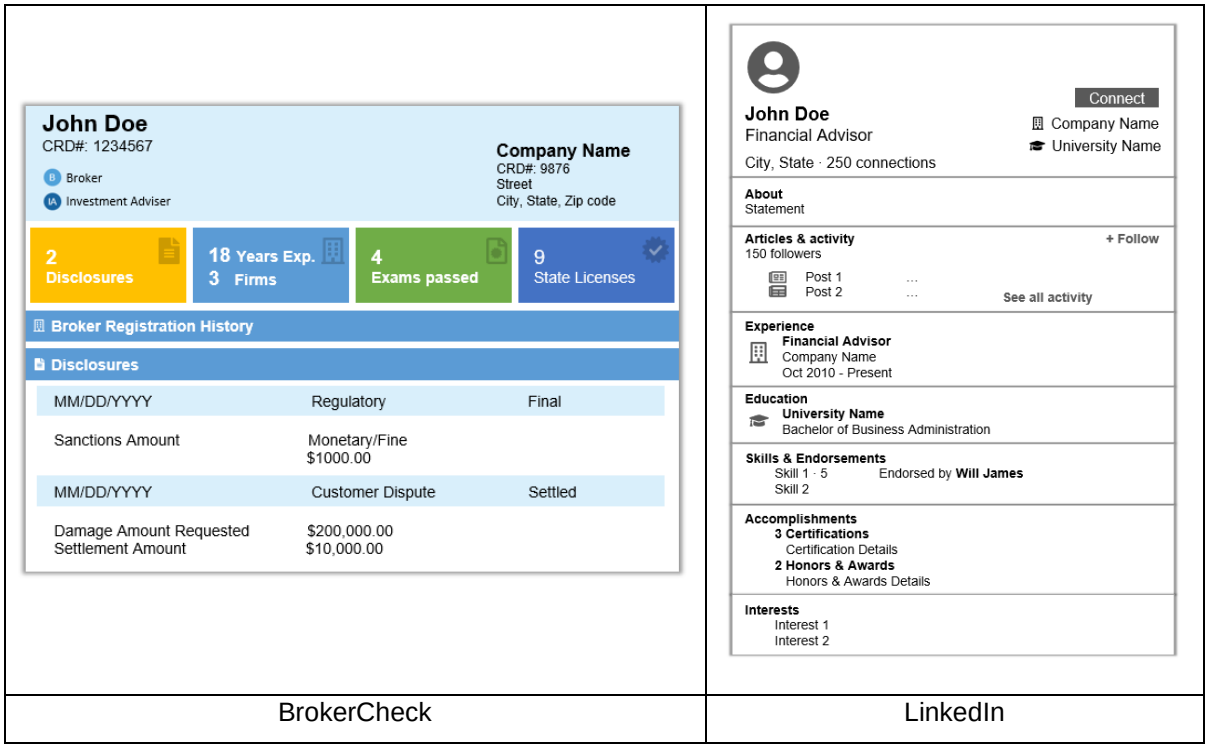


Figure 2: *Publicly available information provided on BrokerCheck and LinkedIn*

As described in Section 3.2, in order to train classifiers to detect financial intermediaries committing misconduct, we separate brokers into two groups: Brokers who already committed misconduct and brokers without any form of misconduct in the past. As dependent variable, we

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

use a binary variable being one for those brokers with misconduct and zero for those without any misconduct event.

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 employments; multiple positions are counted as separate employments
	li_empl_details	Variable equaling 1 if durations of employments are self-disclosed, 0 otherwise
	li_avg_empl_dur	Average duration of employments 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* is information that describes the profile owner, *social network activity* is information on how active an intermediary is on the social network, *professional information* is information on job experience, employments, skills, and education, and *profile summary* represents eye-catching information which gives the profile visitor a first impression of the profile owner. From the profile summary, we extract linguistic cues via textual analysis 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 suppose that self-disclosed information of intermediaries committing misconduct differs from self-disclosed information of intermediaries without any misconduct case. In particular, the first principle, i.e., 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 therefore in each category shown in Table 2. The second principle, i.e., altered information or lies can be expected in the provision of professional information, e.g., past employments or education, and profile summary. The third principle, i.e., relevance of the self-disclosed information, can be assumed 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, i.e., ambiguity of information is represented by the uncertainty and complexity expressed in the profile summary.

Table 3: Features used for classification based on user and regulatory confirmed data

Category	Feature	Description
User confirmed information	li_rec_ob	Number of obtained recommendations on LinkedIn
	li_end_skill	Proportion of endorsements to skills calculated from self-disclosed skills and their endorsements on LinkedIn
Regulatory 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 employments in months calculated based on regulatory disclosed information on BrokerCheck
	bc_jobs	Number of employments 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 employments between LinkedIn and BrokerCheck

Table 3 provides features representing varying levels of confirmed information according to *information manipulation theory* as well as *warranting theory*. *User confirmed information* is self-disclosed information on LinkedIn profiles confirmed by other users. *Regulatory confirmed information* is defined as regulatory information published on BrokerCheck, which can be used to verify self-disclosed information on LinkedIn. Both categories represent information that cannot be easily manipulated. In particular, user recommendations and the proportion of endorsements to skills directly relate external verification to intermediary misconduct since intermediaries who committed misconduct should have significantly fewer recommendations than intermediaries without any misconduct event. Also, while intermediaries with misconduct might try to polish their profiles by stating a variety of skills as suggested by *information manipulation theory*, other users on LinkedIn, e.g., customers or colleagues of the intermediary, only endorse these skills when the intermediary delivers good work, i.e., does not commit misconduct. Thus, there should be a difference in the proportion of endorsements to skills for misbehaving brokers and those without misconduct.

Finally, as financial intermediaries do not have the possibility to manipulate the information provided on BrokerCheck, regulatory confirmed information is even more reliable and discrepancies between the LinkedIn profile and information published on BrokerCheck can be assumed to be valuable for identifying misconduct. In particular, the two features based on deviations between regulatory information on BrokerCheck and self-disclosed information on LinkedIn enable to detect a broker's misrepresentation in practice. First, several brokers with misconduct in our sample conceal frequent job changes on LinkedIn, in particular short employments and those employments where a misconduct occurred. In our sample, we identified brokers that report, e.g., only four out of eight employments or five out of eleven, which reveals a deviation between their full job history on BrokerCheck and their self-disclosed information on LinkedIn. Second, several brokers do not disclose their employment history up to the point where a misconduct occurred. For example, two brokers in our sample only disclose the most recent one out of their three employments as they committed misconduct during their second employment, again revealing a deviation between their information provided on LinkedIn and

BrokerCheck. Third, there are brokers that misrepresent their work experience on LinkedIn and report more work experience on LinkedIn than reported on BrokerCheck to appear more experienced.

3.4 Data Pre-processing and Textual Analysis

Pre-processing is necessary for non-numeric features so that these features can be taken into account by machine learning algorithms. In particular, self-disclosed information regarding the firm where the broker or investment advisor is employed, the job, 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 gathered and categorized the self-provided job titles with respect to their career level according to standard career levels in the financial industry (Eccles & Crane, 1987). Firms are categorized based on their primary business model (e.g., bank or insurance company) and banks are further split into large and small institutions based on the total assets reported in their annual filings.[4] The location of the broker or investment advisor is determined as urban if the city or metropolitan area provided on LinkedIn has more than 200,000 inhabitants and rural otherwise. To make categorial features processable for our machine learning techniques, we use one-hot encoding (also called dummy encoding), which is a standard approach for nominal variables (Wooldridge, 2009).

In order to analyze the profile summaries which brokers and investment advisors provide to describe themselves in more detail on LinkedIn, we perform common text pre-processing steps and generate quantitative linguistic features. First, we remove parts of the text that do not contain relevant information such as email-addresses, website URLs, numbers, single character words, and state abbreviations. In addition, we remove dots that are not part of a sentence and dots that do not represent the end of a sentence such as in company suffixes (e.g., Inc. or Ltd.), common abbreviations (e.g., Mr., No., Jr.), and after a single letter indicating a middle

[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.

name (e.g., Paul J. Smith). Second, we transform the cleaned text into lower cases and split the text into individual words.

We rely on the Harvard IV-4 dictionary in order to calculate common textual analysis measures such as share of positive, negative, strong, and emotional words. Although there also exist specific dictionaries tailored to a financial context such as financial statements (Loughran & McDonald, 2011), we rely on the more general Harvard IV-4 dictionary because brokers and investment advisors introduce themselves in the profile summaries using general and not financial language. In addition, we follow Zhou, Burgoon, Nunamaker, and Twitchell (2004) and determine the share of uncertainty and modal words, both measuring uncertainty in texts. Based on the number of positive and negative words, we also determine the sentiment of the profile summaries (see Equation (1)).

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

For analyzing the complexity of the profile summary, we calculate the average number of words per sentence using the Stanford CoreNLP toolkit (Manning et al., 2014). Besides average words per sentence, we include the share of complex words and the fog index (Li, 2008) as depicted in Equation (2) as readability measures. Thereby, a complex word is understood as a word with more than two syllables.

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

For most of the self-disclosed information gathered from LinkedIn, users deliberately decide whether they want to provide information on a specific category or not. For example, users can either present their number of connections, skills, or interests or decide not to disclose any of them. In the latter case, the variables measuring such information disclosure are set to zero. Also, if a broker has no written profile summary, all variables based on textual analysis of the profile summary are set to zero. Missing values in a narrower sense only exist if employment durations are not self-disclosed on LinkedIn and thus the deviation to regulatory confirmed

information on BrokerCheck is not measurable (less than six percent of the observations). We replace these missing values with zero and include a dummy control variable checking whether employment details are 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 standardize our numerical features with zero mean and unit variance and use a K-Nearest-Neighbor (K = 50) approach based on all features to drop outliers with distances above the 99 percent percentile in our training data to avoid biases in our models (James, Witten, Hastie, & Tibshirani, 2017).

3.5 Machine Learning Techniques Applied

We rely on different machine learning techniques in order to develop classifiers to detect financial intermediaries committing misconduct. Specifically, we consider logistic regression (LOG) as a baseline and support vector machine (SVM), decision tree (DT), random forest (RF), Naïve Bayes (NB), as well as artificial neural networks (ANN)[5] as machine learning techniques that generate 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 (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 and Bhattacharya (2016).

In order to ensure robust and generalizable models, we apply 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 train multiple models using a random bootstrap sample of 80 percent of our data for each single model and perform classification by a major vote of all classifiers. Since RF classifiers are already a specific kind of bagging

[5] We use feed-forward neural networks.

classifier (also using a random subset of features for each single model in the forest), we do not use an additional bagging classifier for RF. We also do not apply bagging for our ANN models because their performance is better using the whole training data set than with using the bagging classifier approach.

3.6 Classifier Configuration and Hyperparameter Tuning

In order to analyze our research hypotheses whether self-disclosed as well as user and regulatory confirmed information is valuable to detect financial intermediaries committing misconduct, we compose different classifiers based on varying levels of verification. Table 4 provides an overview of all composed classifiers of our empirical analysis. Each classifier configuration represents a different level of verification. While classifier A is the baseline using only features based on self-disclosed information, classifiers B, C, and D additionally use different sets of features based on user and regulatory confirmed information. For each classifier, we apply different machine learning techniques as described in Section 3.5. For the sake of completeness, we also include two classifiers based on regulatory confirmed information only as well as regulatory confirmed and user confirmed information. The configuration of these additional classifiers can be found in the appendix (see Table 17).

Table 4: Classifier configuration

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

We train each machine learning technique for all classifiers on the *balanced sample* and tune hyperparameters to optimize F1 score using a grid-search. As described in Section 3.5, we use bagging classifiers, and therefore, further optimize 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 15 in the appendix.

3.7 Evaluation Methodology

Statistical Evaluation

For training and optimization of our models using the balanced sample, we make use of ten-fold 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 data sets (Kohavi, 1995). Afterwards, we evaluate the classification performance of the resulting classifiers on the naturally-distributed sample. In each case, we calculate a confusion matrix and compute 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 make use of the McNemar's test (Everitt, 1977), which compares the performance of two different classifiers. Since the McNemar's Test is a two-sided test, we also report the direction of which classifier outperforms the other.

In addition to assessing one specific configuration of a classifier, we also evaluate our models when considering different classification thresholds that need to be reached to classify an observation as positive. For this purpose, we use 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 data sets (Saito & Rehmsmeier, 2015). The ROC curve simultaneously displays the two classification errors Type-I error (x-axis, false positive rate = 1 – specificity) and Type-II error (y-axis, recall = 1 – Type-II 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).

Economic Evaluation

Besides the above-mentioned machine learning metrics, we additionally perform an *economic evaluation* of the classifiers proposed in this paper. Domain-specific evaluations are important

to assess the value of classifiers composed for specific classification problems and allow for additional statistical analysis (Groth, Siering, & Gomber, 2014). In order to assess the economic gain achievable by a misconduct detection mechanism, we design an evaluation methodology that accounts for interaction between investors and financial intermediaries. Specifically, we derive the economic value of an automated classifier by computing the investor's potential damage that can be avoided by using the classifiers. We use the classification results of the naturally-distributed testing sample representing randomly collected real-world data so that 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 commits misconduct (true positive, TP), an economic loss in the amount of the investor's damage is prevented and can thus be considered as an economic gain. To approximate an investor's damage, we rely on the compensation payment (*cp*) by the financial intermediary. Thereby, we account for the severity of different misconducts. Nevertheless, the investor will select a different intermediary to execute her trade or investment, which leads to additional search costs (*sc*). (2) In case an intermediary is classified incorrectly and actually commits misconduct (false negative, FN), the investor incurs a damage equaling the compensation payment. (3) If the financial intermediary is incorrectly classified as an intermediary committing misconduct (false positive, FP), the investor will unnecessarily select a new intermediary and has to bear additional search costs. (4) In case the intermediary is classified correctly as not-committing misconduct (true negative, TN), the investor sticks to this intermediary and has 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 builds on the compensation payment (*cp*) respectively the average compensation payment if a broker had several

misconduct events in the past. Search costs as defined above refer to costs of finding another suitable intermediary including corresponding opportunity costs (e.g., resulting from non-executed trades). Search costs differ among investors depending on the necessary effort to identify a new intermediary to conduct business with and the individual loss incurred due to opportunity costs, and therefore, are hard to quantify. Still, compared to potential damages an investor might suffer, for example, through losses in retirement plans due to unsuitable advice or false information, search costs are negligible (Egan, 2019). Thus, we assume search costs within our economic evaluation to be zero. Nevertheless, we conduct a sensitivity analysis with different levels of search costs and varying classification thresholds to ensure robustness of our results.

To analyze the economic value of the different classifiers, we compare their average economic gain. Therefore, we average 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 aims for investor protection to ensure fair and efficient markets.

4 Empirical Study

4.1 Descriptive Statistics

Table 5 provides the descriptive statistics for all features of the balanced data set 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. For most features representing self-disclosed information on LinkedIn, we observe differences in means, whereas 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 have significantly longer profile summaries, while the content is significantly harder to read (fog index, words per sentence, share of complex words). This provides indications that the content

is more ambiguous, which is in line with *information manipulation theory* suggesting that dishonest individuals tend to polish their profiles to mislead receivers. Consequently, this indicates that self-disclosed information on LinkedIn is potentially useful to identify misconduct among financial intermediaries.

Manipulation becomes more difficult when information is validated externally as suggested by *warranting theory*, whereas it can be assumed that non-misbehaving intermediaries receive more external validation than misbehaving ones. 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 that committed misconduct. Moreover, also the number of recommendations is higher although the difference is not significant.

For regulatory confirmed information, the WRS test shows highly significant differences for intermediaries with and without misconduct for all features. This is especially true for features taking into account deviations of self-disclosed information from information provided by BrokerCheck. These results provide a first indication that features based on user and regulatory confirmed information are especially valuable for detecting financial intermediaries committing misconduct.

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**

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

Table 5 continued									
	Misconduct N = 400				No misconduct N =400				WRS test
Feature	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**p**
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***
Reg. 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 16 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 five percent level, there is no significant difference between intermediaries with misconduct in the training and testing sample. Those without misconduct in the testing sample show significant but small differences concerning the number of jobs self-disclosed on LinkedIn, indicating that training and testing sample are comparable regarding 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 train and optimize 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 15 in the appendix.

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-disc. + user conf.				
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-disc. + reg. conf.					Self-disc. + user & reg. conf.				
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

The evaluation metrics show meaningful values for all classifiers and machine learning techniques. Specifically, DT and RF show the highest values for most of the metrics and the majority of classifiers, while LOG and ANN follow up closely. Comparing the different classifiers A to D, the results provide first indications regarding 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 percent for balanced data), classifiers based on self-disclosed information can add value to the detection of intermediary misconduct (H1). Second, while classifier B shows overall slightly worse scores than classifier A for most of the machine learning techniques, and as the results of comparing classifier D with classifier C are mixed, we cannot yet provide indications that user confirmed information is valuable for the detection (H2). Third, the comparison of classifiers including regulatory confirmed information (C and D) with classifier A reveals that classifiers C and D achieve higher scores for almost all evaluation metrics suggesting that regulatory confirmed information may add value to the clas-

sification models (H3). Nevertheless, these are only first indications based on the training results from the ten-fold stratified cross-validation and the hypotheses need to be further analyzed based on the evaluation of the naturally-distributed hold-out sample.

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 taking into account self-disclosed personal information without external verification (classifier A) for classifying the naturally-distributed 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. Naïve**	
LOG	64.16	71.63	12.69	63.61	21.56	0.00***	A > Naïve
SVM	69.58	62.41	13.35	70.10	22.00	0.00***	A > Naïve
DT	67.38	73.76	14.13	66.91	23.72	0.00***	A > Naïve
RF	64.02	74.47	13.01	63.25	22.15	0.00***	A > Naïve
NB	58.41	72.34	11.14	57.38	19.30	0.00***	A > Naïve
ANN	67.67	72.34	14.05	67.33	23.53	0.00***	A > Naïve

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

All machine learning techniques achieve meaningful accuracy, recall, and specificity showing that the majority of true misconduct cases (recall of up to 74.47 percent for RF) and true non-misconduct cases (specificity up to 70.10 percent for SVM) is classified correctly. Still, precision and thus F1 score yield lower values (maximum precision score is 14.13 percent for DT). This implies that regarding the share of all predicted misconduct cases, only 14.13 percent (every seventh case) are true misconduct cases. Nevertheless, this precision score is in line with other classifiers that are applied to highly imbalanced data sets (Tan, Tan, Dara, & Mayeux, 2015; Zhang & Mani, 2003) and is regularly observed in case of data sets with very few observations of the class to be predicted (Menzies, Dekhtyar, Distefano, & Greenwald, 2007). Moreover, from the perspective of regulators and/or supervisors, using the classifier as a decision support tool to later manually inspect the predicted misconduct cases improves the hit ratio from every 15th to every seventh case compared to randomly selecting a subset of all

financial intermediaries (given the historical unequal distribution of only 6.83 percent true misconduct cases). Since regulatory/supervisory resources are limited, using the proposed classifiers frees up capacities to further inspect more intermediaries, which is in line with Zhang, Bloedorn, Rosen, & Venese, 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 most important that reliable brokers are predicted correctly (for our classifiers, this holds for 96 to 98 percent 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 more important with rising search costs. We further elaborate on this in the economic evaluation in Section 4.3.

The value of self-disclosed information to detect intermediaries that committed misconduct is also supported by the McNemar's test. For all machine learning techniques, classifier A significantly outperforms a naïve classification algorithm that randomly classifies financial intermediaries whether they committed misconduct or not based on the historical misconduct ratio of 6.83 percent. Consequently, given the high percentage of correctly classified misconduct cases as well as the high percentage of correctly classified non-misconduct cases, self-disclosed personal information can be used to detect financial intermediaries committing misconduct (H1).

H2: User confirmed information

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-disc. + user conf. 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 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 compared to classifier A, accuracy, precision, and specificity are lower. The results of the McNemar's test even show that classifier A significantly outperforms classifier B for five of the six machine learning techniques. Only NB yields an overall better performance. Consequently, the results do not provide support for H2.

Table 9: Detailed summary statistics for features based on user confirmed information

Number of obtained recommendations on LinkedIn (li_rec_ob)

Misconduct	count	mean	std	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

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	std	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

Of the overall 2,051 observations, 890 observations have a value greater than zero.

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 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 have received recommendations from other users, while intermediaries without misconduct have obtained more recommendations. Still, for the large majority of observations, this feature does not provide any information for the classifiers (about 90 percent of the observations).

For the feature li_end_skill representing the number of endorsements per skill on LinkedIn, a lot more observations have values larger than zero (890 out of 2,051). Also, the percentiles show differences for misconduct and non-misconduct cases that confirm the results of the WRS test in Table 5 and Table 16. Nevertheless, for more than 50 percent of the observations,

this feature does not provide any information for the classifiers. Thus, dependent on the machine learning technique applied and the respective hyperparameter configuration of the model, both features based on user confirmed information may not deliver enough information gain to improve the models. Moreover, this might even lead to models that perform worse than models only considering self-disclosed information.

H3: Regulatory 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 regulatory confirmed information is mandatory. Therefore, regulatory confirmed information is available for every broker and provides an even stronger confirmation of self-disclosed information than user confirmations.

Table 10: Classifier evaluation for classifiers using self-disclosed as well as user and regulatory confirmed information (in %, naturally-distributed sample)

	Classifier C					**Classifier D**				
Cues	Self-disc. + reg. conf.					Self-disc. + user & reg. conf.				
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 10 provides the results for applying the trained models on the naturally-distributed sample for classifiers additionally taking self-disclosed information verified by regulatory authorities into account. Compared to classifiers A and B, classifiers C and D show higher scores for all machine learning metrics indicating that self-disclosed information as well as regulatory confirmed information add value to the detection of intermediary misconduct. Classifier D using RF shows the highest overall performance for all machine learning metrics having an accuracy of 77.96 percent, while predicting 75.18 percent of true misconduct cases (recall) and 78.17 percent of true non-misconduct cases (specificity) correctly. Further, classifier D based on RF shows the highest precision and F1 score increasing the share of true misconduct cases in all

predicted misconduct cases compared to all other classifiers. Classifier D based on RF is closely followed by classifier D using DT, while classifier D based on SVM performs best in terms of identifying true misconduct cases (recall of 77.30 percent). Nevertheless, for SVM, accuracy, precision, and specificity are lower compared to the other machine learning techniques, thus leading to classifier D based on RF being the best performing classifier.

Table 11: McNemar's test results on classifier performance for classifiers using self-disclosed as well as user and regulatory confirmed information compared to classifiers A and B as benchmark (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.

Table 11 provides the results of the McNemar's test comparing the performance of the different classifiers. The results show that classifiers C and D significantly add value compared to classifiers A and B, and thus support 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 including user confirmed information (classifier D) for four of the six machine learning techniques. This suggests that user confirmed information can be valuable when combined with regulatory confirmed information, especially for tree-based models and neural networks. For these models, the apparently low availability of user confirmed information together with regulatory confirmed information provides enough information gain to be incorporated in the models, and therefore, improves the performance of the classifiers.

The results for the additional classifiers E and F, which only consider regulatory confirmed as well as user and regulatory confirmed information, are presented in Table 18 to Table 20 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 regulatory confirmed information is valuable. Still, classifier D significantly outperforms classifiers E and F for five respectively four machine learning techniques (see Table 21 in the appendix). These results again support that self-disclosed information is valuable to detect financial intermediary misconduct (H1).

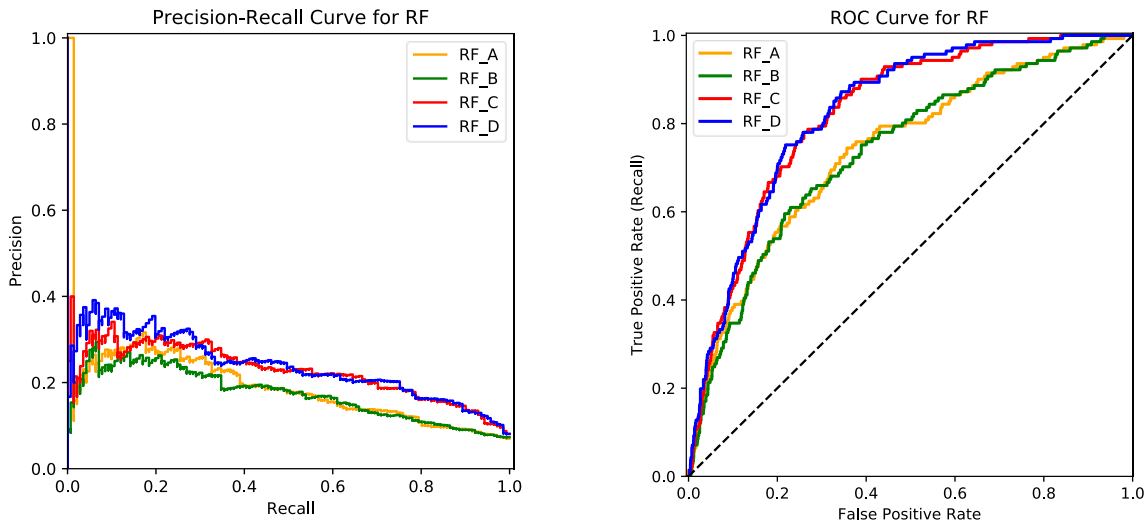


Figure 3: Precision-recall curve and ROC curve comparing classifiers using random forests as machine learning technique

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 Figure 9 and Figure 10 in the appendix show 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 to D. While precision declines with increasing recall, the precision for higher recall scores (above 50 percent) ranges between ten and 20 percent for all machine learning techniques. Considering the ROC curves, we find similar patterns regarding the relation of recall and 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 as well as classifiers C and D. In contrast, we see meaningful differences between classifiers A and C, A and D, B and C, as well as B and D. This

provides further evidence that classifiers including regulatory confirmed information improve classification performance (H3), while we cannot 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.

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%

Although classifiers C and D exhibit comparable AUC scores, classifier D based on RF shows the overall best performance with a score of 82.83 percent. Figure 4 compares the best classifiers for each machine learning technique according to the AUC score. The charts show that classifier D based on 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 has the overall best performance. This result again supports that for classifiers using regulatory confirmed as well as user confirmed information, user confirmed information can add value to the classification.

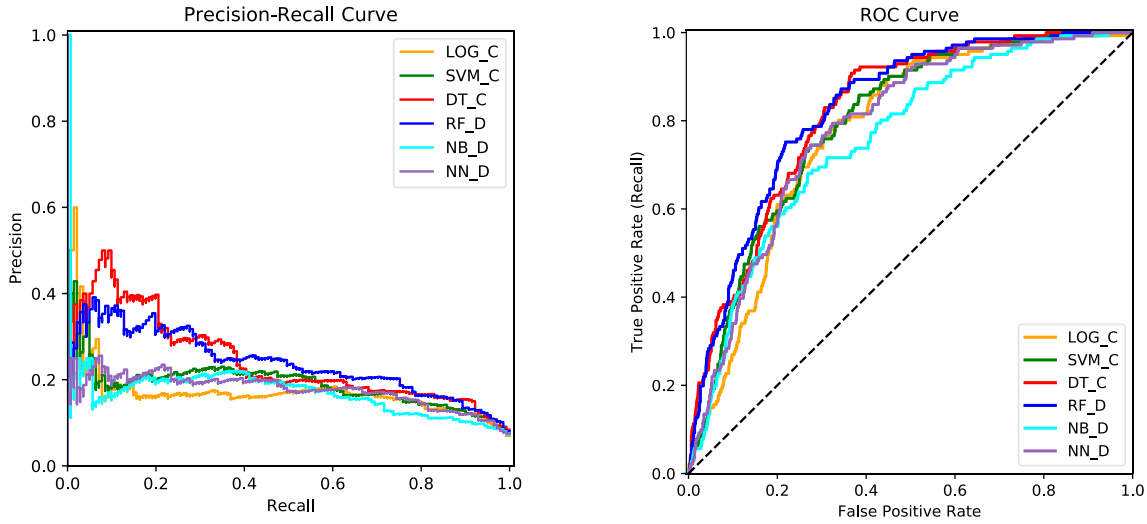


Figure 4: Precision-recall curve and ROC curve comparing the best classifiers for each machine learning technique according to the AUC score

4.3 Economic Evaluation based on the Naturally-Distributed Sample

For the 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, being the 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 losses due to 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 leads to a slightly lower average economic gain of USD 4,280.75. Therefore, classifiers also considering 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 reaching USD 4,695.47 and USD 4,711.58. In particular, classifier D achieves the highest economic value. Thus, classifiers detecting misconduct based on self-disclosed as well as user and regulatory confirmed information are economically valuable, which supports 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. Naïve	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.

In addition to the economic evaluation of the optimized configuration of the classifiers, we perform a sensitivity analysis and 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 potentially committing 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 a realistic level (Egan, 2019; Hortaçsu & Syverson, 2004). For realistic levels of search costs[6], all classifiers add value.

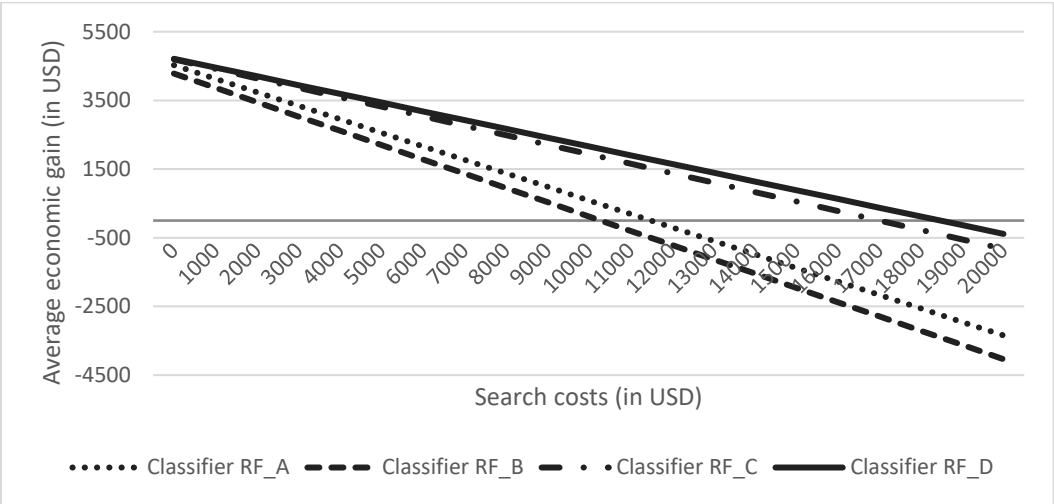


Figure 5: *Sensitivity analysis of different levels of search costs for classifiers A to D*

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

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

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 if compensation payments are high and search costs are low, precision and the avoidance of FP becomes economically more important with higher search costs.

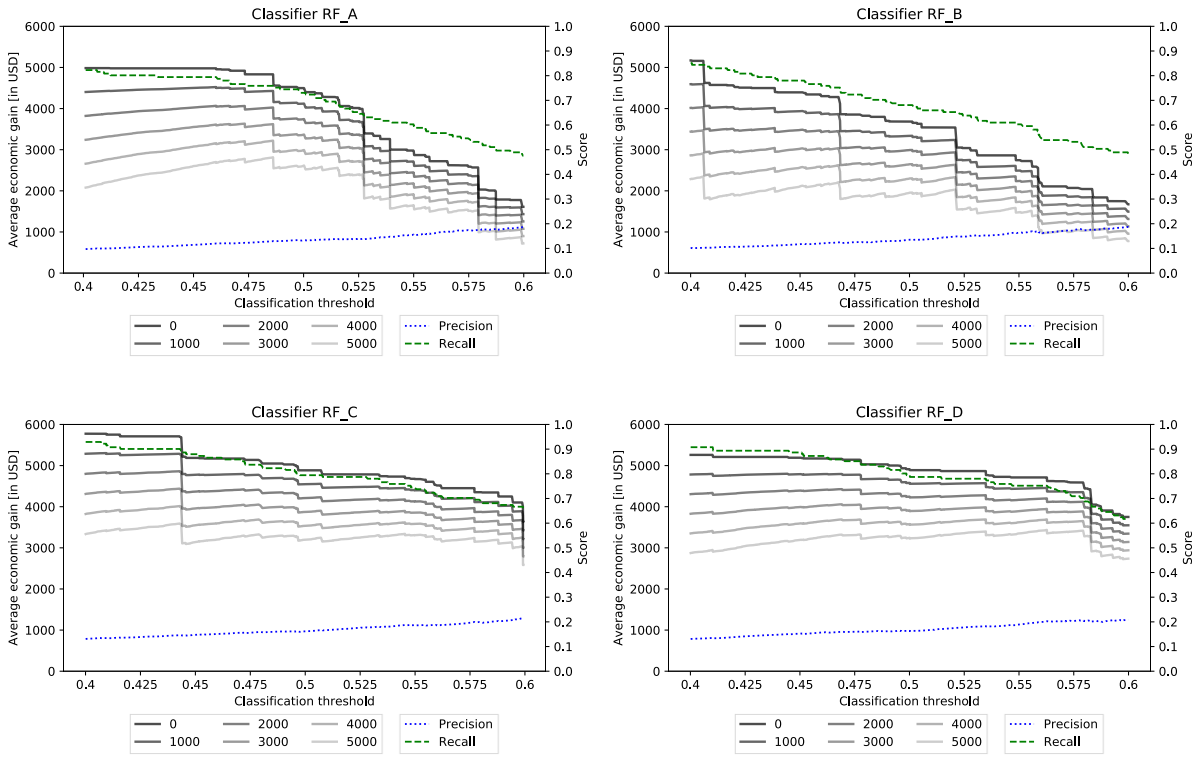


Figure 6: *Sensitivity analysis of economic gain for varying search costs and classification thresholds*

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. Yet, the optimal classification threshold, which leads to the highest possible economic gain of a specific classifier, depends on the amount of search cost being considered. Specifically, the higher the assumed search costs, the higher is 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, that should particularly be avoided in case of high search costs. Nevertheless, for higher classification thresholds, the gain in precision is

not compensating the loss in recall causing lower economic gains. This is especially true for classifiers A and B and needs to be considered when choosing the desired classification threshold. While we observe more jagged lines for classifiers A and B, the lines get smoother for classifiers C and D. The existence of jagged lines and the smoothing from A to D can be explained by the tree-structure of the random forests: more features can 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, with the help of the sensitivity analysis regarding the economic gain of the proposed classifiers, investors and other model users can customize the classifiers 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 classification thresholds can be chosen on a more continuous basis.

In summary, the 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 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 to detect financial intermediaries committing misconduct. Referring to information disclosure, our results show significant differences in means between financial intermediaries with and without misconduct supporting *information manipulation theory*. Further, this self-disclosed information can be used to detect and address intermediary misconduct in financial markets. The proposed approaches achieve a promising classification performance and their application leads to considerable economic gains for the society by preventing misconduct and thus strengthening trust in the financial system. In particular, the results provide evidence that self-disclosed information is valuable to detect financial intermediary misconduct (H1).

Moreover, confirming *warranting theory*, our results show that self-disclosed information with different levels of external verification is valuable to classify intermediaries that do and do not commit misconduct. This is particularly the case 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). Consequently, verifications by reliable third parties such as regulators provide most value for the classification. Still, verifications by third parties such as other users on LinkedIn can lead to a moderate increase in classification performance, and therefore can be valuable for some classifier configurations. Potentially, the performance of classifiers including user confirmed information could be improved if social networks would offer more verification possibilities or incentives to motivate users to provide more mutual verifications.

For all classifiers, the 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 intermediaries committing 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 considering self-disclosed information together with user and regulatory confirmed information still add value for search costs of up to USD 18,477 which are far above the realistic level of costs to identify a new broker or investment advisor even when considering opportunity costs. Moreover, the sensitivity analysis conducted within the economic evaluation shows how investors can customize the classifiers by setting the classification threshold based on their individual level of search costs. Consequently, addressing our research question, classifiers including self-disclosed information with varying levels of external verification are valuable to detect financial intermediary misconduct.

In terms of research methodology, we use a balanced data set for training in order to handle the problem of unequal class distribution of intermediaries with and without misconduct. Our results highlight that the proposed classifiers outperform a naïve classification algorithm as described in Section 4.2.2. Taking into account the historical unbalanced class distribution (only 6.83 percent of the intermediaries actually commit misconduct), a simple classification model would achieve a high accuracy by classifying each intermediary as not committing misconduct. However, recall of such a simple classifier would be zero. Because investors incur immense losses in case of misconduct, it is essential to identify as many intermediaries committing misconduct as practically possible, even if thereby some orderly 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, recall is shown to be very high 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 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 our classifiers fulfill these requirements and show high scores for negative predictive value (between 96 and 98 percent).

From an ethical point of view, our proposed classifiers do not suffer from biases against certain groups or minorities as recently witnessed for machine learning algorithms applied for criminal sentencing in the US (Angwin, Larson, Mattu, & Kirchner, 2016) since we do not include features related to poverty, joblessness, and social marginalization. Moreover, the group of financial intermediaries that is targeted by our proposed machine learning algorithm is very 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 this intermediary more closely but does not directly lead to negative consequences or penalties.

We are aware of certain limitations of our study. For determining financial intermediaries committing misconduct, we rely on disclosures provided by the regulatory authority FINRA's website BrokerCheck. However, intermediaries that have committed misconduct in the past but have not been accused of misconduct do not have a disclosure. 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 data set in case of a yet trustworthy intermediary being accused of misconduct.

Moreover, financial intermediaries might try to 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. Yet, our classifiers cover a wide range of features with varying levels of external verification so that applying countermeasures or the imitation of 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 confirmed information. Thus, when misbehaving intermediaries try to polish their profiles to avoid being detected, classifiers B, C, and D can still identify the misbehaving intermediary due to information which is externally verified by other users on LinkedIn or by the regulator which is hard or even impossible to manipulate (following *warranting theory)*. Consequently, our classifiers should also work in the long run. Nevertheless, the proposed classifiers should regularly be retrained once they are put in place in order to be able 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. Beyond this platform, there are also other business networks such as Maimai, the largest professional social network in China, or 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 these

networks compared to LinkedIn, the proposed classifiers can thus easily be applied to such 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 ensure fair and efficient financial markets and capital provision by investors to corporations. Trust in intermediaries has become particularly important due to increased reliance on electronic communication and less personal interaction between investors and intermediaries, which impedes the process of building trust. Therefore, misconduct by intermediaries needs to be detected and scandals like in case of the "Wolf of Wall Street" have to be avoided to preserve investors from losses and to retain trust in the financial system.

With this paper, we contribute to the literature on financial misconduct and provide evidence 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 are able to detect intermediaries committing misconduct. The best performing classifier including externally verified information is able to detect misconduct among financial intermediaries with a recall of 77.02 percent and an accuracy of 73.74 percent for the balanced training sample and a recall of 75.18 percent and an accuracy of 77.96 percent and for the naturally-distributed validation sample where intermediaries with misconduct represent the minority class with 6.87 percent.

We also contribute to the literature on automated misconduct and fraud detection in general by highlighting the value of self-disclosed information with varying levels of external verification. We show that the provision of self-disclosed information differs between trustworthy financial intermediaries and those committing misconduct. Therefore, our results confirm *information manipulation theory* and provide evidence that self-disclosed information is useful to

classify whether individuals commit misconduct or not. Supporting *warranting theory*, we show that self-disclosed information, which is verified by a third party and thus harder to manipulate, provides additional value to detect 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 for realistic levels of search costs. Furthermore, the sensitivity analysis allows investors to customize the classifiers and optimize the classification results according to their individual level of search costs. Using the classifiers, investors are less likely to be damaged by intermediary misconduct, thereby avoiding severe losses. Moreover, our classifier based on self-disclosed information only already provides sufficient classification accuracy and economic value, so that investors can use the approach 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 which are classified by the system as committing misconduct. Consequently, predictive supervision based on our approach enables authorities to detect potential misconduct earlier and therefore might help to avoid the next “Wolf of Wall Street”. Therefore, the proposed classifiers facilitate investor protection against financial intermediary misconduct, which increases trust in the financial system and is therefore valuable for the society as a whole.

The analysis of self-disclosed information with varying 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 might add value for corporate compliance departments on top of their own, verified data. Moreover, social networks such as Facebook can be investigated, where individuals rather disclose private than job-related information. However, the substantial difference between information provided to friends only and information that is provided to all users would have to be reflected in such an analysis. Future research might also investigate whether classifiers can be developed that are able to detect

specific fraud types committed by intermediaries or whether recidivists that commit multiple misconducts can be identified. Our results show that analytics and machine learning techniques combined with the huge and rising amount of self-disclosed information in social networks can be a powerful combination to provide solutions to important societal challenges.

References

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

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.* Retrieved January 21, 2019, from 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*. Retrieved December 11, 2018, from 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. *Machine 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.). Chichester: 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 *Data Mining and Knowledge Discovery Handbook* (pp. 875–886). Boston, MA: Springer.

Council of Economic Advisors. (2015). *The Effects of Conflicted Investment Advice on Retirements Savings*. Retrieved February 20, 2018, from 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.). New York, NY: 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*. *Monographs on applied probability and statistics*. London: 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.). *The Morgan Kaufmann series in data management systems*. Amsterdam: 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 texts in statistics*. New York, Heidelberg, Dordrecht, London: 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*. New York, NY: 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. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence* (pp. 1137–1143).

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. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 55–60).

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?* FINRA Office of the Chief Economist Working Paper.

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), 1-21 (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. In *Second ACM International Conference on Mobile Software Engineering and Systems (MOBILESoft 2015)* (pp. 99–108). Florence, Italy.

Vapnik, V. N. (1998). *Statistical learning theory*. New York, NY: 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. In *International Conference on Intelligent Computation Technology and Automation* (pp. 50–53).

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.). *ISE - International Student Edition*. Australia: South-Western Cengage Learning.

Zhang, J., Bloedorn, E., Rosen, L., & Venese, D. (2004). Learning Rules from Highly Unbalanced Data Sets. In R. Rastogi (Ed.), *Proceedings of the Fourth IEEE International Conference on Data Mining (ICDM 2004)* (pp. 571–574). Brighton, UK.

Zhang, J., & Mani, I. (2003). kNN approach to unbalanced data distributions: A case study involving information extraction. In *ICML Workshop on Learning from Imbalanced Datasets II* . Washington DC.

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.

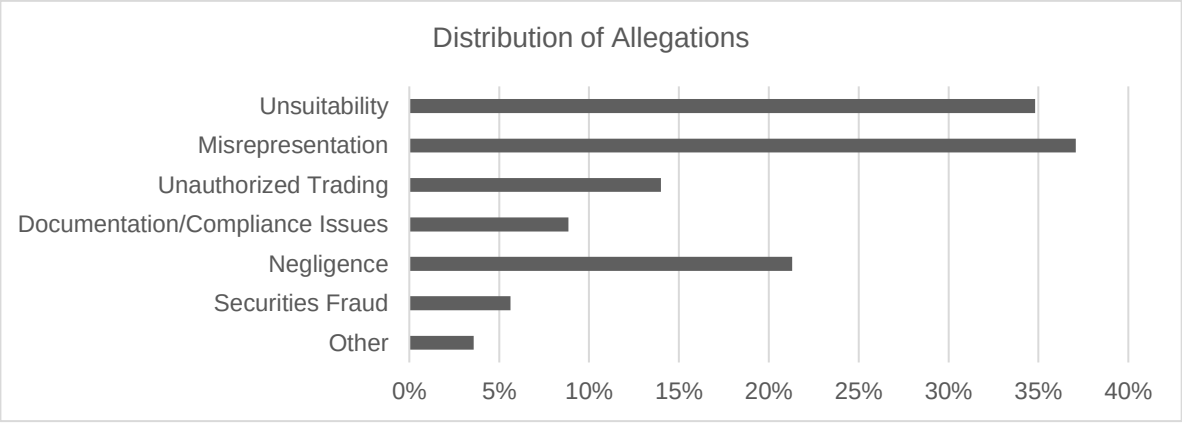
Appendix

A. Descriptive Statistics for Brokers with Misconduct Cases

Table 14: Descriptive statistics for brokers with misconduct cases

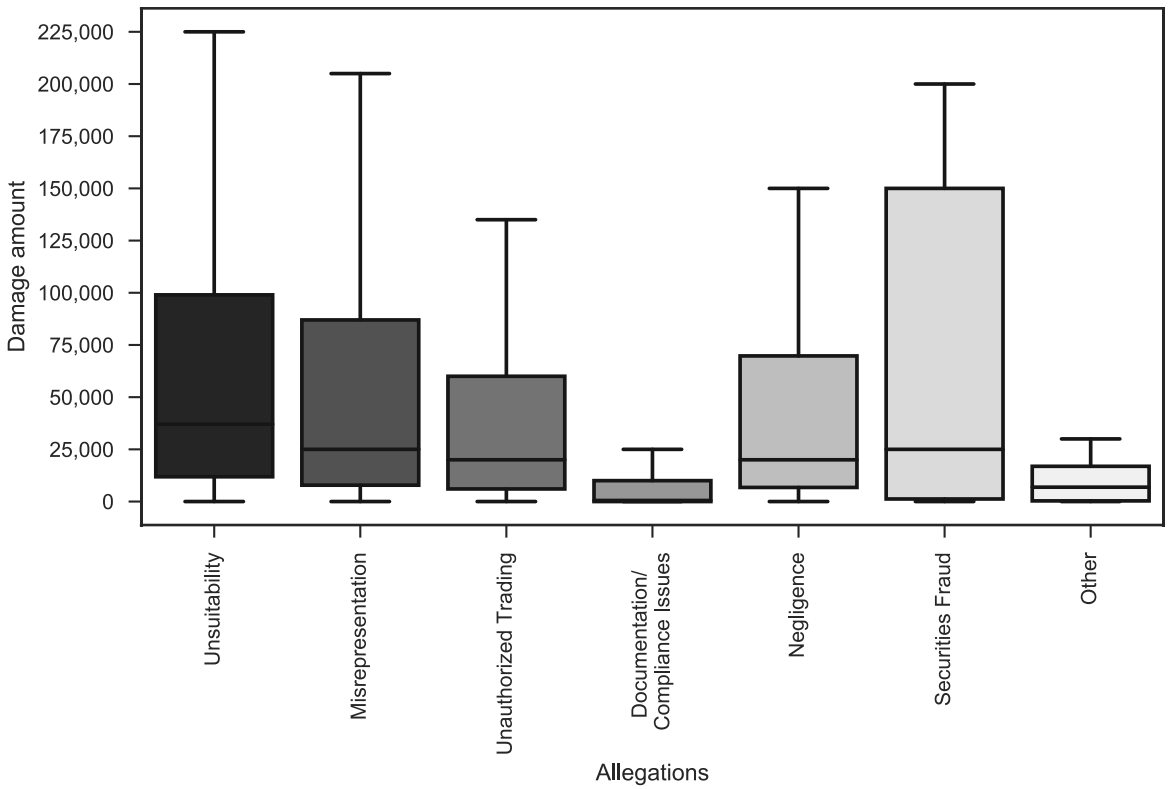
			N = 541						
Feature	**Sum**	**Mean**	**SD**	**Min**	**25%**	**50%**	**75%**	**95%**	**Max**
Disclosures	1,257	2.32	1.97	1.00	1.00	2.00	3.00	6.00	21.00
Customer disputes	983	1.82	1.66	0.00	1.00	1.00	2.00	5.00	21.00
Regulatory actions	109	0.20	0.57	0.00	0.00	0.00	0.00	1.00	5.00
Non job-related disclosures*	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

* 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.



This figure shows the distribution of allegations (types of misconduct) among all observed misconduct cases in our data set. One misconduct case can have multiple allegations, e.g., unsuitability and misrepresentation, therefore the overall sum is not equal to 100 percent. Allegations are categorized by the most prevalent misconduct categories on BrokerCheck. The categories cover the following respective allegations: *Unsuitability* - investment advice unsuitable to the customer's preferences, *misrepresentation* – active misrepresentation or disguise of facts regarding the nature, risks, or fees of a financial product, *unauthorized trading* – trading without permission of the client, *documentation/compliance issues* – practicing without license, failure to document undertaken businesses properly, failure to provide mandatory reportings, *negligence* – failure to execute orders/liquidate assets, failure to maintain/supervise portfolio properly, failure to follow customer's instructions properly, *securities fraud* – gambling with customers' assets, excessive trading, churning, front running, scalping.

Figure 7: Distribution of allegations among all final customer disputes and regulatory actions



This figure 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 data set. Allegations are categorized by the most prevalent misconduct categories on BrokerCheck. The categories cover the following respective allegations: *Unsuitability* - investment advice unsuitable to the customer's preferences, *misrepresentation* – active misrepresentation or disguise of facts regarding the nature, risks, or fees of a financial product, *unauthorized trading* – trading without permission of the client, *documentation/compliance issues* – practicing without license, failure to document undertaken businesses properly, failure to provide mandatory reportings, *negligence* – failure to execute orders/liquidate assets, failure to maintain/supervise portfolio properly, failure to follow customer's instructions properly, *securities fraud* – gambling with customers' assets, excessive trading, churning, front running, scalping.

Figure 8: Distributions of damage amounts per case by allegation category

B. Hyperparameter Tuning

Table 15: 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

* 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).

C. Descriptive Statistics for the Naturally-Distributed Sample

Table 16: Descriptive statistics for the naturally-distributed sample and Wilcoxon Rank-Sum test for equality of means

	Full data set 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

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

Table 16 continued

	Full data set N = 2,051				Misconduct N = 141				No Misconduct N = 1,910				WRS test
Feature	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**	**p**
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***
Reg. 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.

D. Classifier Evaluation

Graphical analysis

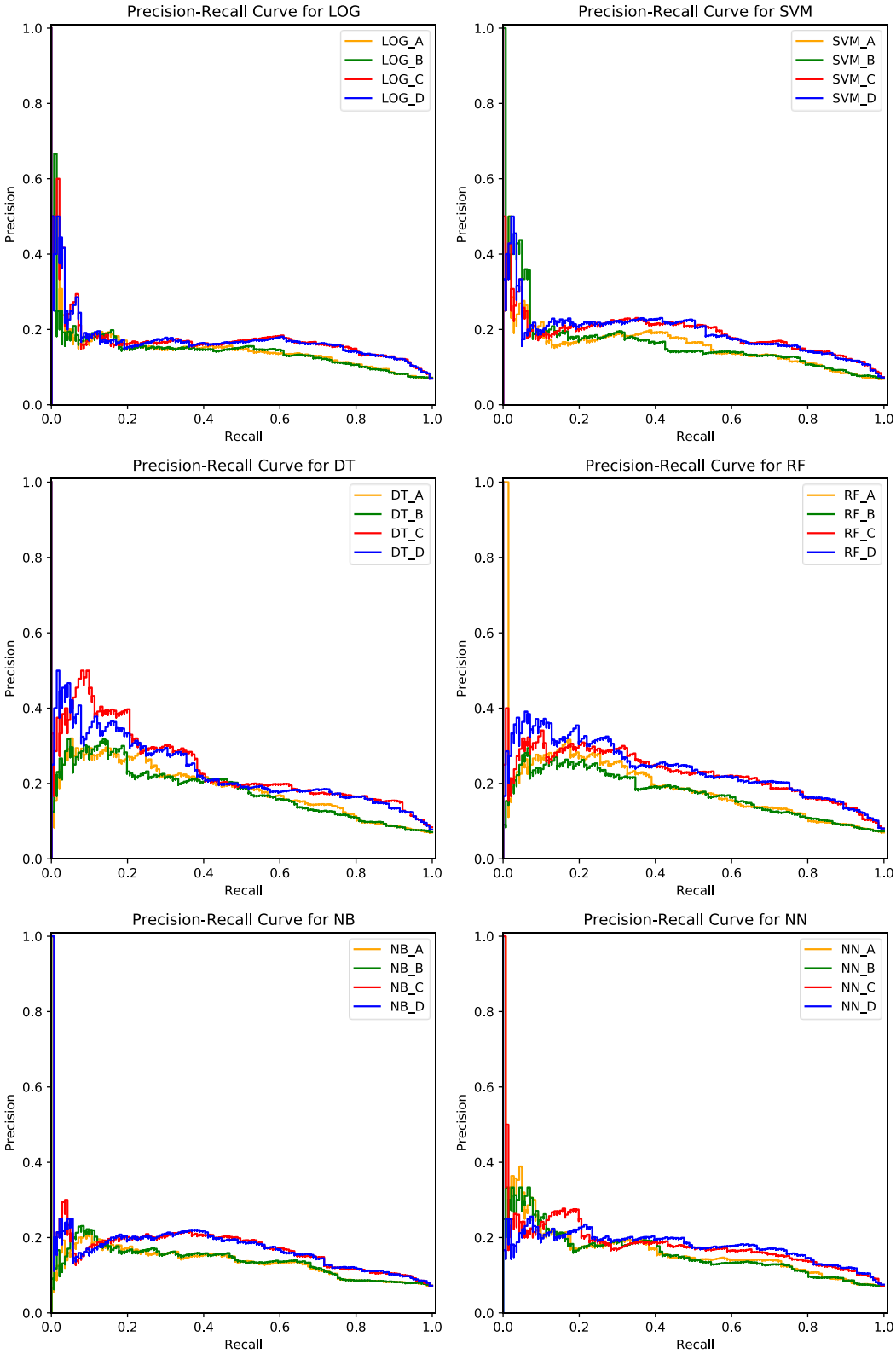


Figure 9: Precision-recall curve for all classifiers and machine learning techniques

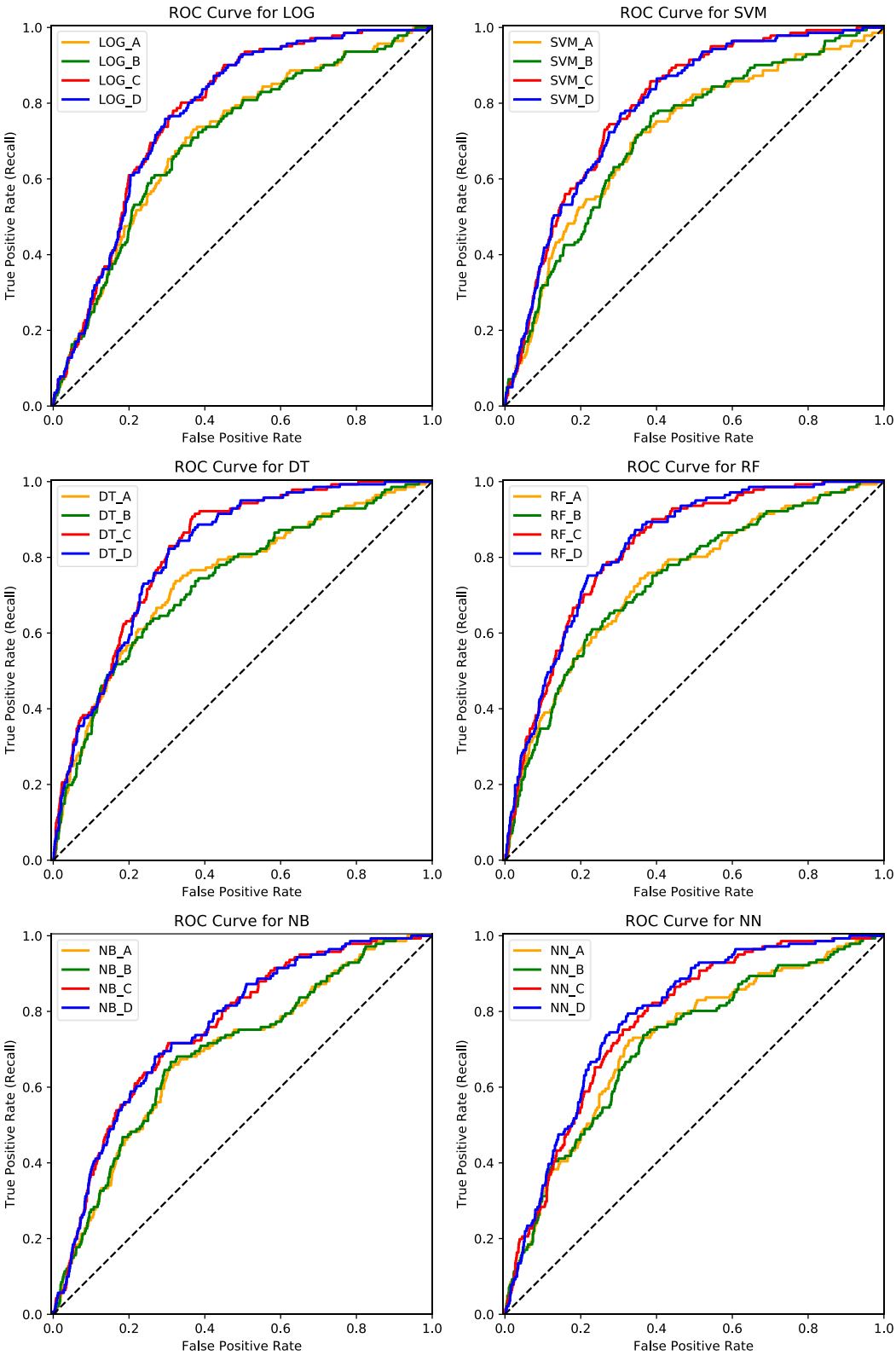


Figure 10: *ROC curves for all classifiers and machine learning techniques*

Classifier evaluation for the additional classifiers

Table 17: Additional classifiers

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

Table 18: Classifier evaluation for classifiers using user and regulatory confirmed information only (in %, training results based on the balanced sample)

	Classifier E					Classifier F				
Cues	Regulatory confirmed information					User & reg. conf. 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 19: Classifier evaluation for classifiers using regulatory confirmed information only and McNemar's test results on classifier performance (in %, naturally-distributed sample)

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

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

Table 20: Classifier evaluation for classifiers using user and regulatory confirmed information only and McNemar's test results on classifier performance (in %, naturally-distributed sample)

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

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

Table 21: McNemar's test results on classifier performance for classifiers using self-disclosed as well as user and regulatory confirmed information compared to classifiers E and F as benchmark (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 include 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 (JAIS), Journal of Information Technology (JIT), and Information Systems Frontiers (ISF).

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 and holds the 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 and 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 Head of Market Development Cash Markets and Xetra Research at Deutsche Börse AG. He graduated in Business Administration and acquired his PhD at the University of Giessen.