Beyond Open Access: Predicting Financial Advisor Misconduct Using Machine Learning

Xin Xu[a], Zhe An[b]

a. School of Business, Sun Yat-sen University, China

b. Monash Business School, Monash University, Australia

August 2024

* For their helpful comments, we thank Chunfang Cao, Sean Cao, Xikai Chen, Shujian Guo, Allen Huang, Sterling Huang, Xian Huang, John (Xuefeng) Jiang, Qinglu Jin, Bin Ke, Dongming Kong, Zhou Lu, Mei Luo, Ting Luo, Bin Miao, Wenqiang Pan, Gordon Phillips, Rui Shen, Zhiyuan Tu, Eric Wang, Lijian Wei, Liping Xu, Jinqiang Yang, Bo Zhang, Tianyu Zhang, and conference and seminar participants at CUHK-Shenzhen Forum of Asian Accounting Scholars 2024, 2024 POMS International Conference in China, Monash University, Sun Yat-sen University, and University of Technology Sydney (scheduled). This paper was previously circulated under the title "*The Missing Information: Predicting Financial Advisor Misconduct Using Machine Learning*." All errors are our own. E-mail: *xuxin67@mail.sysu.edu.cn* (X. Xu); *zhe.an@monash.edu* (Z. An).

Beyond Open Access: Predicting Financial Advisor Misconduct Using Machine Learning

Abstract

This paper employs state-of-the-art machine learning algorithms to assess the adequacy of open-access information for predicting misconduct by financial advisors (FAs). Leveraging the Gradient-Boosted Decision Tree model, we observe that the model, which incorporates 25 input variables sourced from public sources, exhibits a reasonably good predictive performance. Additionally, we show that incorporating supplementary firm (but not county/state or FA) input variables can significantly enhance both the effectiveness and efficiency of our model, in which the overall misconduct status of the advisory firm is a particularly critical factor in this improvement. Further analysis reveals the significance of additional firm input variables in predicting FA misconduct over the next two or three years, the recidivism among FAs with a history of misconduct, and the first instance of misconduct among FAs with a clean record. Lastly, we developed *Advisor Alert* (http://www.luoyaohuang.com:81/advisoralert), an interactive tool that uses our trained machine learning model to predict FA misconduct. Overall, this study offers valuable insights and methodologies for advanced analyses in FA misconduct prediction, providing individuals guidance in selecting reliable and trustworthy FAs.

JEL Classification: C5; G24; D14; D18

Keywords: Machine Learning; Financial Advisor; Financial Misconduct

1. Introduction

Seeking advice from a financial advisor (FA) is prevalent in the United States (U.S.). As per a 2022 retirement survey conducted by the Employee Benefits Research Institute, one in three working adults and retirees currently seek guidance from a professional FA. Among those who now do not have one, almost half expressed their intention to engage with one in the future. Meanwhile, professional misconduct is widespread within the advisory industry, with over 7% of FAs having a documented history of misconduct (Egan, Matvos, and Seru, 2019). Thus, one of the top priorities for individuals selecting their FAs would be to avoid those with a high propensity for misconduct, which elevates their financial risks. However, this is challenging due to the lack of transparency in track records,[1] the knowledge gap between clients and FAs, potential conflicts of interest (i.e., due to compensation structure, such as commissions from selling financial products), varying regulatory standards (i.e., fiduciary standard vs. suitability standard), and inherently difficulty in evaluating ethical conduct. Therefore, understanding the factors that drive FA misconduct serves not only to assist regulators in designing effective industry mechanisms to prevent misbehavior but also offers individuals guidance in selecting reliable and trustworthy FAs.

The availability of a few public data sources offers great convenience and enables users to access FA information with minimal cost. On the one hand, the BrokerCheck and Investment Adviser Public Disclosure (IAPD) provide individual reports for brokers and investment advisors, both referred to as FAs. These reports contain information on the FA's registration history, qualifications, and disclosures (e.g., customer disputes, disciplinary events, and certain criminal and financial matters). A limitation is their restricted provision of FA individual characteristics due to privacy concerns, which are frequently regarded as the primary influencers in explaining personal misbehavior. On the other hand, these data sources also provide similar reports for advisory firms. However, the data is presented in a way that obscures a clear and accurate view of the overall misconduct status of the advisory firm. One concern arises from the fact that certain disclosures, such as pending allegations or settled claims (which may not lead to arbitration), are only accessible at the FA level but not at the firm level, which only reports concluded arbitration. This variance in disclosure standards encourages firms to settle rather than pursue arbitration to its conclusion, which poses a risk of having a bad record

[1] While there are regulatory databases such as FINRA BrokerCheck, not all FAs are required to disclose their complete history of complaints or disciplinary actions, particularly those that were not formalized or settled privately.

on their profiles (Guilford and Kopf, 2020). Thus, to gain a comprehensive understanding of the malfeasance occurring within a particular firm, it is insufficient to rely solely on its profile data. Instead, to obtain an accurate record of firm misconduct, it is necessary to compile the misconduct histories of all individual FAs associated with the firm, including those who have left the industry, and thoroughly analyze the data. Clearly, this is not a straightforward task for individual clients.

In the age of information and technology, enhancing misconduct predictive performance is achievable by embracing the cutting-edge prediction model. Recent advances in machine learning have the potential to untangle these often-nuanced relationships, offering a more timely and accurate means of predicting FA misconduct. This paper endeavors to ascertain if the data available from public sources is adequate for predicting FA misconduct by incorporating state-of-the-art machine learning algorithms and whether/what specific additional information can further improve predictive performance.

In this study, we utilize a widely recognized machine learning model – Gradient-Boosted Decision Tree (GBDT) – to predict FA misconduct.[2] The GBDT model is widely acknowledged for its ability to deliver accurate out-of-sample predictions, particularly excelling in Kaggle competitions.[3] Its strengths include shorter computational time and mitigating the risk of data overfitting (Chen and Guestrin, 2016; Geertsema and Lu, 2023; Guenther, Peterson, Searcy, and Williams, 2023). To implement this model, we gather a sample of 693,195 FAs from 7,136 advisory firms between 2010 and 2019. The dataset is then divided into three subsets: a training sample, a validation sample, and a testing sample. The machine learning models are trained using the designated training sample, and their optimal parameters are selected through the validation sample. Following this, the model leverages the training data covering 2010 to 2016 (2010 to 2017) to identify instances of FA misconduct for 2018 (2019).

Misconduct by a FA is identified when they have at least one misconduct disclosure in a year. This includes customer disputes, internal investigations, and regulatory, civil, and

[2] We also compare the performance of GBDT with other popular machine learning models, including Random Forest, RUSBoost, Logistic Regression, and Artificial Neural Network (ANN). Our findings exhibit that while the models have similar effectiveness, the GBDT model stands out for its superior efficiency and more balanced predictive performance. The results are presented in the robustness section.

[3] Kaggle (*https://www.kaggle.com*) serves as a hub for data scientists and researchers to access public datasets, participate in machine learning competitions to solve data-driven projects, and share insights. It offers tools for data processing, model development and a forum for exchanging ideas and methodologies within the data science community.

criminal events resolved against the FA. The outcome variable is a binary dummy indicating whether a FA commits misconduct the following year. To predict FA misconduct, a set of relevant input variables are identified. Our variable selection begins by considering using information from public data sources, such as IAPD, BrokerCheck, and the SEC's Form ADV filings. Thus, we incorporate 25 easily accessible variables at both FA and firm levels in our baseline model. Obtaining these variables requires minimal effort as they are readily accessible from publicly available sources.

In the initial analysis, our findings reveal that our baseline model with 25 input variables achieves an *AUC* of 0.691,[4] indicating a 69.1% probability that the detected misconduct probability for a randomly chosen misconduct case is higher than that for a case without misconduct. In addition, the *NDCG@k* is 0.030,[5] suggesting that the congruence between the ranking based on the predicted probability of committing fraud and actual misconduct occurrences is 3%. The results of our baseline model indicate that while it displays reasonably good predictive performance, it does not attain an exceedingly high level.

Next, we evaluate the importance of each input variable in contributing to the predictive power of the model. The findings highlight that the FA's gender, the number of FAs within the advisory firm, and the FA's misconduct history are the top three influential variables in predicting FA misbehavior. Following these, the working experience and dually registered status of the FA are identified as the fourth and fifth most significant misconduct predictors. The factors related to a FA's professional qualifications exhibit moderate importance in the prediction model. The formation type and fee structure of the advisory firm play a trivial role in predicting FA misconduct.

Moving forward, we shift our focus to more advanced users who tend to engage in further analyses on the FA and advisory firm or seek additional data at a broader level. The multilevel theory underscores the importance of considering multiple layers of factors to understand complex behaviors (Klein and Kozlowski, 2000). FAs are nested within FA firms, which operate within broader macro environments (i.e., social, economic, and regulatory). Therefore,

---

[4] *AUC* stands for *Area Under the Curve*, representing the total two-dimensional area beneath the ROC (i.e., Receiver Operating Characteristic) curve. It ranges from zero to one, with a higher value indicating a more accurate classifier. The ROC curve is a graphical representation that illustrates the performance of a classification model across all possible classification thresholds. It plots two parameters, including the *True Positive Rate* and the *False Positive Rate*.

[5] *NDCG@k* stands for *Normalized Discounted Cumulative Gain at k*. It compares rankings to an ideal order where all relevant items are at the top *k* of the list. It ranges from zero to one, with a higher value indicating a more efficient classifier.

we incorporate additional county/state, firm, and FA variables into our model to determine whether additional information could improve the model performance in identifying FA prone to misconduct. Specifically, we first incorporate nine additional input variables at the county/state level, bringing the total to 34. Furthermore, we include five additional firm input variables, bringing the total to 30. This entails computing the advisory firm's overall misconduct status, evaluating employee experience, determining the proportion of female FAs, and examining whether the owner/executive of the advisory firm has a history of misconduct. These firm-level measures are not directly accessible in public sources but require additional research effort.[6] Moreover, recognizing that individual characteristics of FAs may offer more profound insights into their misbehavior likelihood, we augment our model with 15 additional FA input variables, leading to a cumulative count of 40 input variables. Finally, we include all the 29 additional input variables together, resulting in a total of 54 input variables. The sample size decreases by approximately 10% and 60% for models incorporating additional county/state and FA input variables, respectively, compared to our baseline model. However, incorporating additional firm input variables does not impact the sample size.

The results indicate that compared to the baseline model, incorporating additional county/state or FA input variables can only slightly improve the *AUC*, while integrating supplementary firm information could significantly enhance the *AUC* from 0.691 to 0.718. These findings imply that incorporating supplementary input variables shows potential for improving the accuracy of our machine learning model, and the most notable improvement is the inclusion of additional firm input variables. Furthermore, the *NDCG@k* of the model incorporating additional firm input variables also exhibits a significant improvement compared to the baseline model, from 0.030 to 0.060. This suggests that conducting further efforts to obtain additional firm information is worthwhile, as it could enhance not only the effectiveness but also the efficiency of our machine learning model. Moreover, we also show that the model integrating all additional input variables achieves the highest *AUC* and *NDCG@k* across all models (i.e., *AUC* = 0.729 and *NDCG@k* = 0.065), showing the potential benefit of incorporating all input variables.

[6] As mentioned in Guilford and Kopf (2020), when asked why FINRA does not provide its own ranking of firms based on their history of misconduct, a spokesperson responded: "*FINRA, in its role as a self-regulatory organization, provides information about firms and registered individuals to help investors make better informed decisions about who they would like to do business with. Rankings are inherently subjective and individuals can disagree regarding the relative importance of different factors. By providing the underlying factual information, FINRA empowers investors to make their own informed decisions*."

Regarding the importance of input variables, our results reveal that additional county/state or FA input variables have a limited contribution in predicting FA misconduct. Conversely, in the model with additional firm input variables, it is observed that the firm-level variable indicating the proportion of FAs engaging in misconduct within the advisory firm (i.e., *Misconduct ratio*) has surpassed all others in importance. This suggests that the overall misconduct status of the advisory firm emerges as the most crucial indicator for misconduct prediction. The results suggest that conducting thorough analyses of the firm and obtaining more difficult-to-access information can enhance both the effectiveness and efficiency of the machine learning model.

Our main analysis focuses on predicting FA misconduct occurring in the next year. In an additional test, we shift our attention to predicting FA misconduct in a longer period, i.e., the next two or three years. Our results show that the *AUC* slightly diminishes with a more extended period across all models. We also find that the model incorporating additional firm input variables surpasses the model integrating all input variables, achieving the highest *NDCG@k* for predicting FA misconduct within the next two or three years. In addition, we note that as the predictive period extends, the cumulative importance of additional firm input variables increases. These findings emphasize the vital role of additional firm information in forecasting FA misconduct over a more extended period.

Next, to evaluate which factors are more critical in predicting recidivism among FAs with a misconduct history, we train the machine learning model using data from FAs with a prior record of misconduct and analyze the importance of each input variable. Our results reveal that the machine learning model's ability to predict recidivism is relatively low, with an *AUC* range of 0.546 to 0.612. However, regarding forecast effectiveness, the model achieves a *NDCG@k* range of 0.095 to 0.180, significantly outperforming our baseline prediction model. Furthermore, we perform similar analyses to predict the first misconduct of FAs with a clean record. Our findings indicate that the *AUC* and *NDCG@k* metrics underperform compared to the baseline model that predicts all types of misconduct.

We perform several additional tests to enhance our analyses. First, we conduct tests employing alternative definitions of misconduct. Our specific focus lies in misconduct initiated by judicial and regulatory authorities, which often entail substantial social costs and consequences. We also examine misconduct that originates from client complaints. The results indicate improved predictive performance with additional input variables, aligning with our main findings. Second, in our study, instances of misconduct represent 0.6% of all observations.

Therefore, we set *k* as 2,000, an amount roughly equivalent to 0.6% of the testing observations. As a robustness test, we examine alternative values for *k*, including 10,000, 5,000, 1,000, and 500. Our findings highlight that when *k* is set to a smaller value (i.e., 1,000 or 500), the model integrating additional firm information could overtake the model with all additional input variables, achieving the highest *NDCG@k* across models. Third, we employ four alternative machine learning models: Random Forest, RUSBoost, Logistic Regression, and ANN to forecast FA misconduct. Overall, our findings exhibit robustness when utilizing alternative machine learning models. Notably, the GBDT model exhibits comparable *AUC* but superior *NDCG@k*. Fourth, we demonstrate that our GBDT model surpasses one of the most advanced and popular large language models, GPT-4o, in accurately predicting FA misconduct. Finally, we have created *Advisor Alert* (*http://www.luoyaohuang.com:81/advisoralert*), an interactive tool based on our machine learning model to forecast FA misconduct. Users can enter details about the FA, the advisory firm, and the county to receive a straightforward prediction of potential future misconduct.

This paper makes several contributions to the existing literature. First, our study contributes to the broader literature on using machine learning techniques in the corporate context. Previous research has employed machine learning to enhance loss estimates for insurance companies (Ding, Lev, and Peng, 2020), assess asset risk premiums (Gu, Kelly, and Xiu, 2020), nominate corporate directors (Erel, Stern, Tan, and Weisbach, 2021), predict future earnings (Chen, Cho, Dou, and Lev, 2022), process information for lending decisions (Liu, 2022), relative valuation and peer firm selection (Geertsema and Lu, 2023), forecast effective tax rates (Guenther, Peterson, Searcy, and Williams, 2023), among others. Our study extends this literature by applying machine learning techniques in the context of the FA industry.[7]

Second, our contribution lies in applying machine learning algorithms to predict individual inappropriate behavior. While previous research has demonstrated the use of machine learning techniques in predicting and detecting improper corporate activities, such as accounting fraud (Cecchini, Aytug, Koehler, and Pathak, 2010; Perols, 2011; Purda and Skillicorn, 2015; Perols, Bowen, Zimmermann, and Samba, 2017; Bao, Ke, Li, Yu, and Zhang,

[7] Drawing from a much more limited sample, several studies have employed machine learning models to detect fraud within the U.S. FA market. For instance, Lausen, Clapham, Siering, and Gomber (2020) show that self-disclosed information from FAs' LinkedIn offers valuable insights for detecting misconduct. Additionally, Lausen and Clapham (2023) delve into the factors influencing FA recidivism. In contrast, our study, using almost the universe of FA data, examines whether publicly accessible sources furnish adequate data to predict future FA misconduct.

2020; Brown, Crowley, and Elliott, 2020; Bertomeu, Cheynel, Floyd, and Pan, 2021; Huang, Kraft, and Wang, 2023; Xu, Xiong, and An, 2023), with a few studies focusing on individual misbehavior like credit card fraud (Dornadula and Geetha, 2019; Ileberi, Sun, and Wang, 2022), our study extends this body of literature by specifically concentrating on the prediction of inappropriate behavior by FAs. We show that the machine learning algorithm can achieve moderate performance in predicting FA misconduct by relying solely on publicly available data. The algorithm's predictive performance improves with increased investment of time and effort in analyzing a broader range of information at the firm level, particularly the overall misconduct status of the advisory firm. Our findings suggest that it is worthwhile for model users to allocate more effort to acquiring the most relevant information to enhance the predictive power of the model.

Third, our research contributes to the prevailing body of literature, reinforcing the established link between FA misconduct and various individual characteristics. For instance, collaborating with colleagues with a history of transgressions (Dimmock, Gerken, and Graham, 2018), possessing criminal records (Law and Mills, 2019), experiencing personal real estate shocks (Dimmock, Gerken, and Van Alfen, 2021), starting a career during recession years (Law and Zuo, 2021), and growing up in counties with less ethical cultures (Clifford, Ellis, and Gerken, 2023) are associated with increased misconduct propensity. Conversely, FAs who complete professional exams focusing on ethics (Kowaleski, Sutherland, and Vetter, 2020) and those certified as top performers (Gerken and Shahraki, 2022) exhibit a diminished likelihood of engaging in misconduct. Our results show that in our baseline model, the FA's gender, misconduct history, experience, and dually registered status are among the most influential factors in predicting misconduct through the machine learning model. This aligns with existing studies employing the linear probability model. Our research thus underscores the noteworthy impact of individual characteristics in shaping the behavior of FAs and their propensity for misconduct. Furthermore, our study reveals that the advisory firm's *Misconduct ratio* assumes a much more pivotal role than individual FA characteristics in predicting misconduct in the model with additional firm input variables. It highlights the significance of considering firm characteristics, particularly the misconduct status of the advisory firm, as a primary factor in the selection of FA. In doing so, our research complements existing literature, which predominantly concentrates on examining the role of individual FA characteristics in influencing FA misconduct.

Fourth, we have developed a user-friendly tool, *Advisor Alert*, that leverages the trained machine learning model, allowing investors to quickly assess the likelihood of FA misconduct and make informed selections. Our results show that publicly available information can indeed be used to predict FA misconduct. When feasible, investors can further enhance their decision-making by gathering additional firm or other information. Investors can choose the most suitable model according to the amount of information they possess, enabling a more accurate evaluation.

The rest of this manuscript is structured as follows: Section 2 outlines the data, sample, and research design. Section 3 unveils the primary findings derived from employing the machine learning model for predicting FA misconduct, while Section 4 furnishes supplementary analyses. Finally, Section 5 offers the concluding remarks for the paper.

2. Data, Sample, and Research Design

2.1 Financial Advisor Data

A FA is usually referred to as an individual who functions as a broker or an investment advisor. Specifically, brokers act as representatives for clients engaged in trading investment products and securities. They are typically employed by broker-dealer firms and are subject to regulation by the Financial Industry Regulatory Authority (FINRA). On the other hand, investment advisors are financial professionals who offer recommendations and furnish financial advice to clients. They operate within registered investment advisor (RIA) firms and are subject to regulation by the Securities and Exchange Commission (SEC) and state securities authorities. A FA can register as a broker, an investment advisor, or choose dual registration, encompassing both roles. Brokers follow a lower suitability standard, recommending products that suit the client's needs. In contrast, Investment advisors adhere to a fiduciary standard, requiring them to prioritize clients' interests over their own. A noteworthy percentage of brokers are also registered as investment advisors.

The BrokerCheck database by FINRA maintains broker data, which can be accessed at *https://brokercheck.finra.org*. On the other hand, investment advisor data is managed by the IAPD database, which is sponsored by the SEC and can be accessed at *https://adviserinfo.sec.gov*. Both datasets utilize the identical Central Registration Depository (CRD) number as unique identifiers for FAs. Although these two data sources offer extensive information about individual FA, they provide limited details at the firm level, furnishing primarily static information. To ensure access to crucial time-variant information for predicting

FA misconduct, our attention is directed toward SEC-registered RIA firms.[8] We are able to acquire significantly more detailed information for these firms directly from the SEC's Form ADV filings. For example, Form ADV filings supply a comprehensive archive of historical annual data for each SEC-registered RIA firm. In contrast, the BrokerCheck and IAPD databases only allow the extraction of static information for active broker-dealer and RIA firms. Furthermore, certain details, such as the compensation/fee structure of a firm, can be accessed through Form ADV filings but may not be available in the BrokerCheck and IAPD databases.

We initiate the data collection by compiling data on SEC-registered RIA firms from SEC's Form ADV filings (accessible at *https://www.sec.gov/foia/docs/form-adv-archive-data#part1*). It provides yearly data encompassing advisory firms' scope (i.e., the number of FAs, accounts, registered states, social network links, and business lines), structure type (i.e., corporation and limited liability), referral status (i.e., whether the advisory firm has referral or financial arrangements with others), client focus (i.e., whether the advisory firm advises non-high-net-worth individuals), and compensation/fee structure (i.e., AUM, hourly, fixed, commission, and performance). We collect an initial sample comprising 11,555 advisory firms.

Then, we gather details about individual investment advisors by accessing the IAPD database. In February 2022, we conducted queries for all potential seven-digit CRD numbers (ranging from 1 to 9,999,999) to retrieve detailed reports for each investment advisor. This approach ensures the creation of a survivorship-bias-free dataset at the FA level, encompassing both presently registered investment advisors and those who have exited the industry. Applying a similar method, we also retrieve information about individual brokers from the BrokerCheck database. This is necessary because a small portion of SEC-registered RIA firms also function as broker-dealer firms, employing brokers not dually registered as investment advisors. The FA reports we obtained from their disclosures include information such as their names, registration details, licenses, employment history spanning the past ten years, misconduct history, and more. By leveraging their first names, we were able to infer the gender of 99.3% of FAs using data from GenderChecker (accessible at *https://genderchecker.com*) and an open

[8] Our sample is limited to SEC-registered RIA firms and does not encompass those registered at the state level and broker-dealer firms. The general rule is that RIA firms with AUM of $100 million or more are required to register with the SEC. However, there are exceptions for firms with less than $100 million AUM, which may still need to register with the SEC if they fall into specific categories. These exceptions include (1) New York RIA firms with $25 million or more in AUM, (2) RIA firms registered under the Investment Company Act of 1940, (3) RIA firms that register in 15 or more states, and (4) online or Robo RIA firms.

Application Programming Interface (API), genderize.io (accessible at *https://genderize.io*). As a consequence, we obtain a sizable individual sample of FAs, totaling 1,529,115 instances.

2.2 Sample and Descriptive Statistics

FINRA mandates that all registered FAs "*disclose customer complaints and arbitrations, regulatory actions, employment terminations, bankruptcy filings, and criminal or judicial proceedings*." Out of the 23 disclosure categories outlined by FINRA, we categorize six as misconduct disclosures. These include *Civil-Final*, *Criminal-Final Disposition*, *Customer Dispute-Award/Judgment*, *Customer Dispute-Settled*, *Employment Separation After Allegations*, and *Regulatory-Final*, following Egan, Matvos, and Seru (2019; 2022). This definition encompasses customer disputes, internal investigations, regulatory actions, and civil and criminal events resolved against the FA. The complete definitions of disclosure are provided in Appendix A. Based on this definition, our outcome variable is constructed as a dummy variable that takes on a value of one if the FA has at least one misconduct disclosure in a year and zero otherwise.

A broad spectrum of information, ranging from readily accessible FA reports to those challenging-to-obtain insights via personal interactions with the FA, can be used for predicting FA misconduct. In the baseline model, we formulate 25 input variables to predict FA misconduct, including 11 FA-level and 14 firm-level variables. These variables are readily accessible from the IAPD, BrokerCheck, and Form ADV filings. Specifically, the 11 FA input variables include measures related to a FA's working experience (*Experience*) and gender (*Female*). Additionally, they encompass variables indicating whether a FA has a misconduct history (*Prior misconduct*), is dually registered as both a broker and an investment advisor (*Dually registered*), and has passed specific professional exams (*Uniform Investment Adviser Law (65), Uniform Combined State Law (66), Securities Agent State Law (63), General Securities Rep. (7), Investment Company Product Rep. (6), General Securities Principal (24),* and *No. of other qualifications*). The 14 firm input variables cover fundamental aspects of the advisory firm's status, including the numbers of advisors, accounts, registered states, social network links, and business lines of the advisory firm (*No. of advisors*, *No. of accounts*, *No. of registered states*, *No. of social network links*, and *No. of business lines*). They also include characteristics such as the advisory firm's formation type (*Corporation* and *Limited liability*), advisory type (*Referral business* and *Retail*), and fee structure (*AUM, Hourly, Fixed, Commission,* and *Performance*). The definitions of each variable are described in Appendix B.

As presented in Appendix C, our initial sample comprises 11,555 advisory firms extracted from the Form ADV filings. Since BrokerCheck and IAPD may eliminate the records of FAs who have not registered within the past decade, we commence our sample from 2010. To ensure data completeness, we exclude the most recent two years, 2020 and 2021, as some complaints or arbitrations filed during this period may not have been resolved at the time of data collection.[9] After merging data on FAs and advisory firms and excluding FAs with missing information to construct baseline input variables, our sample is narrowed down to 693,195 FAs (i.e., 7,136 advisory firms). These FAs contribute to approximately 3.6 million FA-year observations over the sample period from 2010 to 2019.

Table 1 displays descriptive statistics for both the outcome variable and each input variable. Specifically, the average likelihood of a FA engaging in misconduct in a given year is 0.6%. Furthermore, for an average FA in our sample, the probability of having at least one misconduct record is 7.2%, indicating that approximately one in 14 FAs possesses a history of misconduct. The average experience of FAs is 13.402 years since passing their first qualification exam. Female FAs contribute roughly a quarter of the FA-year observations, and most FAs are dually registered as both brokers and investment advisors. Regarding the qualifications of FAs, Individuals who pass the Series 65 and 66 examinations (i.e., *Uniform Investment Adviser Law Exam and Uniform Combined State Law Exam*) are authorized to operate as investment advisors. The Series 63 exam (i.e., *Uniform Securities Agent State Law Exam*) covers the principles of state security regulations, and it is typically required by most states for registered representatives. Successfully completing the Series 7 exam (i.e., *General Securities Representative Exam*) allows individuals to trade various securities products, excluding commodities and futures. The Series 6 exam (i.e., *Investment Company and Variable Contracts Products Representative Exam*) authorizes individuals to sell open-end mutual funds, variable annuities, and insurance. Individuals who pass the Series 24 exam (i.e., *General Securities Principal Exam*) are qualified to supervise and manage branch activities at general securities firms. The summary statistics for these qualification variables exhibit similarities to those reported in prior studies (e.g., Egan, Matvos, and Seru, 2019; 2022).

[Insert Table 1]

[9] We observe that the number of FA misconduct instances in 2020 was approximately half of that in 2019 in our sample. Nevertheless, our analysis yields similar results whether we end our sample in 2020 or 2021.

Regarding firm characteristics, an average advisory firm in our sample has 126 FAs, 5,949 accounts, 12.141 registered states, 2.270 social network links, and 3.960 business lines. Less than half of advisory firms operate in the form of a corporation, and more than half in the form of limited liability. Most advisory firms have referral arrangements with other advisory firms and service non-high-net-worth individuals. Almost all advisory firms charge fees based on AUM, while roughly half of advisory firms charge hourly or fixed fees. Compensation based on commission or performance is not so popular among advisory firms.

2.3 Machine Learning Model

In this study, we employ the GBDT, a widely adopted ensemble learning model, to predict FA misconduct (Bertomeu, Cheynel, Floyd, and Pan, 2021; Geertsema and Lu, 2023). The GBDT model builds on the idea of boosting, where multiple decision trees are trained sequentially. Each subsequent tree in the sequence emphasizes correcting the errors made by its predecessors. Through the iterative addition of decision trees, it discerns the optimal decision rules from the training data, enabling it to make accurate predictions on unseen data (Friedman, 2001). The GBDT model demonstrates superior performance in producing accurate and resilient predictions for classification and regression tasks, particularly within the domains of accounting and finance research (Bertomeu, Cheynel, Floyd, and Pan, 2021; Geertsema and Lu, 2023; Guenther, Peterson, Searcy, and Williams, 2023). We use Scikit-Learn, a powerful and user-friendly Python machine learning module, to construct and train the machine learning model.[10] Scikit-Learn furnishes the foundational framework for machine learning algorithms, allowing us to fine-tune the optimal parameters for each model using a validation dataset.[11] To tackle the issue of sample imbalance in our dataset, we set the *class_weight* parameter to *balanced* in Scikit-Learn to decrease the weight of the majority class and increase the weight of the minority class simultaneously. This adjustment alters the weights assigned to misconduct and non-misconduct cases, promoting a more balanced representation during model training.

In line with established research (e.g., Bao, Ke, Li, Yu, and Zhang, 2020; Bertomeu, 2020; Guenther, Peterson, Searcy, and Williams, 2023), we segment our dataset spanning from

[10] Scikit-Learn originated in 2007 through the initiation of David Cournapeau as a Google Summer of Code project (accessible at *https://scikit-learn.org*) and has since undergone continuous development by a community of volunteers.

[11] The GBDT model is constructed with the parameters set as follows: *n_estimators*, *learning_rate*, and *max_depth* are 135, 0.05, and 32, respectively. These values are determined through a hyperparameter optimization process using the *hyperopt* program, recommended by Guenther, Peterson, Searcy, and Williams (2023). We set the random state to zero in accordance with Bao, Ke, Li, Yu, and Zhang (2020).

2010 to 2019 into three subsets: a training sample (2010-2014), a validation sample (2016-2017), and a test sample (2018-2019). We require a two-year gap (i.e., 2014-2016) between the training and validation samples due to the time-consuming nature of investigating and exposing FA misconduct cases. Implementing this approach helps mitigate the potential issue of utilizing information that is not readily available to predict fraud. The machine learning model undergoes training with the training sample, and optimal parameters are chosen using the validation sample. Subsequently, our machine learning model leverages training data encompassing 2010 to 2016 (2010 to 2017) to predict FA misconduct for 2018 (2019). The adherence to the two-year gap rule is again enforced in the final prediction process.

3. Results and Discussion

3.1 Performance Evaluation Indicators

Misconduct prediction can be conceptualized as a binary classification problem involving the identification of FAs who engage in misconduct and those who do not. In this study, we assess the predictive performance of the machine learning model using the following performance evaluation indicators commonly employed for classification problems.

The *Receiver Operating Characteristics Curve* is a graphical plot that provides insight into the performance of binary classifiers in distinguishing between misconduct and non-misconduct FAs. A classifier is considered more effective when the curve is closer to the top-left corner. The *AUC* represents the area beneath the curve and evaluates the probability that the predicted misconduct probability for a randomly chosen misconduct case is higher than that for a case without misconduct. The *AUC* value ranges from zero to one, with a value of one indicating a perfect classifier that makes no false predictions, while a value of 0.5 suggests a classifier performing the same as random guessing. The higher the value, the more accurate the classifier is deemed to be.

The *Precision rate* assesses the accuracy of the misconduct prediction made by a classifier. It is calculated as the proportion of actual misconduct cases correctly identified by the model to all cases predicted by the classifier as misconduct. The *Recall rate* evaluates the classifier's ability to identify all misconduct cases. It is determined by calculating the proportion of actual misconduct cases correctly identified by the classifier to all actual misconduct cases. A higher *Precision rate* (*Recall rate*) indicates that the classifier is more effective in minimizing false-positive (false-negative) errors. Their values are usually negatively correlated, meaning that as one increases, the other tends to decrease. The *f1 score* is a metric that provides a balance between the *Precision rate* and *Recall rate* by considering both false-positive and false-negative errors. It is calculated using the formula: $2 \times \frac{Precision \times Recall}{Precision + Re}$, and particularly useful in situations where there is an uneven class distribution or when both false positives and false negatives are crucial considerations.

To assess the effectiveness of our misconduct prediction model, we utilize a few additional performance evaluation indicators, including *NDCG@k*, *Precision@k*, and *Recall@k*. These metrics focus on the top *k* observations with the highest predicted probability of committing misconduct. *NDCG@k* is a widely used indicator for evaluating ranking algorithms, commonly applied in contexts such as web search engine algorithms and recommendation algorithms (Järvelin and Kekäläinen, 2002). *NDCG@k* measures the quality of the ranking of the top k items by considering both the relevance and the position of each item in the list and provides a more nuanced assessment of the ranking performance compared to simpler metrics like *Precision rate* and *Recall rate*. To obtain *NDCG@k*, we first calculate *DCG@k* (Discounted Cumulative Gain at *k*) as $\sum_{i=1}^{k} \frac{(2^{rel_i - 1})}{log_2(i+1)}$. Here, rel_i equals one if the i^{th} observation in the ranking list is an actual misconduct case and zero otherwise. The parameter *k* denotes the top *k* observations in the testing dataset with the highest predicted probability of misconduct. *NDCG@k* is computed by taking the *DCG@k* and scaling it with the ideal *DCG@k* achieved when all actual fraud cases are positioned at the top of the ranking list. It provides a standardized measure for comparing ranking performance, especially when actual frauds are prioritized at the top.

We also incorporate two additional performance evaluation metrics by focusing on the top *k* observations with the highest predicted misconduct probability. In particular, *Precision@k* and *Recall@k* gauge the proportion of actual misconduct cases within the top *k* observations relative to *k* and the total number of actual misconduct cases, respectively.

3.2 Out-of-Sample Performance of the Machine Learning Model

In the baseline model, we employ 25 input variables sourced from public sources, including the IAPD, BrokerCheck, and Form ADV filings. The outcome variable is a dummy indicating whether a FA commits misconduct in the next year. Table 2 presents the predictive performance of the GBDT model using the performance evaluation indicators mentioned above. All performance metrics have been averaged across the 2018-2019 test period. As shown in Column 1, the *AUC* of the baseline model is 0.691, indicating a 69.1% probability of correctly identifying a randomly selected misconduct case as more likely than a randomly selected non-misconduct case. In Columns 2-3, we observe a low *Precision rate* of 0.010, while a high *Recall rate* of 0.585. This suggests the model's capacity to accurately recognize 58.5% of actual misconduct instances, and the classifier is more effective in minimizing false-negative rather than false-positive errors. Shifting our focus towards predictive efficiency, Columns 5-7 display the predictive performance focused on the top 2,000 observations (i.e., k = 2,000) with the highest predicted probability of committing misconduct.[12] Specifically, the *NDCG@k* of the baseline model is 0.030, suggesting a 3% alignment between rankings derived from the predicted probability of fraud and the actual occurrences of misconduct. Furthermore, the *Precision@k* is 2.60%, indicating that 2.60% of the top 2,000 observations are actual misconduct cases. Moreover, its *Recall@k* reaches 3.06%, suggesting that 3.06% of actual misconduct cases are accurately identified within the top 2,000 observations. In summary, these findings highlight both the effectiveness and efficiency of the baseline model with 25 baseline input variables in pinpointing cases likely to involve misconduct in the following year.

[Insert Table 2]

Next, we evaluate the importance of each input variable of the baseline model. Following Geertsema and Lu (2023) and Guenther, Peterson, Searcy, and Williams (2023), we assess the significance of each input variable using the *SHAP value* (i.e., SHapley Additive exPlanations). The *SHAP value* of an input variable reveals the extent to which it shifts the expected model output from a baseline expectation, typically represented by the mean of model predictions in the training data, towards the specific prediction made by the model. For any given observation, the sum of *SHAP values* across all input variables, when added to the baseline expectation, yields the predicted target value.

Figure 1.1 illustrates the *SHAP value* of each input variable in a horizontal bar graph. We report the average *SHAP values* for the testing years 2018 and 2019. The y-axis represents each input variable, while the x-axis signifies the corresponding *SHAP value*. The graph highlights that the FA's gender (*Female*), the number of FAs of the advisory firm (*No. of Advisors*), and the misconduct history of the FA (*Prior Misconduct*) stand out as the top three significant misconduct predictors, with *SHAP values* of 0.245, 0.241 and 0.212, respectively. This aligns

[12] In our study, misconduct cases make up 0.6% of the total observations. Considering sample sizes of 371,510 and 372,856 for the testing years 2018 and 2019, respectively, we set k at 2,000, representing approximately 0.6% of the testing observations. This choice is in line with Bao, Ke, Li, Yu, and Zhang (2020), who use a cutoff of 1%, reflecting the proportion of fraudulent firm years being approximately 1% of all observations in their sample.

with consistent observations in previous studies (Egan, Matvos, and Seru, 2019; 2022). In succession, the working experience (*Experience*) and dually registered status (*Dually registered*) of the FA rank as the fourth (*SHAP value* = 0.118) and fifth (*SHAP value* = 0.111) most influential variable in predicting FA misbehavior. Factors associated with the FA's professional qualifications, such as S6, S65, and S24 exams, demonstrate moderate importance in the prediction model. Lastly, the formation type and fee structure of the advisory firm exhibit minimal influence in predicting FA misconduct.

Figures 1.2 and 1.3 present the summary plot for the testing years of 2018 and 2019, respectively. Specifically, these figures illustrate how each input variable influences FA misconduct. Every point on the plot corresponds to the *SHAP value* of an input variable for a particular training observation. The color of each point offers insights into the impact of each input variable on the predictive output of FA misconduct, where red points with positive *SHAP values* (or blue points with negative *SHAP values*) suggest a positive relationship between the input variable and the predictive outcome, and vice versa. Our results reveal that the most influential predictor, *Female*, shows an inverse relationship with misconduct prediction, implying a lower inclination among female FAs to engage in inappropriate behavior. *No. of Advisors* and *Prior Misconduct* are positively linked to the likelihood of predicting misconduct, suggesting that FAs who work in large advisory firms or have a history of misconduct are more prone to misconduct behavior. Additionally, *Experience* and *Dually registered* also exhibit a positive influence on predictive outcomes, aligning with findings from existing research (Egan, Matvos, and Seru, 2019; 2022).

[Insert Figure 1]

3.3 Additional Input Variables

The input variables utilized in the baseline model can be readily obtained from public sources with minimal effort. With an *AUC* of 0.691 and a *NDCG@k* of 0.030 using the baseline model, we achieve a reasonably good predictive performance, although not significantly outstanding. However, whether this information is adequate for accurately predicting FA misconduct depends on various factors, such as the specific context and the tolerance for false positives and false negatives. There is potential to enhance its accuracy and robustness by incorporating additional input variables at various levels. These supplementary variables, albeit requiring additional effort to collect, may offer deeper insights into the predictors of misconduct. Ultimately, the decision to include additional input variables should be weighed against the incremental improvement in predictive performance and the practicality of obtaining and processing the data.

In an exploration of whether additional information enhances the model performance of predicting FAs prone to misconduct, we augment our baseline model by incorporating additional input variables based on the multilevel theory. The multilevel theory is a framework widely utilized in social science and psychology to understand phenomena that operate across multiple levels of analysis. This theoretical approach underscores the importance of considering various layers of factors to achieve a comprehensive understanding of complex phenomena (Klein and Kozlowski, 2000). Any single-level perspective, whether focused on micro or macro aspects, is insufficient to comprehend multifaceted issues fully. Similarly, FA misconduct behavior is caused by multiple factors. FAs are embedded within FA firms, which themselves are situated within larger, encompassing macro contexts, including social, economic, demographic, and regulatory environments. Thus, grounded in the multilevel theory, we enrich our baseline model by incorporating additional input variables at the county/state, firm, and FA levels separately and jointly.

We first augment our baseline model by integrating nine additional input variables. These include eight county and one state input variables, resulting in a total of 34 input variables. Specifically, we gather county-level demographic, economic, and social characteristics from the American Community Survey (accessible at *https://www.census.gov/programs-surveys/acs*). For each county, the collected information includes total population (*Population*), median household income (*Med. household income*), median age (*Med. age*), gender ratio (*Male/Female*), percentage of the white population (*% white*), foreign-born population (*% born foreign*), population with low English proficiency (*% less than "very well" English*), and population with degrees higher than a bachelor (*% bachelor+*). In addition, we obtain data on government administration spending on judicial and legal matters by each state (*Govt legal spending (in Thousands)*) from the annual surveys of State and Local Government (accessible at *https://www.census.gov/en.html*).

In our baseline model, we gather 14 firm input variables from the Form ADV filings, primarily focusing on the advisory firm's size, formation type, and fee structure. However, apart from *No. of Advisors*, none of these variables ranks among the top five predictors. The cumulative *SHAP value* of 14 firm input variables represents only 35.5% of the cumulative *SHAP value* of all 25 input variables. Consequently, we shift our attention to several additional firm input variables. Acquiring these variables necessitates gathering extra information on all

FAs operating within the advisory firm; thus, they are not directly available to the public but require extra research effort, especially for large advisory firms. As a result, we augment our model by introducing five additional firm input variables, thereby increasing the total to 30 input variables. This process includes computing the advisory firm's overall misconduct status, including metrics such as *Misconduct ratio* and *Prior misconduct ratio*, assessing employee experience (*Experience*), and determining the proportion of female FAs (*Female ratio*). Additionally, it entails investigating whether the owner/executive of the advisory firm has a history of misconduct (*Owner/Executive prior misconduct*).

Insights could potentially be enriched by collecting additional information on the personal traits of FAs. Additional individual-level data could be acquired through further data research or in-depth interactions with FAs to access information that is not publicly disclosed. Consequently, we enhance our model by introducing 15 additional FA input variables, bringing the total to 40. These indicators encompass whether the FA possesses a personal website (*Personal website*), the number of languages the FA speaks (*No. of language*), and various aspects of the FA's financial status (*Income*, *Wealth*, *Investment*, *Home owner, Swimming pool*, and *Recreational vehicle*). Additionally, they include whether the FA discloses any interests (*Interest*) or contributions (*Contribution*), has military experience (*Military*), operates a Soho business (*Soho business*), owns a credit card or pets (*Regular credit card*, *Premium credit card*, and *Pets*). We supplement these FA characteristics using data from Discovery Data (*https://discoverydata.com*). It is important to note that Discovery Data only offers information for a subset of currently active FAs, resulting in a substantial reduction in the sample size.

As shown in Appendix C, our baseline sample, utilizing 25 input variables, comprises 693,195 FAs (7,136 advisory firms). Integrating additional firm input variables does not affect the sample size. However, upon incorporating additional county/state or FA input variables, the sample sizes decrease to 653,932 or 201,661 FAs (6,955 or 5,934 advisory firms). Furthermore, conducting a test with all additional input variables further diminishes our sample to 189,491 FAs (5,802 advisory firms).[13]

[13] As the number of input variables has increased, the hyperparameters of the GBDT model undergo a refinement process using the validation sample. With the inclusion of additional county/state, firm, and FA variables, the GBDT model is constructed with specific parameters:

n_estimators set to 110, 125, 75, and 185.

learning_rate set to 0.059, 0.054, 0.007, and 0.064.

max_depth set to 19, 25, 45, and 3.

These values have been determined through the fine-tuning results.

Table 2 presents the predictive performance of the model with additional input variables at various levels. We have uncovered several interesting findings. First, compared to the baseline model, there is a slight enhancement of the *AUC* from 0.691 to 0.695 in the model with additional county/state or FA input variables. Remarkably, the *AUC* of the model incorporating additional firm input variables demonstrates a considerable improvement to 0.718. These results suggest that integrating supplementary input variables holds promise for improving the accuracy of our machine learning model, and the most significant enhancement is observed with the inclusion of additional firm input variables. Second, the *NDCG@k*, *Precision@k*, and *Recall@k* of the model incorporating additional firm input variables increase to 0.060, 4.45%, and 5.24%, representing a notable improvement compared to the baseline model of 0.030, 2.60%, and 3.06%. It indicates that integrating supplementary firm information could enhance not only the effectiveness but also the efficiency of our machine learning model. However, incorporating additional county/state or FA input variables does not yield such a significant improvement. Third, the model with all additional input variables achieves the highest *AUC* of 0.729 and *NDCG@k* of 0.065. This suggests that incorporating all input variables could potentially further enhance the effectiveness and efficiency of our machine learning model.[14]

In Figure 2.1, the *SHAP values* of input variables are portrayed in a horizontal bar graph derived from the model incorporating additional county/state input variables. While baseline indicators, such as *Female*, *Prior misconduct*, and *No. of advisors* continue to rank among the top three important input variables, county/state input variables show a considerably less significant role in predicting FA misconduct. The top-rank county/state input variable, *Govt legal spending*, has a *SHAP value* that merely ranks the eighth amount of all input variables. Column 1 of Table 3 shows that the cumulative *SHAP value* of county/state input variables represents only 13.24% of the cumulative *SHAP value* of all input variables. The findings are similar in Figure 4.1 when incorporating additional FA input variables. Except for *Personal website*, there is no additional FA input variable ranks in the top 10 list. Nevertheless, the *SHAP value* of *Personal website* is merely 0.021. As Column 1 of Table 3 reports, the combined *SHAP value* of additional FA input variables accounts for only 8.23% of the cumulative *SHAP value* of all input variables. These findings indicate that additional county/state or FA input

[14] One concern is that the sample sizes are significantly reduced for models incorporating additional input variables, potentially leading to an unfair comparison between models. To address this issue, we conduct an additional check by utilizing the sample from the model with all input variables to evaluate the performance of all other models. The results remain consistent.

variables play a limited role in predicting FA misconduct, while the input variables in the baseline model remain predominant in these two models.

Next, we shift our focus to additional firm input variables. As shown in Figure 3.1, with the incorporation of these variables, the advisory firm's *Misconduct ratio* emerges as the most crucial indicator for predicting misconduct, surpassing all other input variables. It holds the highest *SHAP value* of 0.299, exceeding the second-highest *SHAP value* of 0.247 for *Female*. Some other firm input variables also exhibit moderate importance in predicting FA misconduct. For instance, the advisory firm's *Prior misconduct ratio* and *Experience* demonstrate *SHAP values* of 0.088 and 0.084, ranking seventh and eighth among all input variables, respectively. Column 1 of Table 3 shows that the combined *SHAP value* of these five additional firm input variables accounts for 29.33% of the cumulative *SHAP value* of all input variables. These results suggest that these additional firm input variables, especially the advisory firm's *Misconduct ratio*, play a crucial role in predicting FA misconduct. Figures 3.2 and 3.3 show a positive correlation between the *Misconduct ratio* and predictive outcomes, suggesting that FAs employed by firms with higher misconduct ratios are more likely to be involved in inappropriate behavior. The observed positive correlation aligns with the conclusions drawn by Egan, Matvos, and Seru (2019). The results lend support to the findings of Dimmock, Gerken, and Graham (2018), which suggest that misconduct tends to spread within the advisory firm.

[Insert Figures 2-4]

[Insert Table 3]

Figure 5.1 reveals similar results with the incorporation of all additional input variables, in which supplementary firm input variables play a much more critical role than county/state or FA input variables. Column 1 of Table 3 shows that the cumulative *SHAP value* of additional firm input variable accounts for 23.13% of the cumulative *SHAP value* of all input variables, which is much higher than additional county/state (cumulative *SHAP value* = 10.43%) or FA input variables (cumulative *SHAP value* = 8.64%).

[Insert Figure 5]

In summary, these findings suggest that obtaining more challenging-to-acquire firm information has the potential to significantly improve the effectiveness and efficiency of the machine learning model. Our results indicate that not all supplementary information is equally valuable for predicting FA misconduct. The overall misconduct status of the advisory firm emerges as a crucial factor in predicting FA misconduct, warranting that investors prioritize their efforts in acquiring this information. In comparison, the incremental benefit of macro and FA information is relatively limited. Therefore, when investors face constraints on time and resources, they should focus primarily on gathering the most relevant information in forecasting FA misconduct.

3.4 Longer Predictive Period

Misconduct prediction is centered around issuing early warnings about potential future misconduct before it happens. This is crucial not only for regulators to prevent misconduct proactively but also for individuals to make informed decisions when choosing their FAs. In our baseline model, we anticipate FA misconduct within the following year. Nonetheless, a more extended predictive timeframe could be pertinent in specific contexts. For example, projecting misbehavior over a longer duration may hold greater relevance for clients seeking extended-term partnerships. Therefore, we train and assess our machine learning model using input variables from the current year to predict FA misconduct in the next two or three years.

Panels A and B of Table 4 report the performance of the machine learning model in predicting FA misconduct over the subsequent two and three years, respectively. The results indicate that the overall predictive accuracy slightly diminishes with a longer period. Specifically, in comparison to the baseline model's prediction for FA misconduct within the next year (with an *AUC* of 0.691), the *AUC* decreases slightly to 0.690 (0.685) for the baseline model predicting FA misconduct within the next two (three) years. Similar trends are observed for other models incorporating additional input variables. In terms of forecasting efficiency, the model incorporating additional firm input variables achieves the highest *NDCG@k* of 0.074 (0.082) for predicting FA misconduct within the next two (three) years. Notably, this figure surpasses even that of the model utilizing all input variables. These results suggest that while the predictive effectiveness diminishes slightly with a longer predictive period, additional firm input variables remain to play an important role in enhancing the efficiency of the prediction model.

[Insert Table 4]

For the models predicting FA misconduct over the following two and three years, we also generate the *SHAP value* of each input variable. The outcomes are consolidated in Table 3. We observe that as the predictive period lengthens, the cumulative importance of additional firm input variables increases. For instance, in the one-year prediction model incorporating additional firm input variables, the cumulative *SHAP value* of these variables represents 29.33% of the total cumulative *SHAP value* of all input variables. This figure rises to 29.93% in the two-year prediction model and further to 30.50% in the three-year prediction model. Similarly, in the one-year prediction model encompassing all additional input variables, the cumulative *SHAP value* of additional firm input variables accounts for 23.13% of the total cumulative *SHAP value* of all input variables. This percentage escalates to 24.78% in the two-year prediction model and eventually to 25.47% in the three-year prediction model. As previously observed, these results are predominantly influenced by the *Misconduct ratio* of the advisory firm. These findings underscore the growing significance of additional firm input variables in predicting FA misconduct over a longer period.

3.5 Predicting Recidivism and First Misconduct

In the baseline model, we observe that a FA's history of prior misconduct stands out as one of the most crucial variables for predicting misconduct. This discovery is consistent with earlier studies employing the linear probability model, indicating that a history of misconduct is among the most influential factors linked to FA misconduct. Notably, research by Egan, Matvos, and Seru (2019) has shown that FAs with prior misconduct are five times more likely to be involved in new misconduct than the average FA. Thus, when individuals are selecting their FAs, it is advisable to avoid choosing someone with a history of misconduct. However, finding FAs with clean records may not always be feasible due to factors such as the demand for specific specialists, location constraints, fee considerations, and more. In our sample, one in 14 FAs has a history of misconduct. Individuals might find themselves in the situation of selecting among FAs with prior misconduct. This raises the question: what factors are more crucial in predicting recidivism? To address this, we train the machine learning model using data from FAs with a prior history of misconduct.

The findings presented in Panel A of Table 5 indicate that the machine learning model's effectiveness in predicting recidivism is relatively low, as evidenced by an *AUC* range of 0.535 to 0.612. This performance is notably inferior to our baseline model for predicting FA misconduct and only marginally superior to random guessing. However, when evaluating forecast efficiency, the model achieves a *NDCG@k* range of 0.095 to 0.180, significantly surpassing our baseline model for predicting FA misconduct. It is also worth noting that the model with additional firm input variables outperforms other models with varying input variables in both *AUC* and *NDCG@k* metrics, highlighting its important role in predicting recidivism.

[Insert Table 5]

While a history of misconduct emerges as the most influential variable in predicting FA misconduct, it does not inherently preclude the possibility that FAs with a previously unblemished record may engage in such behavior. Therefore, it is essential to examine whether our predictive model can effectively identify instances of misconduct among FAs who have no prior history of such behavior. To address this, we train our machine learning model using data exclusively from FAs without any previous misconduct history.

As shown in Panel B of Table 5, the machine learning model's effectiveness in predicting the first instance of misconduct surpasses its effectiveness in predicting recidivism, as indicated by an *AUC* range of 0.631 to 0.680. However, it underperforms relative to our baseline model. When assessing forecast efficiency, the model attains a *NDCG@k* range of 0.018 to 0.039, which is significantly lower than that of the model for predicting FA recidivism and our baseline model. Similarly, the model incorporating all additional input variables outperforms other models with varying input variables in both *AUC* and *NDCG@k* metrics.

4. Additional Analyses

4.1 Alternative Misconduct Definition

In our baseline analyses, our primary aim is to forecast FA misconduct, assisting regulators or individuals in assessing the probability of an FA engaging in misconduct. In line with existing literature, we classify six types of disclosures as misconduct, including *Civil-Final*, *Criminal-Final Disposition*, *Customer Dispute-Award/Judgment*, *Customer Dispute-Settled*, *Employment Separation After Allegations*, and *Regulatory-Final*. Recognizing the varying implications of misconduct types, we conduct an additional test by narrowing our focus to misconduct initiated by judicial and regulatory authorities (i.e., *Civil-Final*, *Criminal-Final Disposition*, and *Regulatory-Final*). These misconducts often carry more significant social costs and consequences compared to those initiated by clients or the firm.

As shown in Panel A of Table 6, the overall predictive performance is improved when predicting misconduct initiated by judicial and regulatory authorities compared to our model predicting all types of misconduct. This is evidenced by an enhanced *AUC* range of 0.738 to 0.795 and a *NDCG@k* range of 0.040 to 0.176. The results also affirm an enhancement in predictive performance with an increase in input variables, aligning with our main analysis conclusions. For instance, the baseline model exhibits an *AUC* of 0.738, which improves with the inclusion of additional input variables, reaching 0.743-0.745 for models with additional state/county, firm, and FA input variables and 0.795 for the model that incorporates all additional input variables. Similarly, other performance metrics, including *NDCG@k*, *Precision@k*, and *Recall@k*, display noteworthy improvements when incorporating additional input variables. The baseline model shows a *NDCG@k* of 0.040, *Precision@k* of 0.80%, and a *Recall@k* of 6.18%, which increase to 0.176, 1.85%, and 29.74%, respectively, in the model with all additional input variables.

[Insert Table 6]

Furthermore, we conduct an additional test by focusing exclusively on misconduct initiated by clients, specifically through customer complaints. This analysis aims to elucidate the underlying factors driving customer complaints.

Panel B of Table 6 shows that the overall performance in predicting customer complaints exceeds that of our baseline model in predicting all forms of misconduct. This is indicated by an *AUC* range of 0.726 to 0.837 and a *NDCG@k* range of 0.026 to 0.158. Furthermore, the predictive performance improves with the inclusion of additional input variables. Specifically, the model incorporating additional firm-specific input variables achieves the highest values for both *AUC* and *NDCG@k* metrics. Overall, our results indicate that the machine learning model demonstrates stronger predictive power when forecasting specific types of misconduct.

4.2 Alternative *k*

In our study, misconduct instances represent 0.6% of all observations. Given an average testing sample size of approximately 370,000 in the baseline model, we choose *k* as 2,000, which corresponds to roughly 0.6% of the testing observations. This selection closely resembles that of Bao, Ke, Li, Yu, and Zhang (2020), who adopt a threshold of 1%, reflecting the proportion of fraudulent firm years at approximately 1% of all observations in their dataset. To ensure the robustness of our findings, we also explore alternative values for *k* at 10,000, 5,000, 1,000, and 500.

The results are presented in Table 7, revealing several key insights. First, when *k* is set to 1,000 or 500 (10,000 or 5,000), the model incorporating additional firm (all additional) input variables demonstrates superior performance, exhibiting the highest *NDCG@k* across models with alternative input variables. Second, when *k* is set to 10,000, our model achieves peak efficiency. For instance, among models with varying *k* and input variables, the model integrating all additional input variables records the highest *NDCG@k* of 0.175. In summary, our results offer additional insights into the role of additional firm input variables. It demonstrates that these variables are particularly impactful in enhancing the efficiency of our machine learning model when *k* is set to a smaller value.

[Insert Table 7]

4.3 Alternative Machine Learning Model

In this section, we employ four alternative machine learning models to forecast FA misconduct: Random Forest (Bertomeu, 2020), RUSBoost (Bao, Ke, Li, Yu, and Zhang, 2020), Logistic Regression (Bao, Ke, Li, Yu, and Zhang, 2020; Chen, Cho, Dou, and Lev, 2022), and ANN (Bertomeu, Cheynel, Liao, and Milone, 2024). Specifically, the Random Forest model is a widely used machine learning algorithm for both classification and regression tasks. It operates by constructing a multitude of decision trees during training and outputting the mode (for classification) or mean prediction (for regression) of the individual trees. Random Forest excels with large, high-dimensional datasets due to its robustness against overfitting. It achieves this by training each tree on a random subset of both the data and features, fostering diversity and enhancing prediction accuracy through decorrelation. RUSBoost, or Random Under-Sampling Boosting, integrates boosting principles with random undersampling of the majority class to tackle class imbalance issues in binary classification tasks within machine learning. RUSBoost excels in managing imbalanced datasets by boosting weak classifiers iteratively and addressing class imbalances through random undersampling. This strategy significantly enhances predictive performance, especially when the minority class is poorly represented. Logistic Regression is a powerful statistical method used for binary classification problems. Unlike linear regression, which predicts continuous values, logistic regression estimates the probability that a given instance belongs to a particular category. It is straightforward and interpretable, making it a popular choice for problems where the relationship between features and binary outcomes needs to be explained. ANN is inspired by biological neural networks and consists of interconnected nodes organized in layers. Each connection has an associated weight that could be adjusted during the training process. ANN excels at capturing complex, non-linear relationships within data, making them highly effective for both classification and regression tasks.

Panels A of Table 8 report the results employing the Random Forest, RUSBoost, Logistic Regression, and ANN models, respectively. These results align with our primary tests using the GBDT model. Specifically, incorporating firm input variables notably enhances *AUC* and *NDCG@k*. When comparing model performances, we observe that the Random Forest (the other three models) slightly outperforms (underperforms) the GBDT model in *AUC*, while all alternative models generally lag behind in *NDCG@k*. This observation is evident in the baseline model and the model with additional county/state input variables in Panel A, the models with additional firm or FA variables in Panels B and C, and all models in Panel D. These results indicate comparable effectiveness among the models; however, the GBDT model demonstrates superior efficiency and exhibits more balanced predictive performance. Overall, we show that our results are robust when using alternative machine learning models.

[Insert Table 8]

4.4 Comparing with GPT-4o

In recent years, artificial intelligence tools based on large language models have emerged in various applications. Their ability to understand and generate human-like text has made them invaluable assets in both commercial and academic settings, enhancing productivity, accuracy, and innovation across diverse sectors. It is worthwhile to assess whether the large language model can perform the same task or potentially even outperform our existing method.

To compare the performance of the large language model with our machine learning model in predicting FA misconduct, we choose the GPT-4o, one of the most advanced and popular large language models nowadays, for the task. First, due to cost considerations, we randomly select 5,000 observations from each of the 2018 and 2019 testing years, resulting in a total of 10,000 observations for the GPT-4o prediction.[15] Second, in the prompt, we provide GPT-4o with the definition of FA misconduct behavior, which is consistent with the definition used in our study, along with all relevant information about the FA used in the GBDT models. We then instruct the GPT-4o model to determine whether the FA could commit misconduct in the following year based on the information.[16] We provide an example of the prompt used for the baseline model in Appendix D. Third, we use the trained GBDT model to predict the same 10,000 randomly selected testing observations and compare its predictive performance with that of GPT-4o.

As shown in Table 9, all models using GPT-4o demonstrate worse predictive performance than those using the GBDT machine learning algorithm, as evidenced by an *AUC*

[15] To ensure that the 10,000 selected observations can be tested across different models, including models with 25, 34, 30, 40, and 54 input variables, we randomly select them from the smallest testing sample, which is for the model with 54 input variables.

[16] The temperature of GPT-4o model is set as zero in our analysis, thus minimize the randomness of responses.

range of 0.598-0.602 in Panel A vs. 0.661-0.699 in Panel B. In addition, we find that GPT-4o outperforms the GBDT model in terms of *Precision rate*, while the GBDT models exhibit better *Recall rate*. These results indicate that GPT-4o is more "lenient" and tends to classify more cases as non-misconduct, resulting in a lower probability of falsely classifying non-misconduct FAs as potential wrongdoers. However, this leniency also leads to a higher likelihood of failing to identify FAs who are prone to misconduct. In contrast, the GBDT model is more "strict," classifying a broader range of cases as potential misconduct. It ensures that the majority of FAs likely to engage in misconduct are identified. Nevertheless, it also results in a higher rate of false positives, where some FAs who would not commit misconduct are erroneously flagged as potential offenders.

[Insert Table 9]

From the client's perspective, each investment decision carries significant weight, and the potential consequences of entrusting funds to a FA who engages in misconduct can be severe, potentially resulting in substantial financial losses or even bankruptcy. Consequently, the opportunity cost of failing to identify an FA prone to misconduct is extraordinarily high, necessitating a more risk-averse approach that prioritizes caution over leniency. Upon careful analysis of the performance metrics, particularly the *AUC* and *Recall rate*, we suggest that large language models such as GPT-4o, while proficient in natural language processing tasks, demonstrate comparatively lower efficacy in prediction tasks that demand extensive computation and complex reasoning. Our findings indicate that for the specific task of predicting FA misconduct—where precision in computational analysis and logical reasoning is crucial—machine learning algorithms offer a more robust and reliable solution.

4.5 Advisor Alert: An Interactive Tool

Lastly, we have developed an interactive tool accessible through our website, *Advisor Alert* (*http://www.luoyaohuang.com:81/advisoralert*). This tool leverages our trained machine learning algorithms to predict FA misconduct. Figure 6 shows its application interface. Users can input relevant data about the FA, the advisory firm, and the county where the firm is located to receive a straightforward prediction of the likelihood that the FA will engage in misconduct in the following year. The tool aims to enhance transparency and trust in financial advisory services by providing valuable insights enabling clients to make informed decisions. Its interactive design ensures that users can easily navigate and utilize all features for optimal outcomes. However, we would also like to mention that *Advisor Alert* is provided "as is" without warranty of any kind, and its developers are not liable for any outcomes or decisions made based on its use.

[Insert Figure 6]

5. Conclusion

This paper conducts an in-depth analysis of FA misconduct prediction, utilizing the GBDT model, an advanced machine learning approach. We start with a baseline model with 25 FA and firm input variables that can be readily obtained from the IAPD, BrokerCheck, and Form ADV filings. Our findings indicate that the baseline model yields an *AUC* of 0.691 and a *NDCG@k* of 0.030. This suggests that while it demonstrates reasonably good predictive performance, it does not reach an extremely high level. According to the *SHAP value*, the FA's gender (*Female*), the number of FAs of the advisory firm (*No. of Advisors*), and the misconduct history of the FA (*Prior Misconduct*) emerge as the top three significant misconduct predictors.

We then incorporate additional input variables at various levels based on the multilevel theory, aiming to improve the predictive performance of our machine learning model. We find that incorporating supplementary firm input variables enhances both the effectiveness and efficiency of our machine learning model, as evidenced by a notably higher *AUC* and *NDCG@k* compared to the baseline model. While the inclusion of additional county/state or FA input variables might slightly enhance the *AUC*, it does not consistently result in a higher *NDCG@k*. We also find that incorporating all input variables could potentially augment the effectiveness and efficiency of our machine learning model, as indicated by the highest *AUC* and *NDCG@k* across all models. Regarding the importance of input variables, our findings emphasize that additional firm input variables, especially the advisory firm's *Misconduct ratio*, play a crucial role, whereas additional county/state and FA input variables play a trivial role in predicting FA misconduct.

Next, we focus on a longer predictive period and use input variables from the current year to predict FA misconduct over the next two or three years. The *AUC* shows a slight decrease with a more extended period. Regarding the model efficiency, we show that the model integrating additional firm information achieves the highest *NDCG@k* for predicting FA misconduct within the next two (three) years across all models. We also show that as the predictive period extends, the cumulative importance of additional firm input variables rises. These findings underscore the increasingly significant role of additional firm input variables in predicting FA misconduct over an extended period. They not only enhance the model efficiency but also serve as important misconduct predictors in this context.

Next, we explore which factors are critical in predicting recidivism among FAs with a history of misconduct and the first misconduct of FAs with a clean record. The results indicate that although our machine learning model's effectiveness in predicting recidivism is low, its efficiency is high. When predicting the first misconduct of FAs with a clean record, both the effectiveness and efficiency are lower than those of our baseline model, which predicts all forms of misconduct.

Lastly, we conduct a few additional tests. First, we employ alternative definitions of misconduct by concentrating on (1) judicial and regulatory-initiated misconduct cases, which typically entail more significant social costs and consequences, and (2) misconduct initialed by clients. The results are consistent with our main findings, wherein the predictive performance is improved with the inclusion of additional input variables. Second, we explore alternative values for *k* at 10,000, 5,000, 1,000, and 500. Our results provide some new insights into the role of additional firm input variables. Third, our results demonstrate robustness when employing alternative machine learning models. Fourth, we show that our GBDT model outperforms one of the most advanced and popular large language models, GPT-4o, in predicting FA misconduct. Lastly, we developed *Advisor Alert*, an interactive tool that uses machine learning algorithms to predict FA misconduct. Users can input data about the FA, the advisory firm, and the county for a straightforward prediction of future misconduct.

Overall, we show that by solely relying on open-access data, the machine learning model performs reasonably well in forecasting FA misconduct of the following year. Nonetheless, to enhance model efficiency, incorporating supplementary firm data, especially the overall misconduct status of the firm, is essential. Our study offers valuable insights and methodologies for advanced analyses in the field of FA misconduct prediction.

Given the prevalence of seeking financial advice in the U.S. alongside the concerning levels of professional misconduct within the advisory industry, clients should exercise great caution when choosing their FAs. To achieve this, there is a critical need to embrace advanced technologies, such as state-of-the-art machine learning algorithms, to predict FA misconduct, thus protecting their own interests. If the existing public data proves insufficient for this purpose, clients are encouraged to invest efforts in accessing additional relevant information. This could involve acquiring comprehensive firm information, enabling clients to make more informed decisions when selecting a FA. By adopting a proactive approach to understanding and utilizing available data, clients can contribute to fostering a more transparent and trustworthy FA landscape, ultimately safeguarding their financial interests.

References

Bao, Y., Ke, B., Li, B., Yu, Y. J., and Zhang, J. (2020). Detecting accounting fraud in publicly traded U.S. firms using a machine learning approach. *Journal of Accounting Research* 58(1), 199-235.

Bertomeu, J. (2020). Machine learning improves accounting: Discussion, implementation and research opportunities. *Review of Accounting Studies* 25(3), 1135-1155.

Bertomeu, J., Cheynel, E., Floyd, E., and Pan, W. (2021). Using machine learning to detect misstatements. *Review of Accounting Studies* 26(2), 468-519.

Bertomeu, J., Cheynel, E., Liao, Y., and Milone, M. (2024). Using machine learning to measure conservatism. *Management Science*.

Brown, N. C., Crowley, R. M., and Elliott, W. B. (2020). What are you saying? Using topic to detect financial misreporting. *Journal of Accounting Research* 58(1), 237–291.

Cecchini, M., Aytug, H., Koehler, G. J., and Pathak, P. (2010). Detecting management fraud in public companies. *Management Science* 56(7), 1146-1160.

Chen, X., Cho, Y. H., Dou, Y., and Lev, B. (2022). Predicting future earnings changes using machine learning and detailed financial data. *Journal of Accounting Research* 60(2), 467-515.

Chen, T., and C. Guestrin. (2016). XGBoost: A scalable tree boosting system. *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 785-794.

Clifford, C. P., Ellis, J. A., and Gerken, W. C. (2023). Childhood exposure to misbehavior and the culture of financial misconduct. *Working Paper*.

Dimmock, S. G., Gerken, W. C., and Graham, N. P. (2018). Is fraud contagious? Coworker influence on misconduct by financial advisors. *The Journal of Finance* 73(3), 1417-1450.

Dimmock, S. G., Gerken, W. C., and Van Alfen, T. (2021). Real estate shocks and financial advisor misconduct. *The Journal of Finance* 76(6), 3309-3346.

Ding, K., Lev, B., and Peng, X. (2020). Machine learning improves accounting estimates: evidence from insurance payments. *Review of Accounting Studies* 25, 1098-1134.

Dornadula, V. N., and Geetha, S. (2019). Credit card fraud detection using machine learning algorithms. *Procedia Computer Science* 165, 631-641.

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

Egan, M., Matvos, G., and Seru, A. (2022). When Harry fired Sally: The double standard in punishing misconduct. *Journal of Political Economy* 130(5), 1184-1248.

Erel, I., Stern, L. H., Tan, C., and Weisbach, M. S. (2021). Selecting directors using machine learning. *Review of Financial Studies* 34(7), 3226-3264.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics* 29(5),1189-1232.

Geertsema, P. and Lu, H. (2023). Relative valuation with machine learning. *Journal of Accounting Research* 61(1), 329-376.

Gerken, W. C., and Shahraki, M. M. (2023). Third party quality certification in the market for financial advice. *Working Paper*.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies* 33(5), 2223-2273.

Guenther, D. A., Peterson, K., Searcy, J., and Williams, B. M. (2023). How useful are tax disclosures in predicting effective tax rates? A machine learning approach. *The Accounting Review* 98 (5), 297-322.

Guilford, G., and Kopf, D. (2020). Wall Street's watchdog is obscuring data that could protect investors. *Quartz*. *https://qz.com/1750840/finra-obscures-brokerage-data-that-could-protect-investors*

Huang, A. H., Kraft, P., and Wang, S. (2023). The usefulness of credit ratings for accounting fraud prediction. *The Accounting Review* 98(7), 347-376.

Ileberi, E., Sun, Y., and Wang, Z. (2022). A machine learning based credit card fraud detection using the GA algorithm for feature selection. *Journal of Big Data* 9, 24.

Järvelin, K., and Kekäläinen, J. (2002). Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS)* 20(4), 422-446.

Klein, K. J., and Kozlowski, S. W. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*, 3-90.

Kowaleski, Z. T., Sutherland, A., and Vetter, F. W. (2020). Can ethics be taught? Evidence from securities exams and investment adviser misconduct. *Journal of Financial Economics* 138(1), 159-175.

Lausen, J., Clapham, B., Siering, M., and Gomber, P. (2020). Who is the next "Wolf of Wall Street"? Detection of financial intermediary misconduct. *Journal of the Association for Information Systems* 21(5), 1153-1190.

Lausen, J., and Clapham, B. (2023). Give them a second chance? Prediction of recurrent financial intermediary misconduct. In: van Hillegersberg, J., Osterrieder, J., Rabhi, F., Abhishta, A., Marisetty, V., Huang, X. (eds) *Enterprise Applications, Markets and Services in the Finance Industry*. FinanceCom 2022. Lecture Notes in Business Information Processing, 467. Springer, Cham.

Law, K. K. F., and Mills, L. F. (2019). Financial gatekeepers and investor protection: Evidence from criminal background checks. *Journal of Accounting Research* 57(2), 491-543.

Law, K. K. F., and Zuo, L. (2021). How does the economy shape the financial advisory profession? *Management Science* 67(4), 2466-2482.

Liu, M. (2022). Assessing human information processing in lending decisions: A machine learning approach. *Journal of Accounting Research* 60(2), 607-651.

Perols, J. (2011). Financial statement fraud detection: An analysis of statistical and machine learning algorithms. *Auditing: A Journal of Practice and Theory* 30(2), 19-50.

Perols, J. L., Bowen, R. M., Zimmermann, C, and Samba, B. (2017). Finding needles in a haystack: Using data analytics to improve fraud prediction. *The Accounting Review* 92(2), 221-245.

Purda, L., and Skillicorn, D. (2015). Accounting variables, deception, and a bag of words: Assessing the tools of fraud detection. *Contemporary Accounting Research* 32(3), 1193-1223.

Xu, X, Xiong, F, and An, Z. (2023). Using Machine learning to predict corporate fraud: Evidence based on the gone framework, *Journal of Business Ethics* 186, 137-158.

Figure 1. The *SHAP Value* and Summary Plot (25 Baseline Inputs)

Figure 1.1 illustrates the *SHAP values* of input variables in a horizontal bar graph based on the GBDT model with 25 baseline input variables. We report the average *SHAP values* for the testing years 2018 and 2019. Figures 1.2 and 1.3 present the summary plot for the testing years 2018 and 2019, respectively.

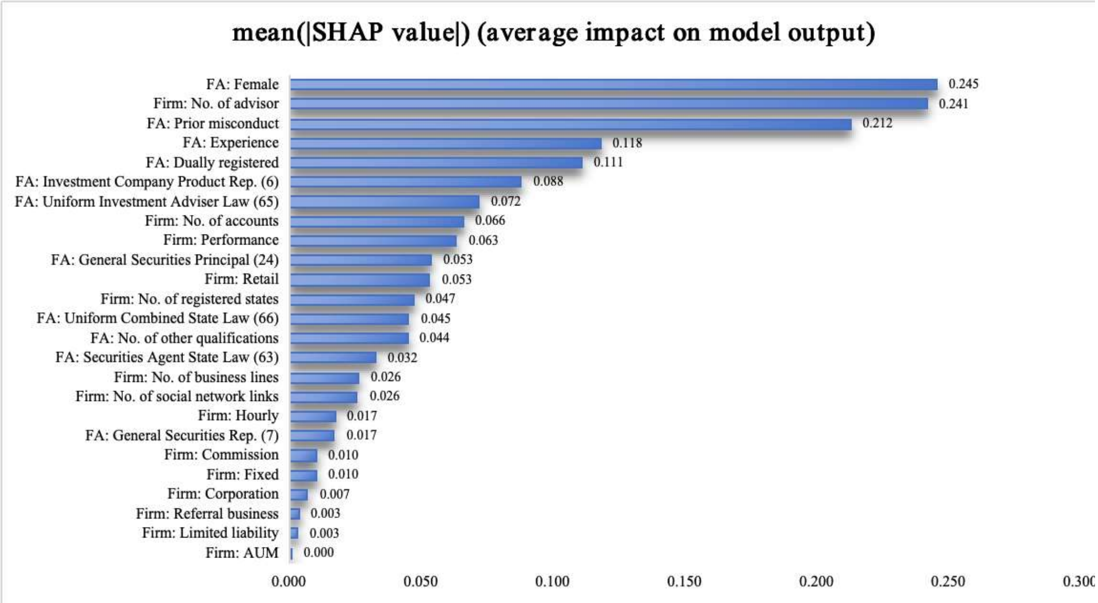


Figure 1.1 - The Importance of Input Variables

Figure 1.2 - Summary Plot (Testing Year = 2018)

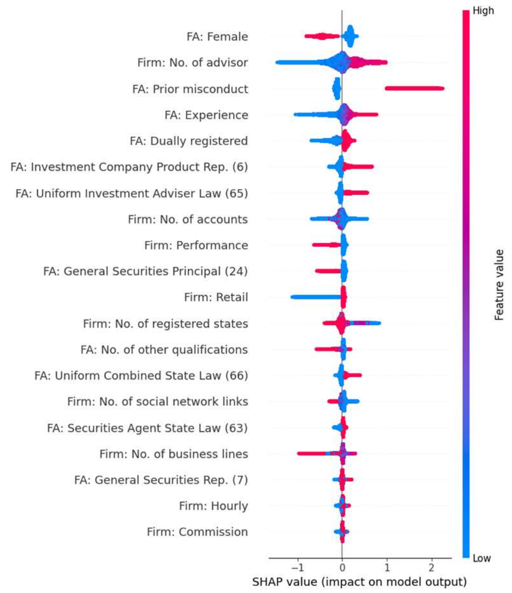


Figure 1.3 - Summary Plot (Testing Year = 2019)

Figure 2. The *SHAP Value* and Summary Plot (25 Baseline Inputs + 9 County/State Inputs)

Figure 2.1 illustrates the *SHAP values* of input variables in a horizontal bar graph based on the GBDT model with 25 baseline input variables and nine additional county/state input variables. We report the average *SHAP values* for the testing years 2018 and 2019. Figures 2.2 and 2.3 present the summary plot for the testing years 2018 and 2019, respectively. For brevity, we report the input variables with the 20 highest *SHAP values*.

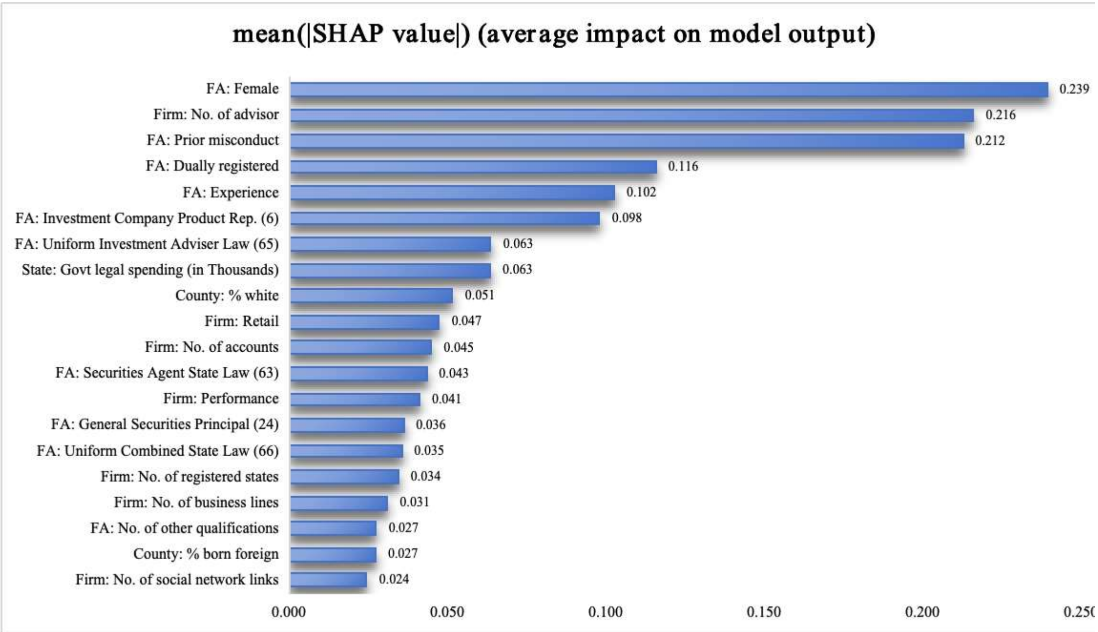


Figure 2.1 - Importance of Input Variables

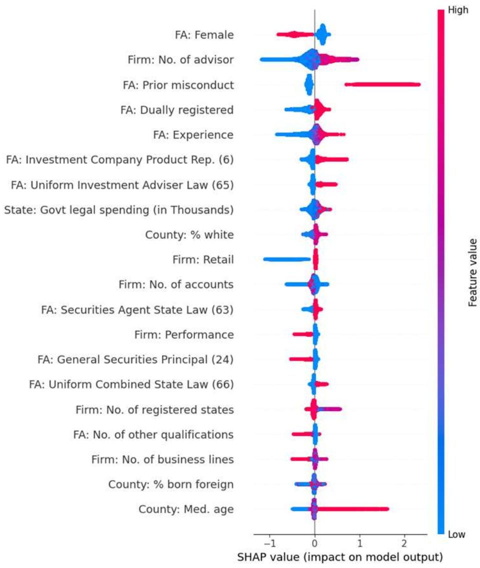


Figure 2.2 - Summary Plot (Testing Year = 2018)

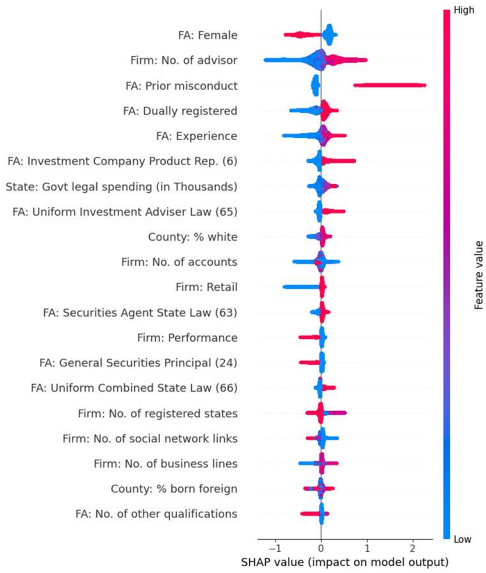


Figure 2.3 - Summary Plot (Testing Year = 2019)

Figure 3. The *SHAP Value* and Summary Plot (25 Baseline Inputs + 5 Firm Inputs)

Figure 3.1 illustrates the *SHAP values* of input variables in a horizontal bar graph based on the GBDT model with 25 baseline input variables and five additional firm input variables. We report the average *SHAP values* for the testing years 2018 and 2019. Figures 3.2 and 3.3 present the summary plot for the testing years 2018 and 2019, respectively. For brevity, we report the input variables with the 20 highest *SHAP values*.

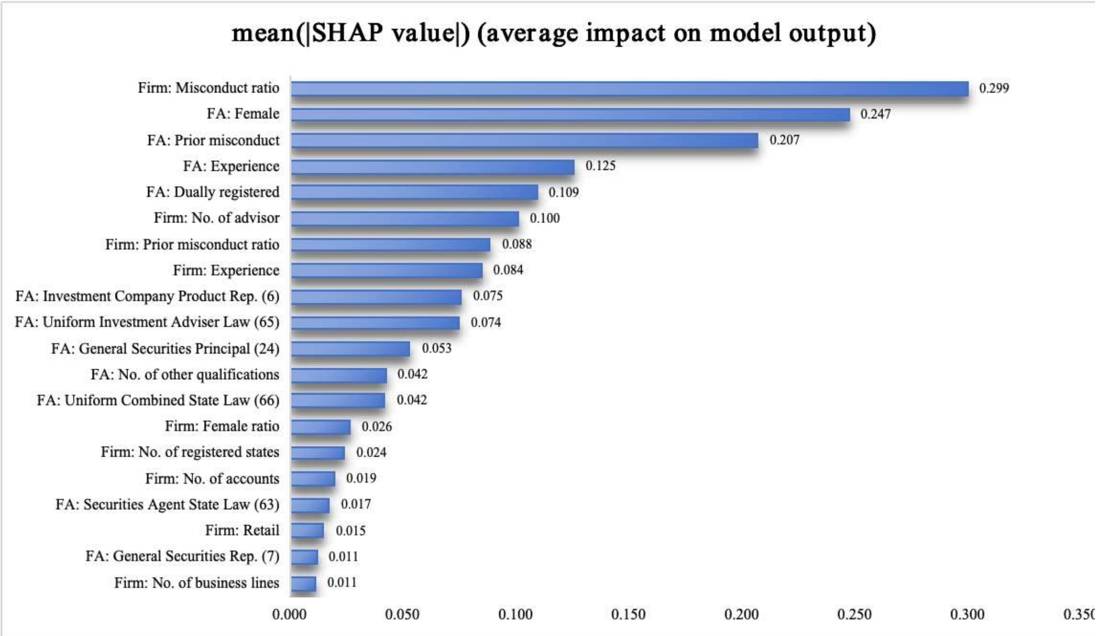


Figure 3.1 - Importance of Input Variables

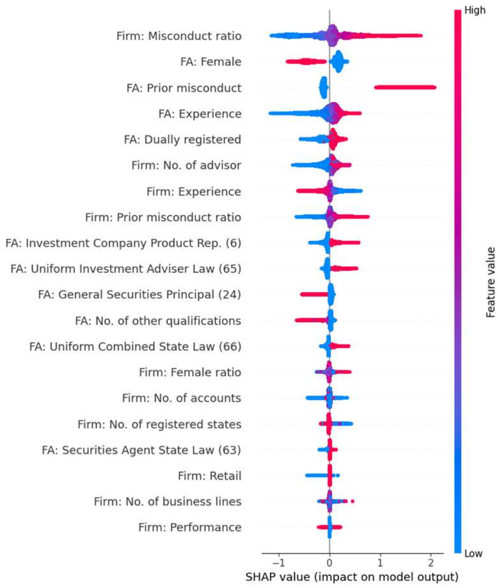


Figure 3.2 - Summary Plot (Testing Year = 2018)

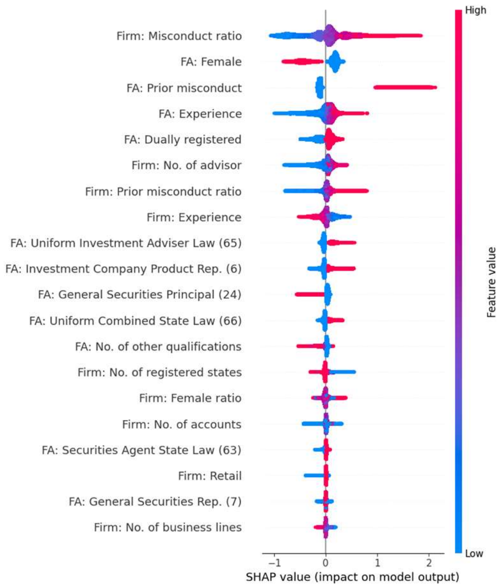


Figure 3.3 - Summary Plot (Testing Year = 2019)

Figure 4. The *SHAP Value* and Summary Plot (25 Baseline Inputs + 15 FA Inputs)

Figure 4.1 illustrates the *SHAP values* of input variables in a horizontal bar graph based on the GBDT model with 25 baseline input variables and 15 additional FA input variables. We report the average *SHAP values* for the testing years 2018 and 2019. Figures 4.2 and 4.3 present the summary plot for the testing years 2018 and 2019, respectively. For brevity, we report the input variables with the 20 highest *SHAP values*.

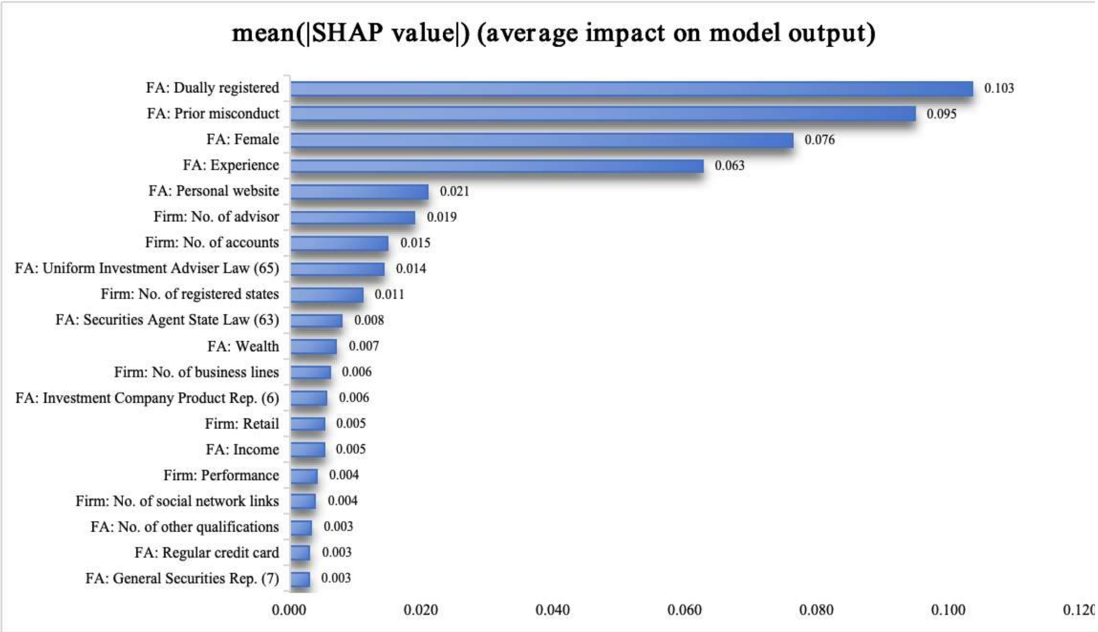


Figure 4.1 - Importance of Input Variables

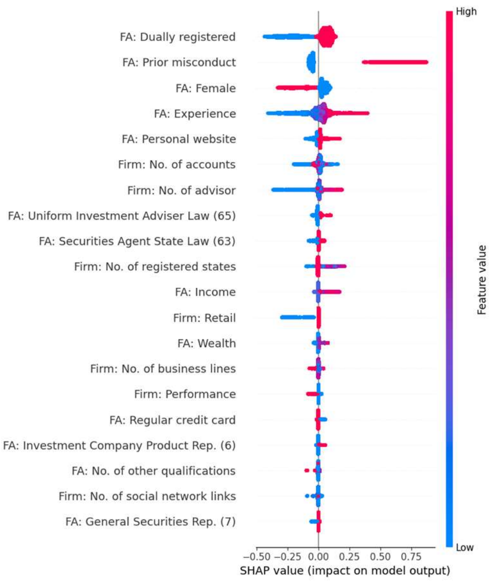


Figure 4.2 - Summary Plot (Testing Year = 2018)

Figure 4.3 - Summary Plot (Testing Year = 2019)

Figure 5. The *SHAP Value* and Summary Plot (25 Baseline Inputs + All Additional Inputs)

Figure 5.1 illustrates the *SHAP values* of input variables in a horizontal bar graph based on the GBDT model with 25 baseline input variables and all additional input variables. We report the average *SHAP values* for the testing years 2018 and 2019. Figures 5.2 and 5.3 present the summary plot for the testing years 2018 and 2019, respectively. For brevity, we report the input variables with the 20 highest *SHAP values*.

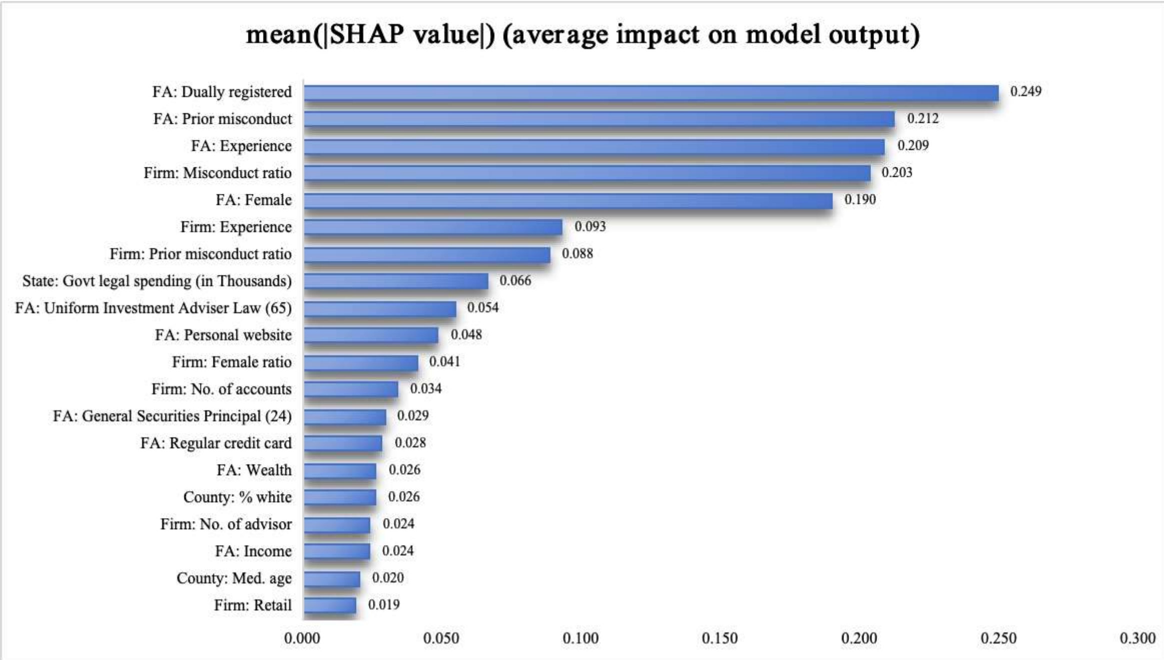


Figure 5.1 - Importance of Input Variables

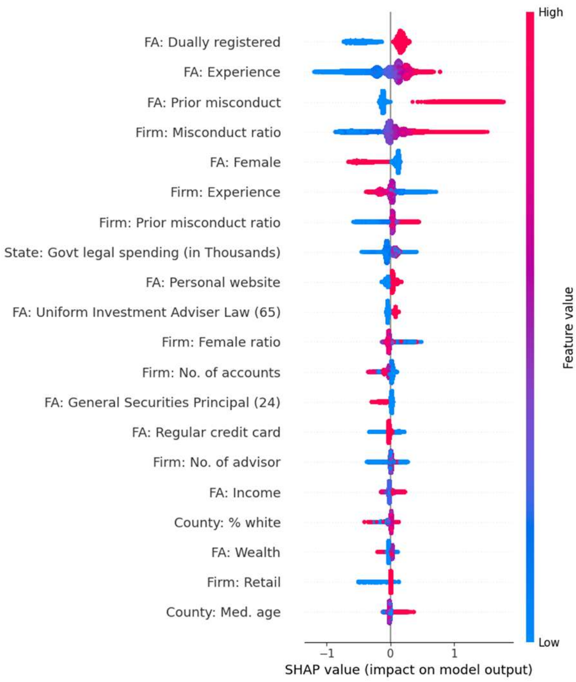


Figure 5.2 - Summary Plot (Testing Year = 2018)

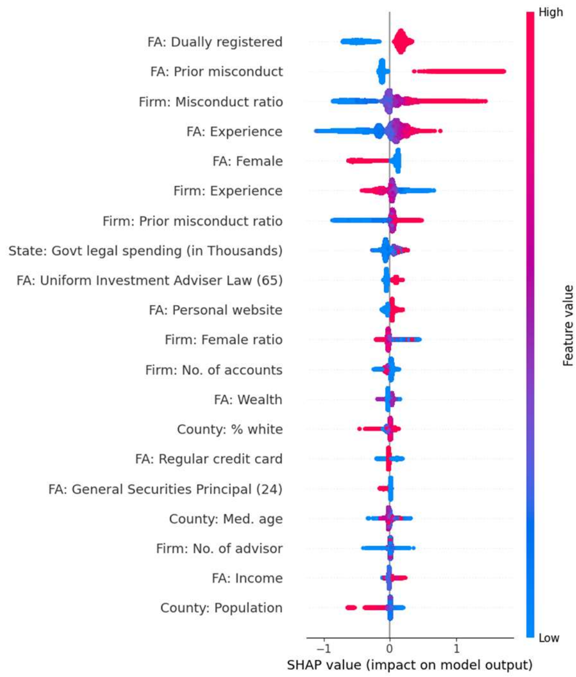


Figure 5.3 - Summary Plot (Testing Year = 2019)

Figure 6. Advisor Alert: An Interactive Tool

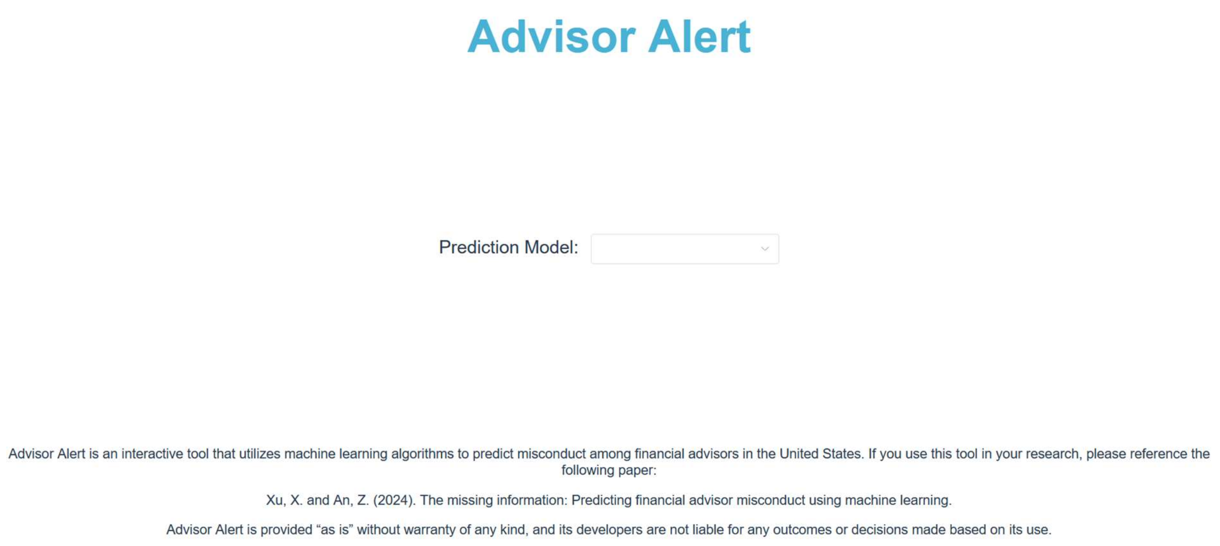
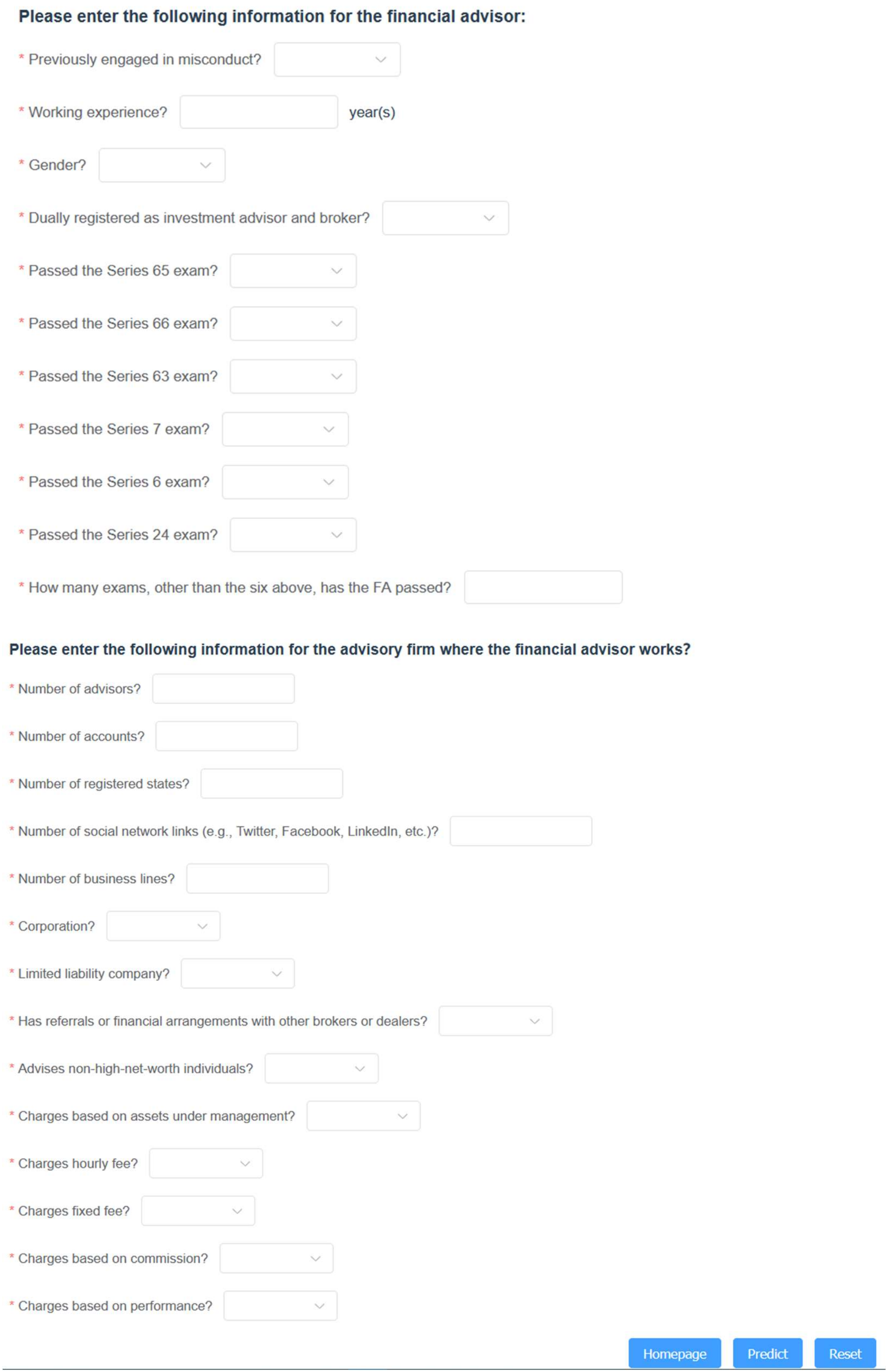


Figure 6. Advisor Alert: An Interactive Tool (Cont.)



Please enter the following information for the financial advisor:

* Previously engaged in misconduct?

* Working experience? year(s)

* Gender?

* Dually registered as investment advisor and broker?

* Passed the Series 65 exam?

* Passed the Series 66 exam?

* Passed the Series 63 exam?

* Passed the Series 7 exam?

* Passed the Series 6 exam?

* Passed the Series 24 exam?

* How many exams, other than the six above, has the FA passed?

Please enter the following information for the advisory firm where the financial advisor works?

* Number of advisors?

* Number of accounts?

* Number of registered states?

* Number of social network links (e.g., Twitter, Facebook, LinkedIn, etc.)?

* Number of business lines?

* Corporation?

* Limited liability company?

* Has referrals or financial arrangements with other brokers or dealers?

* Advises non-high-net-worth individuals?

* Charges based on assets under management?

* Charges hourly fee?

* Charges fixed fee?

* Charges based on commission?

* Charges based on performance?

Homepage Predict Reset

Table 1. Descriptive Statistics

This table presents the descriptive statistics of the outcome variable and each input variable. Columns 1-6 report the number of observations, the mean, standard deviation, and 25th, 50th, and 75th percentile values of each variable.

	N	*Mean*	*Std. Dev.*	*p25*	*p50*	*p75*
	[1]	[2]	[3]	[4]	[5]	[6]
Outcome Variable						
FA: Misconduct	3,639,236	0.006	0.078	0	0	0
Baseline Input Variables						
FA: Prior misconduct	3,639,236	0.072	0.259	0	0	0
FA: Experience	3,639,236	13.402	10.583	4	12	20
FA: Female	3,639,236	0.278	0.448	0	0	1
FA: Dually registered	3,639,236	0.626	0.484	0	1	1
FA: Uniform Investment Adviser Law (65)	3,639,236	0.270	0.444	0	0	1
FA: Uniform Combined State Law (66)	3,639,236	0.310	0.462	0	0	1
FA: Securities Agent State Law (63)	3,639,236	0.662	0.473	0	1	1
FA: General Securities Rep. (7)	3,639,236	0.720	0.449	0	1	1
FA: Investment Company Product Rep. (6)	3,639,236	0.316	0.465	0	0	1
FA: General Securities Principal (24)	3,639,236	0.125	0.331	0	0	0
FA: No. of other qualifications	3,639,236	0.448	0.853	0	0	1
Firm: No. of advisors	38,074	126	1,226	3	6	12
Firm: No. of accounts	38,074	5,949	110,900	221	553	1,249
Firm: No. of registered states	38,074	12.141	15.235	2	5	14
Firm: No. of social network links	38,074	2.270	12.907	1	1	2
Firm: No. of business lines	38,074	3.960	1.756	3	4	5
Firm: Corporation	38,074	0.456	0.498	0	0	1
Firm: Limited liability	38,074	0.527	0.499	0	1	1

Firm: Referral business	38,074	0.832	0.374	1	1	1
Firm: Retail	38,074	0.891	0.312	1	1	1
Firm: AUM	38,074	0.989	0.105	1	1	1
Firm: Hourly	38,074	0.477	0.499	0	0	1
Firm: Fixed	38,074	0.624	0.484	0	1	1
Firm: Commission	38,074	0.072	0.259	0	0	0
Firm: Performance	38,074	0.144	0.351	0	0	0

Table 2. The Predictive Performance of the Machine Learning Model

This table presents the predictive performance of the GBDT model. Columns 1-7 report the *AUC*, *precision rate*, *recall rate*, *f1 score*, *NDCG@k*, *Precision@k*, and *Recall@k*, respectively. The outcome variable is a dummy indicating whether a FA commits misconduct in the next year. We employ 25 baseline input variables and additional county/state, firm, and FA input variables outlined in Appendix B. Performance metrics have been averaged across the 2018-2019 test period. The value of *k* is configured as 2,000.

	AUC	*Precision rate*	*Recall rate*	*f1 score*	*NDCG@k*	*Precision@k*	*Recall@k*
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
25 baseline inputs	0.691	0.010	0.585	0.020	0.030	2.60%	3.06%
25 baseline inputs + 9 county/state inputs	0.695	0.010	0.560	0.020	0.027	2.18%	2.86%
25 baseline inputs + 5 firm inputs	0.718	0.000	0.560	0.020	0.060	4.45%	5.24%
25 baseline inputs + 15 FA inputs	0.695	0.010	0.560	0.010	0.048	3.20%	6.12%
25 baseline inputs + all additional inputs	0.729	0.010	0.570	0.010	0.065	3.90%	8.19%

Table 3. The Importance of Key Input Variables

This table presents the proportion of the *SHAP value* of key input variables out of the *SHAP value* of all input variables in each model. Columns 1, 2, and 3 report the results of predicting misconduct in the next year, the next two years, and the next three years, respectively.

Predicting Misconduct:		*Next Year*	*Next Two Year*	*Next Three Year*
Model	*Key Inputs*	[1]	[2]	[3]
25 baseline inputs + 9 county/state inputs	*9 county/state inputs*	13.24%	11.92%	12.82%
25 baseline inputs + 5 firm inputs	*5 firm inputs*	29.33%	29.93%	30.50%
25 baseline inputs + 15 FA inputs	*15 FA inputs*	8.23%	6.38%	5.52%
25 baseline inputs + all additional inputs	*9 county/state inputs*	10.43%	9.56%	9.14%
25 baseline inputs + all additional inputs	*5 firm inputs*	23.13%	24.78%	25.47%
25 baseline inputs + all additional inputs	*15 FA inputs*	8.64%	8.61%	8.51%
25 baseline inputs + 5 firm inputs	*Firm: Misconduct ratio*	17.51%	16.72%	15.69%
25 baseline inputs + all additional inputs	*Firm: Misconduct ratio*	10.87%	10.93%	10.69%

Table 4. Longer Predictive Period

This table presents the predictive performance of the GBDT model. Columns 1-7 report the *AUC*, *precision rate*, *recall rate*, *f1 score*, *NDCG@k*, *Precision@k*, and *Recall@k*, respectively. The outcome variables are dummies indicating whether a FA commits misconduct in the next two years and the next three years. We employ 25 baseline input variables and additional county/state, firm, and FA input variables outlined in Appendix B. Performance metrics have been averaged across the 2018-2019 test period. The value of *k* is configured as 2,000.

	AUC	*Precision rate*	*Recall rate*	*f1 score*	*NDCG@k*	*Precision@k*	*Recall@k*
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Panel A: Predicting Misconduct in the Next Two Years							
25 baseline inputs	0.690	0.015	0.595	0.025	0.035	3.50%	2.55%
25 baseline inputs + 9 county/state inputs	0.691	0.010	0.560	0.020	0.031	3.13%	2.43%
25 baseline inputs + 5 firm inputs	0.708	0.000	0.560	0.020	0.074	6.48%	4.56%
25 baseline inputs + 15 FA inputs	0.693	0.010	0.560	0.020	0.049	5.25%	5.86%
25 baseline inputs + all additional inputs	0.722	0.010	0.595	0.025	0.062	5.35%	6.38%
Panel B: Predicting Misconduct in the Next Three Years							
25 baseline inputs	0.685	0.015	0.585	0.030	0.030	3.05%	1.91%
25 baseline inputs + 9 county/state inputs	0.690	0.010	0.560	0.020	0.034	3.50%	2.41%
25 baseline inputs + 5 firm inputs	0.707	0.010	0.560	0.020	0.082	6.85%	4.25%
25 baseline inputs + 15 FA inputs	0.686	0.010	0.560	0.020	0.045	5.45%	5.19%
25 baseline inputs + all additional inputs	0.720	0.015	0.625	0.025	0.060	5.70%	6.17%

Table 5. Predicting Recidivism and First Misconduct

This table presents the predictive performance of the GBDT model. Columns 1-7 report the *AUC*, *precision rate*, *recall rate*, *f1 score*, *NDCG@k*, *Precision@k*, and *Recall@k*, respectively. The outcome variable is a dummy indicating whether a FA commits misconduct in the next year. We employ 25 baseline input variables and additional county/state, firm, and FA input variables outlined in Appendix B. Performance metrics have been averaged across the 2018-2019 test period. The value of *k* is configured as 2,000.

	AUC	*Precision rate*	*Recall rate*	*f1 score*	*NDCG@k*	*Precision@k*	*Recall@k*
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Panel A: Predicting Recidivism							
25 baseline inputs	0.546	0.030	0.300	0.050	0.095	2.85%	12.48%
25 baseline inputs + 9 county/state inputs	0.571	0.020	0.560	0.040	0.112	2.88%	14.49%
25 baseline inputs + 5 firm inputs	0.612	0.030	0.560	0.050	0.176	4.85%	20.83%
25 baseline inputs + 15 FA inputs	0.535	0.020	0.560	0.030	0.134	3.30%	21.67%
25 baseline inputs + all additional inputs	0.591	0.020	0.410	0.040	0.180	4.35%	29.81%
Panel B: Predicting First Misconduct							
25 baseline inputs	0.631	0.010	0.550	0.010	0.018	1.20%	1.96%
25 baseline inputs + 9 county/state inputs	0.644	0.010	0.560	0.010	0.020	1.25%	2.23%
25 baseline inputs + 5 firm inputs	0.667	0.010	0.560	0.010	0.024	1.45%	2.36%
25 baseline inputs + 15 FA inputs	0.634	0.000	0.560	0.010	0.017	0.85%	2.24%
25 baseline inputs + all additional inputs	0.680	0.000	0.630	0.010	0.039	1.75%	5.25%

Table 6. Alternative Misconduct Definition

This table presents the predictive performance of the GBDT model. Columns 1-7 report the *AUC*, *precision rate*, *recall rate*, *f1 score*, *NDCG@k*, *Precision@k*, and *Recall@k*, respectively. The outcome variable in Panel A is a dummy indicating whether a FA commits misconduct in the next year based on the misconduct initiated by judicial and regulatory authorities (i.e., *Civil-Final*, *Criminal-Final Disposition*, and *Regulatory-Final*). The outcome variable in Panel B is a dummy indicating whether a FA commits misconduct in the next year based on the misconduct initiated by clients. We employ 25 baseline input variables and additional county/state, firm, and FA input variables outlined in Appendix B. Performance metrics have been averaged across the 2018-2019 test period. The value of *k* is configured as 2,000.

	AUC	*Precision rate*	*Recall rate*	*f1 score*	*NDCG@k*	*Precision@k*	*Recall@k*
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Panel A: Predicting Severe Misconduct							
25 baseline inputs	0.738	0.000	0.510	0.005	0.040	0.80%	6.18%
25 baseline inputs + 9 county/state inputs	0.743	0.000	0.560	0.000	0.043	0.85%	6.88%
25 baseline inputs + 5 firm inputs	0.743	0.000	0.560	0.000	0.135	2.25%	17.40%
25 baseline inputs + 15 FA inputs	0.745	0.000	0.560	0.000	0.048	0.75%	10.56%
25 baseline inputs + all additional inputs	0.795	0.000	0.605	0.005	0.176	1.85%	29.74%
Panel B: Predicting Compliant							
25 baseline inputs	0.784	0.010	0.730	0.015	0.043	2.10%	5.33%
25 baseline inputs + 9 county/state inputs	0.774	0.000	0.645	0.010	0.026	1.10%	3.40%
25 baseline inputs + 5 firm inputs	0.837	0.010	0.795	0.010	0.158	5.43%	13.67%
25 baseline inputs + 15 FA inputs	0.726	0.000	0.675	0.010	0.039	1.80%	6.02%
25 baseline inputs + all additional inputs	0.786	0.005	0.685	0.010	0.115	3.90%	14.52%

Table 7. Alternative *k*

This table presents the *NDCG@k* of the GBDT model with alternative *k* values. The outcome variable is a dummy indicating whether a FA commits misconduct in the next year. We employ 25 baseline input variables and additional county/state, firm, and FA input variables outlined in Appendix B. Performance metrics have been averaged across the 2018-2019 test period. The value of *k* is configured as 10,000, 5,000, 2,000, 1,000, and 500, respectively.

	NDCG@10,000	*NDCG@5,000*	*NDCG@2,000*	*NDCG@1,000*	*NDCG@500*
	[1]	[2]	[3]	[4]	[5]
25 baseline inputs	0.093	0.055	0.030	0.031	0.033
25 baseline inputs + 9 county/state inputs	0.091	0.058	0.027	0.019	0.019
25 baseline inputs + 5 firm inputs	0.131	0.097	0.060	0.073	0.090
25 baseline inputs + 15 FA inputs	0.156	0.102	0.048	0.031	0.016
25 baseline inputs + all additional inputs	0.175	0.106	0.065	0.043	0.026

Table 8. Alternative Machine Learning Model

This table presents the predictive performance of alternative machine learning models, including the Random Forest, RUSBoost, Logistic Regression, and ANN. Columns 1-7 report the *AUC*, *precision rate*, *recall rate*, *f1 score*, *NDCG@k*, *Precision@k*, and *Recall@k*, respectively. The outcome variable is a dummy indicating whether a FA commits misconduct in the next year. We employ 25 baseline input variables and additional county/state, firm, and FA input variables outlined in Appendix B. Performance metrics have been averaged across the 2018-2019 test period. The value of *k* is configured as 2,000.

	AUC	*Precision rate*	*Recall rate*	*f1 score*	*NDCG@k*	*Precision@k*	*Recall@k*
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Panel A: Random Forest							
25 baseline inputs	0.703	0.010	0.565	0.020	0.028	2.33%	2.75%
25 baseline inputs + 9 county/state inputs	0.704	0.010	0.560	0.020	0.026	2.08%	2.73%
25 baseline inputs + 5 firm inputs	0.721	0.010	0.560	0.020	0.035	3.05%	3.61%
25 baseline inputs + 15 FA inputs	0.700	0.010	0.560	0.010	0.026	1.70%	3.24%
25 baseline inputs + all additional inputs	0.722	0.010	0.540	0.010	0.039	2.35%	4.90%
Panel B: RUSBoost							
25 baseline inputs	0.671	0.010	0.515	0.020	0.014	1.28%	1.51%
25 baseline inputs + 9 county/state inputs	0.672	0.010	0.560	0.020	0.016	1.23%	1.61%
25 baseline inputs + 5 firm inputs	0.710	0.010	0.560	0.020	0.063	4.80%	5.65%
25 baseline inputs + 15 FA inputs	0.691	0.010	0.560	0.010	0.030	1.90%	3.65%
25 baseline inputs + all additional inputs	0.728	0.010	0.565	0.010	0.074	4.25%	8.89%
Panel C: Logistic Regression							
25 baseline inputs	0.683	0.010	0.550	0.020	0.013	1.13%	1.32%
25 baseline inputs + 9 county/state inputs	0.675	0.010	0.525	0.015	0.014	1.15%	1.51%
25 baseline inputs + 5 firm inputs	0.706	0.010	0.545	0.020	0.061	4.15%	4.87%
25 baseline inputs + 15 FA inputs	0.710	0.010	0.615	0.010	0.035	1.18%	4.48%
25 baseline inputs + all additional inputs	0.728	0.010	0.600	0.010	0.079	1.68%	7.02%

Panel D: ANN							
25 baseline inputs	0.645	0.010	0.330	0.020	0.023	1.95%	2.30%
25 baseline inputs + 9 county/state inputs	0.569	0.010	0.405	0.010	0.009	0.70%	0.92%
25 baseline inputs + 5 firm inputs	0.644	0.010	0.320	0.015	0.037	2.50%	2.90%
25 baseline inputs + 15 FA inputs	0.574	0.010	0.070	0.010	0.017	0.53%	2.00%
25 baseline inputs + all additional inputs	0.615	0.010	0.100	0.010	0.041	0.85%	3.56%

Table 9. GPT-4o

This table presents the predictive performance of the GPT-4o model. Columns 1-7 report the *AUC*, *precision rate*, *recall rate*, and *f1 score*, respectively. The outcome variable is a dummy indicating whether a FA commits misconduct in the next year. We employ 25 baseline input variables and additional county/state, firm, and FA input variables outlined in Appendix B, as well as interactions of each pair of input variables. Performance metrics have been averaged across the 2018-2019 test period.

	AUC	*Precision rate*	*Recall rate*	*f1 score*
	[1]	[2]	[3]	[4]
Panel A: GPT-4o				
25 baseline inputs	0.602	0.015	0.295	0.020
25 baseline inputs + 9 county/state inputs	0.602	0.015	0.295	0.020
25 baseline inputs + 5 firm inputs	0.610	0.010	0.320	0.025
25 baseline inputs + 15 FA inputs	0.602	0.015	0.295	0.020
25 baseline inputs + all additional inputs	0.598	0.010	0.295	0.020
Panel B: GBDT				
25 baseline inputs	0.661	0.010	0.585	0.010
25 baseline inputs + 9 county/state inputs	0.669	0.010	0.610	0.010
25 baseline inputs + 5 firm inputs	0.691	0.010	0.610	0.010
25 baseline inputs + 15 FA inputs	0.643	0.005	0.625	0.010
25 baseline inputs + all additional inputs	0.699	0.010	0.600	0.015

Appendix A. Disclosure Definitions

This table presents the definitions of 23 disclosure categories in FINRA (*http://brokercheck.finra.org*). * indicates the misconduct disclosure categories defined by Egan, Matvos, and Seru (2019; 2022).

Disclosure Category	**Definition**
**Civil – Final*	This type of disclosure event involves (1) an injunction issued by a court in connection with the investment-related activity, (2) a finding by a court of a violation of any investment-related statute or regulation, or (3) an action brought by a state or foreign financial regulatory authority that is dismissed by a court pursuant to a settlement agreement.
Civil - On Appeal	This type of disclosure event involves an injunction issued by a court in connection with the investment-related activity or a finding by a court of a violation of any investment-related statute or regulation currently on appeal.
Civil - Pending	This type of disclosure event involves a pending civil court action that seeks an injunction in connection with any investment-related activity or alleges a violation of any investment-related statute or regulation.
Civil Bond	This type of disclosure event involves a civil bond for the adviser that has been denied, paid, or revoked by a bonding company.
**Criminal - Final Disposition*	This type of disclosure event involves a criminal charge against the adviser that has resulted in a conviction, acquittal, dismissal, or plea. The criminal matter may pertain to any felony or certain misdemeanor offenses, including bribery, perjury, forgery, counterfeiting, extortion, fraud, and wrongful taking of property.
Criminal - On Appeal	This type of disclosure event involves a conviction for any felony or certain misdemeanor offenses, including bribery, perjury, forgery, counterfeiting, extortion, fraud, and wrongful taking of property that is currently on appeal.
Criminal - Pending Charge	This type of disclosure event involves a formal charge for a crime involving a felony or certain misdemeanor offenses, including bribery, perjury, forgery, counterfeiting, extortion, fraud, and wrongful taking of property is currently pending.

**Customer Dispute - Award/Judgment*	This type of disclosure event involves a final, consumer-initiated, investment-related arbitration or civil suit containing allegations of sales practice violations against the adviser that resulted in an arbitration award or civil judgment for the customer.
**Customer Dispute - Settled*	This type of disclosure event involves a consumer-initiated, investment-related complaint, arbitration proceeding, or civil suit containing allegations of sales practice violations against the adviser that resulted in a monetary settlement to the customer.
Customer Dispute - Closed - No Action/Withdrawn/Dismissed/Denied/Final	This type of disclosure event involves (1) a consumer-initiated, investment-related arbitration or civil suit containing allegations of sales practice violations against the individual adviser that was dismissed, withdrawn, or denied or (2) a consumer-initiated, investment-related written complaint containing allegations that the adviser engaged in sales practice violations resulting in compensatory damages of at least $5,000, forgery, theft, or misappropriation, or conversion of funds or securities, which was closed without action, withdrawn, or denied.
Customer Dispute - Pending	This type of disclosure event involves (1) a pending consumer-initiated, investment-related arbitration or civil suit that contains allegations of sales practice violations against the adviser or (2) a pending, consumer-initiated, investment-related written complaint containing allegations that the adviser engaged in sales practice violations resulting in compensatory damages of at least $5,000, forgery, theft, or misappropriation, or conversion of funds or securities.
**Employment Separation After Allegations*	This type of disclosure event involves a situation in which the adviser voluntarily resigned, was discharged, or was permitted to resign after being accused of (1) violating investment-related statutes, regulations, rules, or industry standards of conduct; (2) fraud or the wrongful taking of property; or (3) failure to supervise in connection with investment-related statutes, regulations, rules, or industry standards of conduct.
Financial - Final	This type of disclosure event involves a bankruptcy, compromise with one or more creditors, or Securities Investor Protection Corporation liquidation involving the adviser or an organization the adviser controlled that occurred within the last ten years.
Financial - Pending	This type of disclosure event involves a pending bankruptcy, compromise with one or more creditors, or Securities Investor Protection Corporation liquidation involving the adviser or an organization the adviser controlled that occurred within the last ten years.

Investigation	This type of disclosure event involves any ongoing formal investigation by an entity such as a grand jury, state or federal agency, self-regulatory organization, or foreign regulatory authority. Subpoenas, preliminary or routine regulatory inquiries, and general requests by a regulatory entity for information are not considered investigations and, therefore, are not included in a BrokerCheck report.
Judgment/Lien	This type of disclosure event involves any unsatisfied and outstanding judgments or liens against the adviser.
**Regulatory - Final*	This type of disclosure event involves (1) a final, formal proceeding initiated by a regulatory authority (e.g., a state securities agency, self-regulatory organization, federal regulatory agency such as the SEC, foreign financial regulatory body) for a violation of investment-related rules or regulations or (2) a revocation or suspension of an adviser's authority to act as an attorney, accountant, or federal contractor.
Regulatory - On Appeal	This type of disclosure event involves (1) a formal proceeding initiated by a regulatory authority (e.g., a state securities agency, self-regulatory organization, federal regulatory agency such as the SEC, foreign financial regulatory body) for a violation of investment-related rules or regulations that is currently on appeal or (2) a revocation or suspension of an adviser's authority to act as an attorney, accountant, or federal contractor that is currently on appeal.
Regulatory - Pending	This type of disclosure event involves a pending formal proceeding initiated by a regulatory authority (e.g., a state securities agency, self-regulatory organization, federal regulatory agency such as the SEC, foreign financial regulatory body) for alleged violations of investment-related rules or regulations.

Appendix B. Variable Definition

This table describes the definitions and data sources of the outcome variable and each input variable. The prefix of the variable indicates its level (i.e., FA, firm, county, and state).

Variable	*Definition*	*Source*
Outcome Variables		
FA: Misconduct	Dummy variable takes on a value of one if the FA has at least one misconduct disclosure in a year.	IAPD; BrokerCheck
Baseline Input Variables		
FA: Prior misconduct	Dummy variable takes on a value of one if the FA has at least one prior misconduct record.	IAPD; BrokerCheck
FA: Experience	The number of years since the FA passed the first qualification exam.	IAPD; BrokerCheck
FA: Female	Dummy variable takes on a value of one if the FA is female.	IAPD; BrokerCheck; GenderChecker; genderize.io
FA: Dually registered	Dummy variable takes on a value of one if the FA is registered as both a broker and an investment advisor.	IAPD; BrokerCheck
FA: Uniform Investment Adviser Law (65)	Dummy variable takes on a value of one if the FA passed the *Uniform Investment Adviser Law Exam* in or before the current year.	IAPD; BrokerCheck
FA: Uniform Combined State Law (66)	Dummy variable takes on a value of one if the FA passed the *Uniform Combined State Law Exam* in or before the current year.	IAPD; BrokerCheck
FA: Securities Agent State Law (63)	Dummy variable takes on a value of one if the FA passed the *Securities Agent State Law Exam* in or before the current year.	IAPD; BrokerCheck
FA: General Securities Rep. (7)	Dummy variable takes on a value of one if the FA passed the *General Securities Representative Exam* in or before the current year.	IAPD; BrokerCheck

FA: Investment Company Product Rep. (6)	Dummy variable takes on a value of one if the FA passed the *Investment Company Product Representative Exam* in or before the current year.	IAPD; BrokerCheck
FA: General Securities Principal (24)	Dummy variable takes on a value of one if the FA passed the *General Securities Principal Exam* in or before the current year.	IAPD; BrokerCheck
FA: No. of other qualifications	The number of qualifications other than Series 6, 7, 24, 63, 65, and 66 that the FA possesses.	IAPD; BrokerCheck
Firm: No. of advisors	The number of FAs of the advisory firm.	Form ADV filings
Firm: No. of accounts	The number of accounts under the advisory firm's management.	Form ADV filings
Firm: No. of registered states	The number of states in which the advisory firm is registered.	Form ADV filings
Firm: No. of social network links	The number of social network links (e.g., Twitter, Facebook, LinkedIn, etc.) the advisory firm has.	Form ADV filings
Firm: No. of business lines	The number of types of businesses the advisory firm conducts.	Form ADV filings
Firm: Corporation	Dummy variable takes on a value of one if the advisory firm's formation type is *Corporation*.	Form ADV filings
Firm: Limited liability	Dummy variable takes on a value of one if the advisory firm's formation type is *Limited liability*.	Form ADV filings
Firm: Referral business	Dummy variable takes on a value of one if the advisory firm has referrals or financial arrangements with other brokers or dealers.	Form ADV filings
Firm: Retail	Dummy variable takes on a value of one if the advisory firm advises non-high-net-worth individuals.	Form ADV filings
Firm: AUM	Dummy variable takes on a value of one if the advisory firm charges fees based on AUM.	Form ADV filings
Firm: Hourly	Dummy variable takes on a value of one if the advisory firm charges hourly fees.	Form ADV filings
Firm: Fixed	Dummy variable takes on a value of one if the advisory firm charges fixed fees.	Form ADV filings
Firm: Commission	Dummy variable takes on a value of one if the advisory firm charges fees based on commission.	Form ADV filings

Firm: Performance	Dummy variable takes on a value of one if the advisory firm charges fees based on performance.	Form ADV filings
Additional County/State Variables		
County: Population	The total population of the county.	American Community Survey
County: Med. household income	The median household income of the county.	American Community Survey
County: Med. age	The median age of the county.	American Community Survey
County: Male/Female	The ratio of male to female population of the county.	American Community Survey
County: % white	The percentage of the white population of the county.	American Community Survey
County: % born foreign	The percentage of the population of the county born in a foreign country.	American Community Survey
County: % less than "very well" English	The percentage of the population of the county speaks less than "very well" English.	American Community Survey
County: % bachelor+	The percentage of the population of the county with an above bachelor degree.	American Community Survey
State: Govt legal spending (in Thousands)	The annual government administration spending on judicial and legal matters is by each state (in thousands).	U.S. Census Bureau, Annual Surveys of State and Local Government
Additional Firm Variables		
Firm: Misconduct ratio	The proportion of FAs with at least one misconduct disclosure in the advisory firm in a year.	IAPD; BrokerCheck
Firm: Prior misconduct ratio	The proportion of FAs with at least one prior misconduct record in the advisory firm.	IAPD; BrokerCheck

Firm: Owner/Executive prior misconduct	Dummy variable takes on a value of one if at least one of the owners or executives of the advisory firm has at least one prior misconduct record.	IAPD; BrokerCheck
Firm: Experience	The average working experience of FAs in the advisory firm.	IAPD; BrokerCheck
Firm: Female ratio	The proportion of female FAs in the advisory firm.	IAPD; BrokerCheck; GenderChecker; genderize.io
Additional FA Variables		
FA: Personal website	Dummy variable takes on a value of one if the FA has a personal website.	Discovery
FA: No. of language	The number of languages the FA speaks.	Discovery
FA: Income	Take the value of 5, 4, 3, 2, and 1 if the inferred income of the FA is [$1m+], [$501 -$1m], [$251k-$500k], [$101k-$250k], and [<$100k], respectively.	Discovery (Interests & Lifestyle)
FA: Wealth	Take the value of 4, 3, 2, and 1 if the inferred wealth segment of the FA is [$2m-5m], [$500k-2m], [$100k-500k], and [<$100k], respectively.	Discovery (Interests & Lifestyle)
FA: Investment	Dummy variable takes on a value of one if the FA owns investments in any form (e.g., stock, bonds).	Discovery (Interests & Lifestyle)
FA: Home owner	Dummy variable takes on a value of one if the FA owns (rather than rents) a home.	Discovery (Interests & Lifestyle)
FA: Swimming pool	Dummy variable takes on a value of one if the FA owns a swimming pool.	Discovery (Interests & Lifestyle)
FA: Recreational vehicle	Dummy variable takes on a value of one if the FA owns a recreational vehicle.	Discovery (Interests & Lifestyle)
FA: Interest	Dummy variable takes on a value of one if the FA discloses an interest in any form (e.g., reading, music, and sport).	Discovery (Interests & Lifestyle)
FA: Contribution	Dummy variable takes on a value of one if the FA contributes in any form (e.g., political, religious).	Discovery (Interests & Lifestyle)
FA: Military	Dummy variable takes on a value of one if the FA has military experience.	Discovery (Interests & Lifestyle)

FA: Soho business	Dummy variable takes on a value of one if the FA operates a business out of the home.	Discovery (Interests & Lifestyle)
FA: Regular credit card	Dummy variable takes on a value of one if the FA owns a regular credit card.	Discovery (Interests & Lifestyle)
FA: Premium credit card	Dummy variable takes on a value of one if the FA owns a premium credit card.	Discovery (Interests & Lifestyle)
FA: Pets	Dummy variable takes on a value of one if the FA owns a pet.	Discovery (Interests & Lifestyle)

Appendix C. The Sample

This table outlines the process of data filing to obtain our estimation sample.

	No. Obs.	*No. of FAs*	*No. of Advisory Firms*
	[1]	[2]	[3]
The initial sample extracted from the Form ADV filings			11,555
Exclude observations with missing FA information to construct:			
25 baseline inputs	3,639,236	693,195	7,136
25 baseline inputs + 9 county/state inputs	3,309,736	653,932	6,955
25 baseline inputs + 5 firm inputs	3,639,236	693,195	7,136
25 baseline inputs + 15 FA inputs	1,357,224	201,661	5,934
25 baseline inputs + all additional inputs	1,224,730	189,491	5,802

Appendix D. An Example of GPT-4o Prompt

FINRA mandates that all registered FAs "*disclose customer complaints and arbitrations, regulatory actions, employment terminations, bankruptcy filings, and criminal or judicial proceedings*." Out of the 23 disclosure categories outlined by FINRA, we categorize six as misconduct disclosures. These include Civil-Final, Criminal-Final Disposition, Customer Dispute-Award/Judgment, Customer Dispute-Settled, Employment Separation After Allegations, and Regulatory-Final. If an FA is predicted to commit misconduct, output one. If not, output zero. Only output the number zero or one. Please predict whether the FA will engage in misconduct behavior in the next year based on the following information:

(1) The FA has not engaged in misconduct before;

(2) The FA has worked for 1.0 years;

(3) The FA is a male;

(4) The FA is not registered as an investment advisor and broker;

(5) The FA has not passed the S65 exam;

(6) The FA has not passed the S66 exam;

(7) The FA has not passed the S63 exam;

(8) The FA has not passed the S7 exam;

(9) The FA has not passed the S6 exam;

(10) The FA has not passed the S24 exam;

(11) The FA has passed 1.0 exams other than the six mentioned above;

(12) There are 1132.0 advisors in the company the FA works for;

(13) There are 57712.0 accounts the company has;

(14) The company has registered in 49.0 states;

(15) There are 7.0 business links the company has;

(16) There are 11.0 business lines the company has;

(17) The company is a corporation;

(18) The company is not a limited liability company;

(19) The company has referrals or financial arrangements with other brokers or dealers;

(20) The company advises non-high-net-worth individuals;

(21) The company charges based on assets under management;

(22) The company charges an hourly fee;

(23) The company charges a fixed fee;

(24) The company charges based on commission;

(25) The company charges based on performance.