Culture ‘profiling’, AI and AML: Efficacy vs Ethics

John W. Goodell[a], Cal Muckley[b], Parvati Neelakantan[c], Darragh Ryan[b], Pei-Shan Yu[b]

[a] *The University of Akron, Akron, Ohio, USA*
[b] *University College Dublin, Belfield, Dublin, Republic of Ireland*
[c] *Indian Institute of Technology Kanpur, Kanpur, India*

Abstract

Using extensive transaction and money laundering detection data, at a globally important financial institution, we investigate the efficacy of including aspects of national culture in formulating anti-money laundering predictions. For corporate and individual accounts, Hofstede individualism scores of the country in which a customer is resident, or from which a wire is sent/received, are of first-order importance. When combined with account and transaction data; as well as even a proprietary institutional algorithm, individualism scores continue to determine the models’ predictive performances. Our finding of the efficacy of profiling in AML compliance underscores the need for stringent and enforced data protection safeguards, which can serve to ensure an individual’s fundamental right to privacy.

JEL Classification: C52, C55, D12, G17, G21

Keywords: Financial Institutions, National Culture, Machine Learning, Anti-Money Laundering

---

"Transaction monitoring using AI and machine learning tools may allow regulated entities to carry out traditional functions with greater speed, accuracy and efficiency"

Financial Action Task Force

Opportunities and Challenges of new technologies for AML/CFT

July, 2021

1. Introduction

Money laundering legitimizes dirty money, i.e., money generated from illegal activities.[1] The improvement of anti-money laundering (AML) risk management systems is of paramount importance.[2] However, this should not occur in breach of the fundamental human right of an individual to privacy. In this paper, we test a commercial incentive, in AML, to profile customers. We examine whether national culture traits, in particular, the trait of individualism (Hofstede, 2001), can serve to improve the efficacy of bank-level machine learning informed alert models to mitigate money laundering. Evidence of successful profiling would suggest that there is a commercial incentive to conduct it. This would underscore the importance of improved guidance, in AML compliance, in the implementation of data privacy protection law, and the enforcement of this guidance.

We begin by explaining why national culture facets might prove efficacious regarding the detection of money laundering. There are two channels which show that national culture facets can influence the decision making of an individual, whether a private person or a corporate representative. These channels comprise a dual process understanding of cognition for an individual: personal attitudes and values (Fischer et al., 2010; Peterson and Barreto, 2018), at one level, together with societal culture facets (Peterson and Barreto, 2018; Watts et al., 2020), at another level. In

[1] In so doing, it integrates such monies into an established financial system for subsequent use without evoking suspicion (IMF, 2021).

[2] Although difficult to measure, the United Nations Office on Drugs and Crime (UNODC) estimates that the amount of money laundered globally every year is 2 to 5 percent of global GDP. The invidious upshot of money laundering includes the perpetuation of crime, the misallocation of capital, the compromising of customer welfare, and even the potential to undermine economies and international financial stability (e.g. Financial Stability Board (2018)).

this theoretically insightful work, it is shown that these channels inform an individual's cognition and, thus, we argue, they can inform opportunistic financial decisions.

An expanding body of empirical scholarship, indeed, indicates that national culture can explain financial decisions (Chui et al., 2002; Kwok and Tadesse, 2006; Chui et al., 2010; Lievenbrück and Schmid, 2014; Fidrmuc and Jacob, 2010; Orlova et al., 2017), and that it can, moreover, inform opportunistic decision making (Fisman and Miguel, 2007). In particular, it can account for opportunistic decision making in relation to financial services (DeBacker et al., 2015; Liu, 2016a; Conlon et al., 2024). To elaborate, national culture facets have been shown to relate to corporate capital structure decisions (Chui et al., 2002), the configuration of financial systems (Kwok and Tadesse, 2006), momentum strategies (Chui et al., 2010), corporate financial hedging (Fidrmuc and Jacob, 2010), and cash holding decisions (Orlova et al., 2017). Opportunistic decision making has been shown to be related to culture, such as those of United Nations' diplomats to illegally park in New York (Fisman and Miguel, 2007).[3] In respect to financial services, national culture has been shown to influence individuals' opportunistic decisions, such as tax evasion (DeBacker et al., 2015), financial wrongdoing (Liu, 2016a) and bank employees' financial misconduct (Conlon et al., 2024). Due to this theoretical and empirical evidence, we argue that it is insightful to assess the scope for national culture traits to improve the estimation of individual-level money laundering propensities.

To examine the efficacy of national culture, as a lead indicator of money laundering, we examine a major global financial institution's proprietary dataset. It contains cross-border wire transactions made during 2009-2018. The dataset pertains to 153,917 money laundering alerts and 2,206 analyst money laundering suspect issue cases associated with international wire transfers to and from customers of the institution. The financial institution's monitoring system generates an alert for a wire transfer, if the wire breaches rules related to jurisdiction and the amount of funds. The alert is then investigated by a team of expert analysts. In their judgement, if the

[3] Studies indicate that national culture dimensions influence perceptions of what constitutes unethical behaviour (Vitell et al., 1993; Volkema, 2004).

corresponding wire transaction is suspicious, then they escalate it to an issue case and refer the matter to higher authorities for further investigation. The issue cases can be regarded as a precursor to the establishment of a money laundering incident.

We employ supervised machine learning techniques to predict money laundering at the financial institution i.e. which alerts are escalated to issue cases regarding money laundering instances. We examine the contribution of the features, including national culture traits, to the machine learning predictions of instances of money laundering. Our study incorporates as our principal point of focus, Hofstede's individualism cultural dimension of bank customers.

Following prior literature that relates national culture with financial decisions, ethical values and discernment, and corporate misconduct, we control for other cultural and related variables. We include masculinity, uncertainty avoidance, and power distance Hofstede cultural traits of the bank customers' resident countries and the origin/destination countries of the wire transaction. We, further, include two internationally recognized indices that measure the levels of corruption and financial secrecy of a country, namely, the Corruption Perception Index and the Financial Secrecy Index. We also control for account- and transaction-level predictors.

We provide a brief outline of our findings. We find country-level factors, specifically the individualism national culture trait, as comprising strong predictive capacity to identify suspect bank wire transfers. We report high performing models (e.g. TPR and AUC performance metrics using versions of the cross validation resampling procedure), across sets of lead indicators and machine learning algorithms. Turning to the ranking of our lead indicators of money laundering, for corporate accounts, Hofstede individualism scores of the country in which a customer is resident, *and* from which a wire is sent/received are the most important factors. Of the twelve country-level factors, the two Hofstede individualism scores rank, across machine learning models, as the most important and second most important lead indicators of money laundering issue cases. For individual accounts, individualism scores of the country in which a customer is resident rank, across machine learning models, in the top four of the twelve factors.

We show, further, that when national culture is combined with extensive country, account, transaction, and even proprietary risk score data[4], its inclusion still markedly enhances the predictive capacity of the models. While the Hofstede individualism scores of the country of residence of individual customers ranks only moderately well (about half way up the table of 25 factors), the same scores for corporate accounts are ranked, across machine learning models, consistently in the top three of the twenty five factors. If we exclude the opaque proprietary risk score data, the corporate representative's individualism score rankings remain of first order importance (top four of 24 factors). The individualism scores of the country in which an individual customer is resident rank, across the machine learning algorithms, well inside the top tercile of factors. It achieves first rank importance in our best performing model, the gradient boosting machine learning specification.

Our paper's principal contributions are fourfold. *First*, with the exponential growth of proprietorial transaction-level data (FATF and Egmont Group, 2020), financial institutions are increasingly applying machine learning techniques to fulill 'know your customer' protocols and expedite and scale the detection and prevention of money laundering (FATF, 2021).[5] We contribute to this work and test if publicly available data on national culture can supplement proprietorial account- and transaction data to improve the predictive capacity of models to detect money laundering. Our models have been able to successfully detect suspicious transactions, which can save analysts' a significant amount of time and resources which they might otherwise spend going through vast amounts of customer data to detect suspicious transactions. The use of demographic inputs, particularly country-level factors in machine learning models, touches on a wide variety of literature that incorporate cultural and demographic variables that ascribe value to these characteristics.[6] We establish, thus, the potential importance, in AML protocols, of country level and publicly available variables, in particular the national culture trait of individualism. We show that such variables can inform AML machine learning protocols even in proprietorial customer account and transaction

[4]The proprietary risk score is the risk score assigned to the alerts by the financial institution's proprietary alert algorithm.

[5]The total cost of financial crime compliance across financial institutions worldwide, largely in relation to countering money laundering, stands at about $274.1 billion in 2022.

[6]See, for instance, the area of microfinance, where the notion of female borrowers being more trustworthy undergirds the industry (Aggarwal et al., 2015).

level data.

Second, we contribute relative to Liu (2016b) and DeBacker et al. (2015) who find that when individuals emigrate their beliefs and values accompany them. These beliefs and values, inherent in the individuals' cultures, can influence their decisions and, thus, impact the decisions of corporations which employ them. Given this reported inter-generational (Liu, 2016b) impact of national culture, even in the context of a new host country economic, social and institutional environment, on corporate outcomes, we look to test whether national culture facets of customers who have not necessarily emmigrated can influence decisions. Specifically, we examine if the national culture, in the country where a customer resides, or in the customer's counterparty country in the wire transaction, can influence an individual's decision to breach money laundering regulation.

Third, national culture has also been shown to influence individuals' opportunistic decisions related to bank employees' financial misconduct (Conlon et al., 2024). Unlike Conlon et al. (2024) who show that national culture, an immutable aspect of organisational culture, accounts for regulatory sanctions regarding malfeasance in banks globally. Our focus is on a single bank's team of money laundering analysts. This serves to alleviate confounding concerns about the heterogeneity of the rule of law and enforcement internationally (Cumming et al., 2018).

Finally, we describe the ethical implications of any such enhanced AML system. We elaborate on an ethical tension between the collective need to mitigate money laundering, and its harmful consequences, and an individual's right to privacy, regarding their national culture traits. If the inclusion of national culture in a fraud detection model materially improves the performance of that model and alleviates the suffering that can ensue due to money laundering, then its inclusion is ethically defensible. In this paper, we nonetheless abstract from this ethical tension. Enshrined in law is the fundamental right to privacy of individuals and yet, at least in an AML compliance context, there is a paucity of guidance regarding the nature of requisite safeguards, and the degree to which any such safeguards are enforced.[7] We acknowledge, hence, that the national culture trait

[7]The European Union's 2015 fourth anti-money laundering directive, for instance, required financial institutions to apply data protection safeguards to their anti-money laundering/countering the financial terrorism compliance programs. There was, however, inadequate guidance provided regarding how this should be conducted. See e.g.

is a sensitive personal trait. Its use can compromise a fundamental right of an individual to privacy. Our paper serves to show the potential efficacity of including national culture profiling in AML despite data privacy concerns. Our work hence underscores the importance of guidance regarding how AML systems should protect an individual's fundamental right to privacy.

The rest of this paper is organized as follows. Section 2 discusses the relation between national culture, individual decision making and, in particular, it descibes our individualism-malfeasance conjecture. Section 3 discusses the proprietary dataset, country-specific culture, and institution quality indices from which we have drawn our predictors/features. Section 4 outlines the various data resampling methods used in the paper for meaningfully sourcing information from the data, and discusses the machine learning methodologies, performance evaluation metrics, and feature importance metrics of the study. Section 5 presents the empirical findings. Section 6 discusses ethical implications of ascribing values, against a global standard, to dimensions of national culture to inform anti-fraud alert models. Finally, section 7 concludes.

2. Hofstede's individualism, individual decision making and money laundering

In this section, we examine Hofstede's national culture trait of 'individualism'. We describe, in particular, how this trait can inform individual decision making and result in financial malfeasance, e.g. money laundering, or the mitigation of the laundering of criminal proceeds.

2.1. National culture and individual risk-taking and decisions

Our rationale for investigating the utility of national culture to mitigate money laundering stems from prior literature. That literature identifies two channels that show how national culture facets can influence an individual's decision-making, and risk-taking behavioour. *First*, the contextual effects of societal institutions and norms can inform "individuals experience and, hence, what people unconsciously intuit and consciously understand." This can, in turn, contribute to an individual's cognition and decision-making (Peterson and Barreto, 2018; Goodell, 2019). In line with argumen-

https://iapp.org/news/a/data-protection-and-the-eus-anti-money-laundering-regulation/

tation in Peterson and Barreto (2018), national culture facets can indicate contextual characteristics that more strongly shape an individual's cognition, than do consciously expressed personal values. For instance, these authors find that it can be comparatively insightful to know of the societal context in which an individual was raised, as opposed to her self-professed attitudes and values, with a view to predicting, in a corrupt system, her interaction with regulations.

Previous studies show, *secondly*, that national culture facets can approximate values at the individual level (for example, Fischer et al. (2010)). These studies report strong evidence of the structural similarity in values at the individual and country levels. While the assignment of country level scores to our samples necessarily neglects within-country variability (Kirkman et al., 2006), Fischer et al. (2010) report such values data can serve as a robust approximation for the cultural knowledge, resources, structures and norms held by society (Peterson and Barreto, 2018). These national culture constructs can be primed and made temporarily accessible (Leung and Morris, 2015), indicating that they do manifest at the individual level. The highlighted channels testify to how national culture facets can influence an individual's decision-making and risk taking.

2.2. Hofstede's Individualism and money laundering

Hofstede's individualism culture index provides insights into people's level of interdependency in a society. In an individualistic society, people are expected to fend for themselves and their immediate families; whereas in a collectivist society, in-groups, to which people belong, look after individuals in exchange for unquestioning loyalty. Individualism is linked to behavioral attributes of over confidence and self-attribution bias i.e., people's tendency to attribute positive events to their own character but attribute negative events to external factors (for example, Chui et al. (2010); Li et al. (2013)). Due to these behavioural attributes, such individuals show low levels of self-monitoring (Biais et al., 2005), and they are over-optimistic in respect to the precision of their predictions (Van den Steen, 2004). This can give individuals over-optimistic views of the future (Fischer and Chalmers, 2008) and lead to their inaccurate evaluation of bad news (Kim et al., 2016). As a result, we expect that bank customers, in more individualistic countries, can overestimate their abilities (Heine et al., 1999; Markus and Kitayama, 1991) to opportunistically (Chen et al., 2002) disguise malfeasance, e.g., money laundering, so that financial institutions and regulators will not detect their behavior.

Further, because individualistic cultures value self-reliance, freedom, achievement, and tend to consider its people's actions as beyond reproach, prior studies argue that individualistic cultures promote ethically questionable behavior (Bame-Aldred et al., 2013; Cullen et al., 2004; Martin et al., 2007; Vitell et al., 1993). Additionally, individualism is also linked to risk-taking behavior (Gaganis et al., 2019; Mourouzidou-Damtsa et al., 2019), and one example of risk taking is taking action which increases likelihood of a regulatory sanction.

Also, Chui et al. (2010) and Kreiser et al. (2010) suggest that in more individualistic countries, decisions, in general, are more likely to be taken by individuals rather than the group. In such countries, people have a strong belief in individual choices and decisions (Markus and Kitayama, 1991). This is critical as, according to Shupp and Williams (2008), in high risk situations, individuals are more risk tolerant than groups. Similarly, the incentive to perform in individualistic countries is underpinned by compensation practices that focus on individual recognition (Schuler and Rogovsky, 1998). As incentivized and risky decision making is more likely in individualistic cultures, we expect that money laundering can also be more likely in this setting.[8]

Collectively, the above arguments suggest that a banking customers' predilections for committing money-laundering can be due to cross-country cultural differences linked to that facet of national culture known as individualism. Individualism scores pertaining to a customer's country of residence and/or the country of wire origination/destination can prove in detecting money-laundering.

However, it is noteworthy that prior literature also identifies a countervailing outcome between individualistic cultures and its peoples' ethically questionable behavior. For instance, some studies find a negative association between individualism and tax evasion (Bame-Aldred et al., 2013;

[8]National culture also figures prominently in assessing ethical values and discernment in business ethics research. For instance, Vitell et al. (1993) propose a theoretical framework to understand the association between culture and ethical decision-making in business. They argue that in individualistic societies, business practitioners are less likely to adhere to both formal codes of ethics and informal norms than their counterparts in countries that value collectivist ethos.

Tsakumis et al., 2007; Richardson, 2008). Further, Cullen et al. (2004) detect a negative relationship between individualism and managers' ethically questionable decisions. Similarly, Armstrong (1996) notes a positive association between individualism and higher ethical standards. Therefore, one can put forward an alternative hypothesis that the individualism score of a bank customer's country is negatively related to her predilection for money-laundering. It is, therefore, a moot question whether individualism is associated with fraud.

3. Data

This study employs a major global financial institution's proprietary dataset consisting of cross-border wire transactions made during 1 January 2009- 31 December 2018. The data pertains to alerts generated by international wire transfers both to and from customers of that institution. An alert is generated for a wire transfer in the financial institution's monitoring system, if the wire amount exceeds a predetermined threshold and if the country from which the wire is sent and/or received falls in the list of countries blacklisted by the financial institution. The alert is then investigated by a team of experts. In their judgement, if the corresponding wire transaction seems highly suspicious, then they escalate it to an issue case and refer the matter to higher authorities for further investigation.[9] The *issue case* is the dependent variable employed in our study.

Alerts are generated for more than 60,000 customer accounts in the data. These accounts can be broadly classified into six account registration types.[10] Among the six account registration types, we focus only on two, the corporate- and people-related account registrations, since these pertain to 78.23% of the alerts and 93.77% of the issue cases. Table 1, Panel A reports the number of alerts and subsequent issue cases over time, associated with the corporate- and people-related accounts.[11]

[9]Far from efficient, this method of screening transactions for suspicious activity remains standard across the industry, since financial governing bodies enforce harsh penalties on institutions that they deem to be lax in detecting money laundering. The proportion of false alarms typically exceeds 99%. To avoid confusion, we reserve the use of the term "false positives," for reference to the model results.

[10]See Internet Appendix A.

[11]The total number of alerts generated during the ten-year period is 206,751. In considering only the corporate- and people-related account registration types, however, the total number of alerts reduces to 153,917. See Panel B in Table

3.1. Sample Selection

The dataset provides information on the wire transactions that generated the alerts and the corresponding customers' accounts and transaction history. However, we do not have complete information on customers' account- and transaction-level data. Considering this deficiency, only 60% of the alerts could be matched to the wire transactions that triggered the alerts. Table 1, Panel B reports the number of alerts and subsequent issue cases that can be successfully matched with the corporate- and people-related account registration types.

[Please insert Table 1 about here.]

3.2. Feature Selection

In creating features from the data of the financial institution, we account for customers' account and transaction history. However, our chief novelty consists in creating features from country-specific culture and institution quality indices which go into investigating if banking customers' socio-cultural matrix influences their predilections for committing financial misconduct, namely, money-laundering. We further group the features into three categories: (1) Country-level, (2) Account-level, and (3) Transaction level. Below, we discuss the features included in our study. Concise definitions are provided in Table 2.

[Please insert Table 2 about here.]

3.2.1. Country-level Predictors

We create quantifiable country-level features from internationally recognised country-specific culture and institution quality indices. Corresponding to each of the indices, we create two features which we distinguish as the origin/destination country of the wire transaction and the residence country of the customer receiving/sending the wire that triggered an alert. We create two sets of features for each index, since in the financial institution's data a customer's residence country is documented precisely, though often the data is unclear on the country to/from which the customer

1.

is sending/receiving the wire.[12] Thus, we distinguish the features constructed from a particular index by using the subscripts R and W; where R and W denote customer's country of residence and the country of wire origin/destination, respectively. Below, we define the four country-specific culture indices and two institution quality indices employed to create features in our study.[13]

1. **Individualism Index** (IDV_R, IDV_W): This index provides insights into the level of people's interdependency in a society. In an individualistic society, people are expected to fend for themselves and their immediate families; whereas in a collectivist society, in-groups to which people belong look after them in exchange for unquestioning loyalty. A country scoring high on this index exhibits individualistic trait, whereas a country with a low score cherishes a collectivist ethos.

2. **Masculinity Index** (MAS_R, MAS_W): This culture dimension quantifies the extent to which a society values achievement, success, and competition (masculine traits) over modesty and compassion towards others (feminine traits). A country scoring high on this index indicates that its people privilege masculine over feminine traits.

3. **Power-Distance Index** (PDI_R, PDI_W): The index provides insights on the level of inequality endorsed and accepted by the less powerful members of a society. A country with a low score on this index shows its citizens as having a lower tolerance for social inequality and vice-versa.

4. **Uncertainty Avoidance Index** (UAI_R, UAI_W):The index quantifies the impact of national culture on its peoples' tolerance to deal with uncertainty. Cultures that try to minimize ambiguity rank high on this index and vice-versa.

[12] For instance, in some wire transfers the IBAN of the customer sending/receiving the wire is documented, whereas in others we retrieve this information from the address/information field documented in the data.

[13] In this study we employ four national culture dimensions proposed by Hofstede (2001) and two internationally recognized indices that measure the levels of corruption and financial secrecy of a country.

5. **Corruption Perception Index** (CPI_R, CPI_W): Drawing on thirteen different data sources, Transparency International's composite index ranks countries/territories based on the perceived corruption in public sector by experts in governance and business climate analysis. The index ranks 180 countries on a scale of 0 to 100, where 0 corresponds to high perceived level of corruption and 100 to low perceived level of corruption, respectively.[14]

6. **Financial Secrecy Index** (FSI_R, FSI_W): Proposed by Tax Justice Network, the index ranks jurisdictions based on the scale of their offshore financial activities and the regulatory framework providing legal and financial secrecy to businesses and individuals based elsewhere. The index provides insights on global financial secrecy, tax havens, and illicit financial flows (Puspitasari et al., 2019; Houqe et al., 2015; Michalos and Hatch, 2019; Hassan and Giorgioni, 2015).[15]

3.2.2. Account-level Predictors

We employ four account-level features from the financial institution's dataset, namely *Customer Age*, *Account Age*, *Customer Net Worth*, and *Alert Supplier Code*. The feature *Customer Age* is defined as the age of the customer, a private individual, on the date when the alert is generated. We include this feature when detecting money-laundering exclusively for people-related accounts. Similarly, *Account Age* is the length of time an account stood registered from the date of alert. The feature *Customer Net Worth* is the aggregate balance on all the accounts of the customer. The *Alert Supplier Code* records the type of systemic method that the financial institution employs to collect alerts.

[14] Liu (2016) indicates that the Transparency Internationals Corruption Perception Index can "capture a general attitude toward opportunistic behavior at the country level."

[15] Hassan and Giorgioni, 2015 indicate that the Tax Justice Network's Financial Secrecy Index at country-level "indicates a lack of transparency and unwillingness to engage in effective information exchange, which makes a secretive country a more attractive location for routing illicit financial flows and for concealing criminal and corrupt activities."

3.2.3. Transaction-level Predictors

For each wire transaction that triggered an alert, we have information on, over a 180-day period preceding the alert, the number of incoming and outgoing wires, and transfers to and from the corresponding customer's account (TFI_{180}, TFO_{180}); the aggregated amount of incoming and outgoing wires and transfers ($\sum TFI_{180}, \sum TFO_{180}$); the number of incoming and outgoing checks (CKI_{180}, CKO_{180}), and the aggregated amount of incoming and outgoing checks ($\sum CKI_{180}, \sum CKO_{180}$).

4. Methodologies

In this section of the paper, we discuss the various data resampling methods for meaningfully inferring information from the data. It then focuses on the machine learning methodologies and the performance evaluation metrics used to evaluate them. Finally, it discusses the feature importance method. We discuss the data resampling techniques in subsection 3.5.1, machine learning methodologies in subsection 3.5.2, performance evaluation metrics in subsection 3.5.3, and feature importance in subsection 3.5.4.

4.1. Data Balancing

The dependent variable suffers from severe class imbalance. In other words, the number of observations that belong to the positive class (issue case) is significantly lesser than those that belong to the negative class (generated alert is not an issue case). Models trained on such data in prioritizing the prevalent class over the minority class leads to an overly optimistic measure of accuracy (Batista et al., 2004). While such models can detect a non-fraudulent transaction with high level of accuracy, they often fail to detect highly suspicious transactions. Failure to detect highly suspicious transactions could pose a threat to the financial institutions' professional credibility and may also lead to significant regulatory penalties.

In this study, we avail of various data-resampling techniques for overcoming the challenges posed by the imbalanced class distribution. Below, we discuss the resampling techniques employed in our study.

1. **Under sampling:** This technique randomly discards observations from the majority class to better balance the skewed distribution. In reducing the majority class's size to match the minority class, this technique, however, forgoes potentially useful information from the majority class.

2. **Hybrid sampling:** Combining under-sampling and over-sampling methods,[16] this technique applies under-sampling technique to the majority class and over-sampling technique to the minority class to balance the class distribution.

3. **Synthetic sampling:** This technique works like over-sampling. However, instead of randomly duplicating observations from the minority class, it introduces artificial noise to perturb its predictor values to avoid over-fitting. In our study, we use ROSE (Random Over-sampling Examples) synthetic sampling method. This method utilizes the hybrid-sampling technique besides synthetic sampling to overcome the computational challenges of a much larger data set.

4.2. Machine Learning Methodologies

In this subsection, we discuss the machine learning algorithms, namely logistic regression, random forests, support vector machines, and gradient boosted machines, employed in our study to detect money-laundering at the financial institution.

4.2.1. Logistic Regression

Logistic regression (LR) models the probability of an observation belonging to a particular class. It employs the logistic function,

$$p(X) = \frac{e^{\beta_0+\beta_1 X_1+\ldots+\beta_p X_p}}{1+e^{\beta_0+\beta_1 X_1+\ldots+\beta_p X_p}} \tag{1}$$

[16]Over-sampling: This technique randomly duplicates observations from the minority class to match the majority class size. We refrain from employing this technique as it can be computationally expensive (in cases of severe class imbalance, it may almost double the size of the dataset) and it often leads to overfitting the model.

to model the probability of the categorical response variable, Y. In the above logistic function $X_1, X_2, \ldots, X_p$ are the p features and we estimate the following equation

$$ln(\frac{p(X)}{1-p(X)}) = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p \tag{2}$$

We estimate the coefficients using the Maximum likelihood method. After the coefficient estimation, we select a suitable probability threshold to classify observations to the two distinct classes.

4.2.2. Random Forest

Random Forest (Breiman, 2001) algorithm in generating multiple decorrelated decision trees averages their predictions to yield a single prediction. Relying on the premise that averaging a set of independent observations having equal variances, this algorithm decreases the variance of the mean of the observations. The algorithm first generates a large number, say 'B,' bootstrapped samples from the training dataset. It then fits and trains the Decision Tree model on each of these B bootstrapped samples. The algorithm fits the decision trees on to the bootstrapped samples such that a random sample of 'm' features are considered as split candidates every time a split is made, rather than the entire set of features.[17] By drawing a fresh sample of 'm' features, the algorithm allows every feature to be considered for a split. This in turn produces uncorrelated decision trees which result in uncorrelated predictions. Averaging these uncorrelated predictions leads in a reduction of the variance of the ensemble method.

4.2.3. Support Vector Machines

Support Vector Machine (SVM) builds on the Maximal Margin Classifier algorithm applied in classifying linearly separable observations. Since most datasets cannot separate the observations by a linear boundary, the Maximal Margin Classifier has limited applications. By introducing Soft Margin and Kernel concepts to the Maximal Margin Classifier, SVM can classify observations with non-linear decision boundaries.

[17] Anytime a split is made, a fresh sample of random 'm' features are chosen for split consideration. Generally, 'm' is the square root of the total number of features.

SVM is a classifier/hyperplane such that an observation belonging to one side of the hyperplane is classified as class 1; if the observation belongs to the other side, then it is classified as class 2. Thus, an observation X belongs to class 1 if, say, for example,

$$f(X) > 0 \tag{3}$$

And it belongs to class 2 if,

$$f(X) < 0 \tag{4}$$

Additionally, the magnitude $f(X)$ acts as a measure of confidence in the class assignment. If $f(X)$ is far from zero, then we can be confident about the class assignment. Whereas if $f(X)$ is close to zero, then the class assignment may not be reliable.

4.2.4. Gradient Boosted Models

GBM algorithm consists in fitting weak predictive models sequentially to the ensemble such that their inclusion improves the predictive performance of the whole ensemble. The weak predictive models are constructed such that these models and the negative gradient of the loss function associated with the whole ensemble are maximally correlated (Friedman, 2001). The ensemble is updated as follows,

$$f_m(\mathbf{x}) = f_{m-1}(\mathbf{x}) + \beta_m h(\mathbf{x}; \mathbf{a}_m) \tag{5}$$

In the above function, $h(\mathbf{x}; \mathbf{a})$ is a parameterized function of the explanatory variables $\mathbf{x}$, characterized by the parameters $\mathbf{a} = \{a_1, a_2, \dots\}$. In our case, $h(\mathbf{x}; \mathbf{a}_m)$ is a shallow classification tree and therefore the parameters $\mathbf{a}_m$ are the split variables, split locations, and the modes of the terminal node for the individual trees.

Each "boost" is updated as follows,

$$f_m^*(\mathbf{x}) = -\rho_m g_m(\mathbf{x}) \tag{6}$$

where

$$g_m(\mathbf{x}) = [\frac{\partial \phi(f(\mathbf{x}))}{\partial f(\mathbf{x})}]_{f(\mathbf{x})=f_{m-1}(\mathbf{x})} = [\frac{\partial E_y[L(y, f(\mathbf{x}))|\mathbf{x}]}{\partial f(\mathbf{x})}]_{f(\mathbf{x})=f_{m-1}(\mathbf{x})} \tag{7}$$

is the gradient[18], and

$$\rho_m = argmin_\rho E_{y,\mathbf{x}} L(y, f_{m-1}(\mathbf{x}) - \rho g_m(\mathbf{x})) \tag{8}$$

is the "line search" along the direction of $-g_m$, and $L(y, f(\mathbf{x}))$ is the loss function.[19]

4.3. Model Evaluation

[Please insert Figure 1 about here.]

We now define our metrics, true positive rate (TPR)[20] and false positive rate (FPR)[21], with reference to the confusion matrix.

TPR measures the proportion of positive observations (*Issue Case*) correctly classified by a model:

$$TPR = \frac{TP}{(TP + FN)} \tag{9}$$

FPR measures the proportion of negative observations (*Non-Issue Case*) misclassified by a model:

$$FPR = \frac{FP}{(FP + TN)} \tag{10}$$

Both TPR and FPR lie between 0 and 1. Typically, we want TPR to be as high as possible and FPR to be as low as possible. However, these two metrics do not vary independently of each other, unless we deal with a perfect model. To achieve high TPR, we require a more sensitive model, though its inclusion would mean higher false positives, i.e., higher FPR. This trade-off is a general feature of any classification model.

---

[18] $\phi(f(\mathbf{x})) = E_y[L(y, f(\mathbf{x}))|\mathbf{x}]$ and $f_{m-1}(\mathbf{x}) = \sum_{i=0}^{m-1} f_i^*(\mathbf{x})$

[19] Since our response variable is binary, we consider the binomial loss function.

[20] Also referred to as sensitivity and recall

[21] Also referred to as fall-out

Most ML classification algorithms estimate the probability of an observation belonging to the positive class. Typically, a value of 50% is used as the probability threshold, i.e., an observation whose estimated probability greater than the threshold is assigned the positive class; whereas, if the estimated probability is less than the threshold, it is assigned the negative class. Lowering the threshold increases the number of true positives, though it also increases the number of false positives. Raising the threshold lowers the number of false positives, though it comes at the expense of reducing the number of true positives. Therefore, to measure the overall performance of a model, we plot the receiver operator characteristic (ROC) curve. ROC curve is the graphical representation of the relationship between the true positive rate and false positive rate, when the probability threshold is varied.

Figure 2 shows a typical ROC curve for a classification model. Each point on the curve provides the TPR (y-coordinate) and FPR (x-coordinate) corresponding to a probability threshold. Ideally, a model with TPR equal to 1 and FPR equal to 0 yields the best predictive capacity. However, in practice, we choose a model that hugs the top left corner of the ROC curve. Additionally, to measure the model's out-of-sample predictive performance we compute the area under the ROC curve (AUC). AUC is the probability that a random positive example (*Issue Case*) will be ranked above (*Non-Issue Case*) a random negative example (someone who is up to 4 years later not clinically diagnosed with ADRD). AUC lies between 0 and 1. A model with AUC of 0.5 is no better than randomly guessing (random classifier) the class for an observation; a model with AUC less than 0.5 performs worse than the random classifier; and a model with AUC greater than 0.5 demonstrates predictive capacity.

[Please insert Figure 2 about here.]

4.4. Predictor Importance

Finally, we investigate the relative importance of features in determining whether a transaction is fraudulent. In case of logistic regression, we use the statistical significance, and the magnitude of coefficient estimates to infer the relative importance of features. For random forests and gradient boosted machines, we estimate the total decrease in node purity corresponding to each predictor. Given that the SVM algorithm does not naturally extend itself towards estimating feature contribution, constructing a heuristic is in order. This method, unfortunately, does not provide consistent and reliable estimates. Therefore, we do not compute feature importance for the SVM model. We choose the models estimated on hybrid-sampled dataset to compute feature importance, since these models outperform the models fitted on datasets resampled by other techniques employed in this study.

5. Results

This section presents results of both our baseline empirical and robustness tests. We discuss the baseline results in subsection 3.6.1. The results of the robustness tests are discussed in subsections 3.6.2 and 3.6.3.

5.1. Model Performance and Interpretation

We first determine whether the various country-level features employed in our study can detect money-laundering at the financial institution. To meaningfully gauge the predictive capacity of these features, we first decompose our dataset into transactions involving private customers (people-related) and corporate clients (corporate-related).[22] We then train our models on the country-level attributes of the people-related, corporate-related, and combined dataset to estimate the out-of-sample performance of our models. We train 48 models; 4 machines learning algorithms trained on 3 datasets (people-related, corporate-related and the combined dataset) that are balanced by 4 balancing techniques. A randomized 50:50 split is performed on the datasets to create training and test datasets.[23] We further perform cross-validation to test the validity of our models and estimate relative importance of various country-level features.

[22] See Table 1 and Table A (Internet Appendix A).
[23] Except in the case of cross validation, where 80:20 and 90:10 splits are performed.

5.1.1. *Predictive capacity of Country-level features*

Table 3 shows the TPR, FPR, and AUC results of the models trained on the country-level features of the three datasets. Our features include the Hofstede country-specific culture dimensions and two institution quality indices for the customer's country of residence and origin/destination country of the wire.[24] These features are: CPI_R, CPI_W, FSI_R, FSI_W, IDV_R, IDV_W, MAS_R, MAS_W, PDI_R, PDI_W, UAI_R, and UAI_W. For models trained on the combined dataset, we note that the AUCs are in the 0.70-0.80 range. This demonstrates that our models can discern between suspicious and legitimate transactions. We find that our models can discern better for the corporate-related dataset with AUCs as high as 0.88. We further find evidence for predictive capacity for the country-level features for the people-related data. However, compared to the combined and corporate-related data, these results are modest with AUCs in the 0.65-0.72 range.[25]

We further note that all the models trained on datasets balanced by the hybrid-sampling technique consistently provide significant out-of-sample performance. Additionally, we find that the RF and GBM models have the best out-of-sample performance for all the three datasets balanced by the under- and hybrid-sampling techniques.

[Please insert Table 3 about here.]

5.1.2. *Determining the validity of our models using cross-validation techniques*

To determine the validity of our models we perform K-fold cross-validations. K-fold cross-validation estimates how well a model generalizes an independent dataset by dividing the dataset into K equal parts, using one part as a hold-out test set, and training the model on the remaining K-1 parts. This is then repeated K times, such that each of the K equal parts is considered for a test dataset. The out-of-sample model performance is then computed as the average of the K results.

[24]Please see Table 2 for concise definitions.

[25]We further train our models on the Hofstede country-specific culture dimensions, excluding the institution quality indices. We report the models that include only the national culture dimensions are comparable to models including the institution quality indices as well. Please see Tables B1-B3 of the Internet Appendix B.These may reflect that national culture and national governance are endogenously related. However, our objective is to determine whether national culture is an effective predictor. We do not aim to identify a causal relationship between national culture traits and malfeasance in banks.

We perform 5- and 10-fold cross-validations, and for each of the K instances, we train our models on 80% and 90% of the datasets, balanced by the hybrid-sampling method, respectively.[26] We then estimate the out-of-sample performance on the remaining 20% and 10% of the datasets. In Table 4 we report the AUC metric, estimated by the cross-validation technique, to measure performance for all the models. The results demonstrate that the predictive capacity of country-level variables remain similar to that reported in Table 3. The low standard deviation (σ) further attests to the reliability of our models.

[Please insert Table 4 about here.]

5.1.3. Investigating the relative importance of country-level features in detecting money-laundering

Table 5 presents the relative importance of our country-level features for the models trained on the three datasets.[27] We find that for both corporate-related and combined alerts, the individuality rating of both the customer's residence country (IDV_R) and country of wire origination/destination (IDV_W) are of paramount importance. This is followed by the corruption perception score of the country of wire origination/destination (CPI_W) and the customer's residence country (CPI_R) for the corporate-related alerts; and (CPI_W) and the financial secrecy score of the customer's resident country (FSI_R) for the combined alerts. For people-related alerts, the corruption perception score for the country of wire origination/destination (CPI_W) and the financial secrecy score of the resident country (FSI_R) are the two most important features, followed by the CPI_R and IDV_R.

[Please insert Table 5 about here.]

5.2. Can we improve the predictive capacity of our models by enlarging the feature space?

In this section, we extend our feature space to include account- and transaction-level variables. We further include the proprietorial risk score (PROP Score) in our enlarged feature space to assess the predictive capacity of our models.[28]

[26]We train our models on the three datasets balanced by the hybrid-sampling method since this method results in models with high predictive performance across all the three datasets.
[27]We do not report feature importance results for the SVM model since there does not exist a reliable model-specific feature importance method for SVM algorithm.
[28]All features are defined in Table 2.

5.2.1. *Predictive capacity of country-, account-, and transaction-level features in detecting money-laundering*

We further extend our feature space to include customers' account- and transaction-level information to determine whether we could improve the predictive capacity of our models. This extends our feature space to include 24 features with 12 country-level features, 4 account-level features, and 8 transaction-level features.[29]

Table 6 presents the TPR, FPR, and AUC scores for models trained on the enlarged feature space. These models enhance the predictive capacity across all the models reported in Table 3, with AUCs ranging between 0.72-0.91, 0.83-0.94, and 0.60-0.85, on the combined, corporate- and people-related datasets, respectively. We further note that the models trained on the datasets balanced by the hybrid technique are better able to discern between a fraudulent and non-fraudulent transaction with AUC scores between 0.75-0.91, 0.85-0.94, and 0.71-0.85 for the combined, corporate-, and people-related datasets, respectively. We report a significant increase in the predictive capacity of our models across all the three datasets. We again find that the RF and GBM models with under- and hybrid-sampling are the optimal models.

[Please insert Table 6 about here.]

5.2.2. *Predictive capacity of country-, account-, and transaction-level features along with the proprietorial risk score in detecting money-laundering*

Finally, we include the proprietorial risk score, PROP Score, to our enlarged feature space to determine whether its inclusion markedly enhances the predictive capacity of the models reported in Table 6.[30]

We report the out-of-sample performance of these models in Table 7. Interestingly, we find only a slight improvement, of approximately 1-2% on average, in performance. This indicates

[29] We include an additional feature, Customer Age, in the people-related models which extends the feature set to include 25 predictors.

[30] It is reasonable to surmise that the financial institution's proprietary algorithm employs data to which we do not have access.

that models with the country-, account- and transaction-level information provide useful predictive power.

[Please insert Table 7 about here.]

5.3. Does national culture traits remain useful in the extended dataset?

In this section, we investigate whether the country-specific culture and institution quality indices pertaining to customer's residence country and the country of wire origination/destination remain useful in detecting money-laundering in the enlarged feature space.

5.3.1. Does national culture traits remain useful in comparison with account-level and transaction-level variables?

We estimate feature importance for models reported in Table 6 to determine whether country-level features of the customers provide useful predictive capacity in detecting fraudulent wire transactions in the enlarged feature space. We present these results in Table 8. We note that for corporate-related alerts, the county-level features that rank among the top five features are the individuality rating of the customer's country of residence (IDV_R), individuality rating of the country of wire origination/destination (IDV_W), and the uncertainty avoidance cultural trait of the customer's residence country (UAI_R). We further find that the power-distance index score of the customer's residence country (PDI_R) informs the customer's predilections for committing financial misconduct. For people-related alerts, the individuality score of the customer's residence country (IDV_R), corruption perception score of the country of wire origination/destination (CPI_W), and financial secrecy score of the customer's country of residence (FSI_R) are the most important county-level features that rank among the top ten features. These results provide evidence of the usefulness of culture traits of customers for detecting both corporate and individual malfeasance. However, the country-level features are more pronounced in detecting corporate malfeasance than individual malfeasance. For the combined alerts, we note that IDV_R, IDV_W, and FSI_R rank among the top ten features. This further provides evidence of the usefulness of the country-level features in detecting malfeasance.

[Please insert Table 8 about here.]

5.3.2. Does national culture traits remain useful in comparison with a proprietorial risk score along with account- and transaction-level features?

Table 9 reports the feature importance for models reported in Table 7. For corporate-related alerts, we again find that the individuality scores of both the country of the wire origination/destination (IDV_W) and customer's resident country (IDV_R) are important country-level features. These features also rank among the top five features influencing a customer's predilections for committing financial misconduct. We further note that the corruption perception score of the country of wire origination/destination (CPI_W) and power-distance index score of the customer's residence country (PDI_R) are among the top ten features. Interestingly, we find that IDV_W, IDV_R, and CPI_W have higher predictive capacity than the proprietorial risk score. However, in case of people-related alerts, the PROP Score is the most important feature. This suggests that the proprietary algorithm, used by the financial institution, is more effective in detecting fraudulent transactions pertaining to individual accounts than for corporate accounts. Further, in case of people-related alerts, the financial secrecy score of the customer's residence country (FSI_R), corruption perception score of the customer's residence country (CPI_R), and corruption perception score of the country of wire origination/destination (CPI_W) rank among the top ten features in detecting money-laundering in our models. For the combined alerts, the features that influence the models in decreasing order are IDV_W, *PROP Score*, IDV_R, and FSI_R. These features also rank among the top ten features. In addition to results reported in Table 8, these results further underline the usefulness of adopting country-specific features to complement current account and transaction variables for AML monitoring.

[Please insert Table 9 about here.]

6. Discussion and Ethical Framework

The potential for AI applications' unethical repercussions, especially those that impact people's well-being is immense. Examples include recruitment, promotion, flight risk, and cessation of employment algorithms and credit extension, insurance risk-scoring, and dynamic pricing algorithms, among others. Fraud detection, mediated through machine learning, arguably falls on the lower end of the spectrum of potentially unethical AI, given its goal of mitigating financial malfeasance. Nevertheless, it is critically important to consider the ethical implications of factoring in nationality as a prompt for scrutinizing individuals. The inclusion of national culture traits in fraud detection models is a sensitive issue, since it can potentially compromise an individual's right to privacy. However, if its inclusion in the model can help financial institutions to detect and prevent money laundering, then it can be ethically justifiable. Countering money laundering is a global concern. In facilitating the generation and disbursement of illicit proceeds from criminal activities, money laundering paves the way for further financial illegal activity, compounding the problem. The upshot of money laundering is hence the perpetuation of associated crime, the misallocation of capital and the possibility of international financial instability. This paper aims to provide an empirical estimation of the impact of national culture in a money laundering model. Our results suggest that the inclusion of the national culture individuality trait significantly improves the predictive capacity of our models to detect money laundering, even relative to country-, account-, and transaction-level features along with the financial institution's proprietorial money laundering risk score.

We acknowledge the obvious tension between the right to privacy and the objective to counter money laundering. We, therefore, present below our arguments justifying the use of national culture traits in models to detect and prevent money laundering.

6.1. Do public good concerns to counter money laundering offset 'collective treatment' in algorithmic national profiling?

Alter and Darley (2009) define collective treatment as 'the act of behaving toward more than one individual uniformly.' Contrastingly, in individualized treatment, individuals are treated differently based on certain criteria. An example of collective treatment is punishing a group for being offenders as opposed to prosecuting its individual members commensurate with their level of crime. According to Alter and Darley (2009), collective treatment is predicated on a group's shared salient features (such as race and ethnicity, among others) that are used to stereotype them. Thus, people belonging to the same group are treated as interchangeable members of the group. As noted by Brewer and Harasty (1996), Campbell (1958), Dasgupta et al. (1999) among others, such salient features can include race, ethnicity, socio-economic status, religion, physical appearance, nationality, and even debility. Clearly, national culture can also be featured to characterize and homogenize a people, especially in alert models that excavate cultural factors to detect fraud. However, the chief danger of collective treatment lies in the prospect of individuals in positions of authority administering it to reward, punish, or restrict the rights of a group within a population. For instance, a judge who sentences a criminal gang rather than its individuals etc. One advantage of machine learning-based alert models is in sidestepping arbitrary individual choices to impose or not to impose collective treatment.

6.2. Are those issuing the alerts permitted to use the personal data?

The legality of gathering and mining certain data does not in itself make it ethical. After all, 'Ethics' constitutes a set of moral codes beyond legally stipulated minimums. Apropos mining data inhering in machine learning procedures, questions of ethicality invariably arise. Whether the institution conducting machine learning is allowed to use the data in its algorithms becomes an important legal and ethical issue. While there may be legal restrictions to using particular data, ethical issues too loom over issues of legality. In many cases, machine learning may eventually employ data without proper permissions. As noted by Adomavicius and Tuzhilin (2001), data mining in the context of individuals is viewed as either 'factual' viz., who the customer or 'transactional,' what the customer has done or is doing (see also Cook (2008)). Adomavicius and Tuzhilin (2001) suggest the latter is more commonly used for identifying a criminal as well as more commonly contested for intruding on individual privacy. However, the money laundering alert model of this paper suggests that simply using data about who the customer is, i.e., the customer's home country, can generate area-under-the-curve predictions that are almost 90 percent successful. So, using national culture as a predictor brings the potential advantage of circumventing intrusive gathering of customer behavior. Use of national culture avoids issues of employing personal data. The sweeping aggregate generality of national culture helps avoid invasive use, more often than not without permission, of individual characteristics. Thus, in respect of including national culture in machine-learning models, a relevant question arises, if not national factors, then what other factors? And what would be the alternative set of implications? Overall, as regards permissions to use data, national culture, while arguably a rough profiling of people from respective nations, avoids the use of more individual and likely personal data. In effect, an ethical minefield opens up in the interstices of national culture profiling's sweeping generalities, and the potential invasion of privacy in the collection of personal, often transactional, data.

7. Conclusion

High-profile scandals at Danske and Swedbank call for an urgent need to innovate in AML protocols. A UN study estimates that law enforcement intercepted less than 1 percent ($40 billion) of the approximately $4tn of illicit funds in circulation, across the financial system (UN, 2011). As money laundering enables the perpetuation of crime (e.g. human and drug trafficking) and is associated with the mis-allocation of capital and financial instability, the enhancement of AML systems is of paramount human rights, societal and capital market importance.

Our paper examines the efficacity of incorporating national culture profiling, at a globally important financial institution, in machine-learning informed alert models to detect money-laundering. The models specify the national cultural measure of individualism, and they also account for other national culture measures such as masculinity, power distance, and uncertainty avoidance, as well as variables relating to levels of corruption and financial secrecy in a country. Our models, further comprise industry standard account- and transaction-level variables. The models report, across samples of corporate and individual customers, high out-of-sample predictive performances of detecting money laundering incidents, before analysts do so.

We estimate the relative importance of the predictors in the most successful models involving corporate accounts. We discover that individualism scores for the customer's resident country and of the country of origin / destination of the wire transfer, are the most important of the country-level variables and, indeed, are among the most important variables specified in the models. As regards personal account models, the national culture individualism trait of where a customer is resident is highly important in contributing to the models predictions. Our study, hence, suggests that a banking customers' socio-cultural matrix informs their predilections for committing financial malfeasance, namely, money-laundering. Our findings indicate that the inclusion of the national culture individualism trait considerably contributes to the predictive capacity of our models to detect money laundering, even relative to country-, account-, and transaction-level features, together with the proprietorial risk score. This suggests compelling evidence that culture profiling can prove efficacious and thus incentivated in AML compliance.

If the inclusion of national culture in a fraud detection model materially improves the performance of that model and alleviates the suffering that can ensue due to money laundering, then its inclusion is ethically defensible. In this paper, we abstract from this ethical upshot. Enshrined in law is the fundamental right to privacy of individuals, an opposing ethical argument. In an AML compliance context, there is inadequate guidance regarding the nature of requisite safeguards for the protection of data privacy. Our paper serves to show the potential efficacity of including national culture profiling in AML, despite data privacy concerns. Our work hence underscores the importance of guidance regarding how AML systems should protect an individual's fundamental right to privacy, and, critically, the degree to which this guidance is enforced.

References

Adomavicius, G. and A. Tuzhilin (2001). Using data mining methods to build customer profiles. *Computer 34*(2), 74–82.

Aggarwal, R., J. W. Goodell, and L. J. Selleck (2015). Lending to women in microfinance: Role of social trust. *International Business Review 24*(1), 55–65.

Alter, A. L. and J. M. Darley (2009). When the association between appearance and outcome contaminates social judgment: A bidirectional model linking group homogeneity and collective treatment. *Journal of Personality and Social Psychology 97*(5), 776.

Armstrong, R. W. (1996). The relationship between culture and perception of ethical problems in international marketing. *Journal of Business Ethics 15*(11), 1199–1208.

Bame-Aldred, C. W., J. B. Cullen, K. D. Martin, and K. P. Parboteeah (2013). National culture and firm-level tax evasion. *Journal of Business Research 66*(3), 390–396.

Batista, G. E., R. C. Prati, and M. C. Monard (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD explorations newsletter 6*(1), 20–29.

Biais, B., D. Hilton, K. Mazurier, and S. Pouget (2005). Judgemental overconfidence, self-monitoring, and trading performance in an experimental financial market. *The Review of economic studies 72*(2), 287–312.

Brewer, M. B. and A. S. Harasty (1996). Seeing groups as entities: The role of perceiver motivation.

Campbell, D. T. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science 3*(1), 14.

Chen, C. C., M. W. Peng, and P. A. Saparito (2002). Individualism, collectivism, and opportunism: A cultural perspective on transaction cost economics. *Journal of Management 28*(4), 567–583.

Chui, A. C., A. E. Lloyd, and C. C. Kwok (2002). The determination of capital structure: is national culture a missing piece to the puzzle? *Journal of International Business Studies 33*(1), 99–127.

Chui, A. C., S. Titman, and K. J. Wei (2010). Individualism and momentum around the world. *The Journal of Finance 65*(1), 361–392.

Conlon, T., X. Huan, and C. B. Muckley (2024). Does national culture influence malfeasance in banks around the world? *Journal of International Financial Markets, Institutions and Money 90*, 101888.

Cook, J. (2008). Ethics of data mining. In *Information Security and Ethics: Concepts, Methodologies, Tools, and Applications*, pp. 211–217. IGI Global.

Cullen, J. B., K. P. Parboteeah, and M. Hoegl (2004). Cross-national differences in managers' willingness to justify ethically suspect behaviors: A test of institutional anomie theory. *Academy of Management Journal 47*(3), 411–421.

Cumming, D., A. Groh, and S. Johan (2018). Same rules, different enforcement: Market abuse in europe. *Journal of International Financial Markets, Institutions & Money 54*, 130–151.

Dasgupta, N., M. R. Banaji, and R. P. Abelson (1999). Group entitativity and group perception: Associations between physical features and psychological judgment. *Journal of Personality and Social Psychology 77*(5), 991.

DeBacker, J., B. T. Heim, and A. Tran (2015). Importing corruption culture from overseas: Evidence from corporate tax evasion in the united states. *Journal of Financial Economics 117*(1), 122–138.

FATF (2021). Opportunities and challenges of new technologies for aml/cft. *FATF, Paris, France*.

FATF and Egmont Group (2020). Trade-based money laundering: Risk indicators. *FATF, Paris, France*.

Fidrmuc, J. and M. Jacob (2010). Culture, agency costs, and dividends. *Journal of Comparative Economics 38*(3), 321–339.

Financial Stability Board (2018). Strengthening governance frameworks to mitigate misconduct risk: A toolkit for firms and supervisors.

Fischer, R. and A. Chalmers (2008). Is optimism universal? a meta-analytical investigation of optimism levels across 22 nations. *Personality and Individual Differences 45*(5), 378–382.

Fischer, R., C.-M. Vauclair, J. R. Fontaine, and S. H. Schwartz (2010). Are individual-level and country-level value structures different? testing hofstede's legacy with the schwartz value survey. *Journal of cross-cultural psychology 41*(2), 135–151.

Fisman, R. and E. Miguel (2007). Corruption, norms, and legal enforcement: Evidence from diplomatic parking tickets. *Journal of Political economy 115*(6), 1020–1048.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.

Gaganis, C., I. Hasan, P. Papadimitri, and M. Tasiou (2019). National culture and risk-taking: Evidence from the insurance industry. *Journal of Business Research 97*, 104–116.

Goodell, J. W. (2019). Comparing normative institutionalism with intended rationality in cultural-finance research. *International Review of Financial Analysis 62*, 124–134.

Hassan, O. A. and G. Giorgioni (2015). Analyst coverage, corruption and financial secrecy: a multi-country study. *Corruption and Financial Secrecy: A Multi-Country Study (February 18, 2015)*.

Heine, S. J., D. R. Lehman, H. R. Markus, and S. Kitayama (1999). Is there a universal need for positive self-regard? *Psychological review 106*(4), 766.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage publications.

Houqe, N., R. M. Monem, M. Tareq, and T. van Zijl (2015). Secrecy and mandatory ifrs adoption on earnings quality.

IMF, I. M. F. (2021). The imf and the fight against money laundering and the financing of terrorism. *Policy Brief July 2021*.

Kim, J.-B., Z. Wang, and L. Zhang (2016). Ceo overconfidence and stock price crash risk. *Contemporary Accounting Research 33*(4), 1720–1749.

Kirkman, B. L., K. B. Lowe, and C. B. Gibson (2006). A quarter century of culture's consequences: A review of empirical research incorporating hofstede's cultural values framework. *Journal of International Business Studies 37*(3), 285–320.

Kreiser, P. M., L. D. Marino, P. Dickson, and K. M. Weaver (2010). Cultural influences on entrepreneurial orientation: The impact of national culture on risk taking and proactiveness in smes. *Entrepreneurship theory and practice 34*(5), 959–984.

Kwok, C. and S. Tadesse (2006). National culture and financial systems. *Journal of International Business Studies 37*, 227–247.

Leung, K. and M. W. Morris (2015). Values, schemas, and norms in the culture–behavior nexus: A situated dynamics framework. *Journal of International Business Studies 46*(9), 1028–1050.

Li, K., D. Griffin, H. Yue, and L. Zhao (2013). How does culture influence corporate risk-taking? *Journal of corporate finance 23*, 1–22.

Lievenbrück, M. and T. Schmid (2014). Why do firms (not) hedge?—novel evidence on cultural influence. *Journal of Corporate Finance 25*, 92–106.

Liu, X. (2016a). Corruption culture and corporate misconduct. *Journal of Financial Economics 122*(2), 307–327.

Liu, X. (2016b). Corruption culture and corporate misconduct. *Journal of Financial Economics 122*(2), 307–327.

Markus, H. R. and S. Kitayama (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review 98*(2), 224.

Martin, K. D., J. B. Cullen, J. L. Johnson, and K. P. Parboteeah (2007). Deciding to bribe: A cross-level analysis of firm and home country influences on bribery activity. *Academy of Management Journal 50*(6), 1401–1422.

Michalos, A. C. and P. M. Hatch (2019). Good societies, financial inequality and secrecy, and a good life: from aristotle to piketty. *Applied Research in Quality of Life*, 1–50.

Mourouzidou-Damtsa, S., A. Milidonis, and K. Stathopoulos (2019). National culture and bank risk-taking. *Journal of Financial Stability 40*, 132–143.

Orlova, S., R. Rao, and T. Kang (2017). National culture and the valuation of cash holdings. *Journal of Business Finance & Accounting 44*(1-2), 236–270.

Peterson, M. F. and T. S. Barreto (2018). Interpreting societal culture value dimensions. *Journal of International Business Studies 49*(9), 1190–1207.

Puspitasari, E., C. Sukmadilaga, H. Suciati, R. F. Bahar, and E. K. Ghani (2019). The effect of financial secrecy and ifrs adoption on earnings quality: A comparative study between indonesia, malaysia and singapore. *International Journal of Innovation, Creativity and Change 5*(2), 863–886.

Richardson, G. (2008). The relationship between culture and tax evasion across countries: Additional evidence and extensions. *Journal of International Accounting, Auditing and Taxation 17*(2), 67–78.

Schuler, R. S. and N. Rogovsky (1998). Understanding compensation practice variations across firms: The impact of national culture. *Journal of International business studies 29*(1), 159–177.

Shupp, R. S. and A. W. Williams (2008). Risk preference differentials of small groups and individuals. *The Economic Journal 118*(525), 258–283.

Tsakumis, G. T., A. P. Curatola, and T. M. Porcano (2007). The relation between national cultural dimensions and tax evasion. *Journal of international accounting, auditing and taxation 16*(2), 131–147.

UN, U. N. (2011). Estimating illicit financial flows resulting from drug trafficking and other transnational organized crimes.

Van den Steen, E. (2004). Rational overoptimism (and other biases). *American Economic Review 94*(4), 1141–1151.

Vitell, S. J., S. L. Nwachukwu, and J. H. Barnes (1993). The effects of culture on ethical decision-making: An application of hofstede's typology. *Journal of Business Ethics 12*(10), 753–760.

Volkema, R. J. (2004). Demographic, cultural, and economic predictors of perceived ethicality of negotiation behavior: A nine-country analysis. *Journal of Business Research 57*(1), 69–78.

Watts, L. L., L. M. Steele, and D. N. Den Hartog (2020). Uncertainty avoidance moderates the relationship between transformational leadership and innovation: A meta-analysis. *Journal of International Business Studies 51*(1), 138–145.

Figure 1: Confusion Matrix

		Actual Class	
		positive	negative
Predicted Class	positive	# **T**rue **P**ositives	# **F**alse **P**ositives
	negative	# **F**alse **N**egatives	# **T**rue **N**egatives

Figure 2: ROC Curve

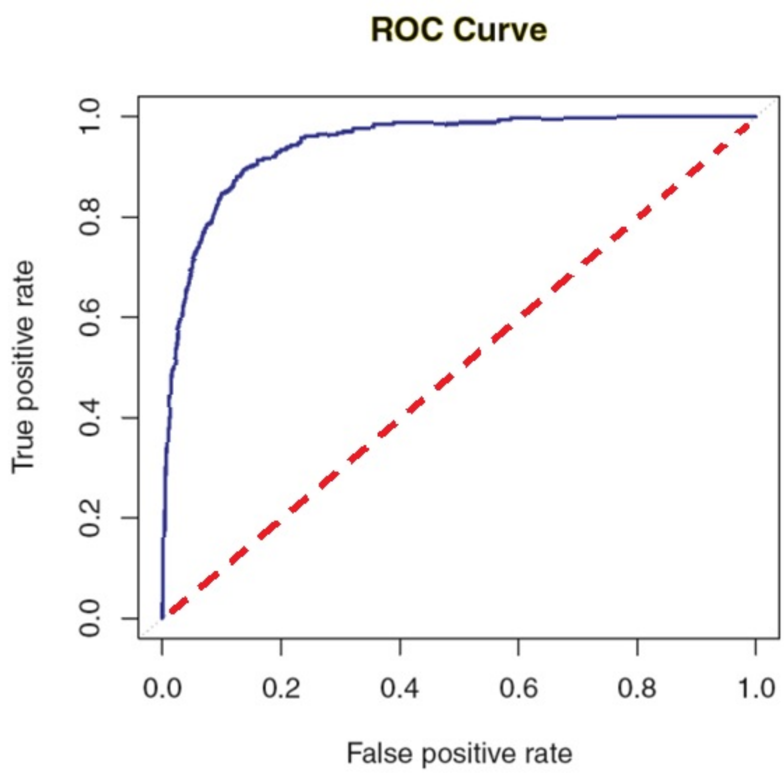

Table 1: Data Cross-section and Sample Selection

Panel A: Alerts and Issue Cases by Year

	Combined		Corporate		People	
Year	**#Alerts**	**#Issues**	**#Alerts**	**#Issues**	**#Alerts**	**#Issues**
2009	22,183	878	5,752	448	16,431	430
2010	23,154	485	6,643	215	16,511	270
2011	20,335	216	6,193	68	14,142	148
2012	18,572	143	5,298	29	13,274	114
2013	21,088	205	5,984	71	15,104	134
2014	11,098	87	2,617	41	8,481	46
2015	11,468	34	2,937	7	8,531	27
2016	11,779	71	2,841	14	8,938	57
2017	9,885	76	2,771	19	7,114	57
2018	4,355	11	1,236	2	3,119	9
Total	**153,917**	**2,206**	**42,272**	**914**	**111,645**	**1,292**

Panel B: Sample Selection

	Combined		Corporate		People	
Selection Criteria	**#Alerts**	**#Issues**	**#Alerts**	**#Issues**	**#Alerts**	**#Issues**
All Alerts	206,751	2,440	42,272	914	111,645	1,292
Corp & Ppl Accounts	153,917	2,206	42,272	914	111,645	1,292
Country-level Variables	74,832	1,183	30,303	524	44,529	659
Account/Transaction-level Variables	74,246	1,172	30,292	524	43,954	648

Notes: The table reports the cross-section of our data (Panel A) and the sample selection (Panel B). An alert is raised when a customer's wire activity raises certain flags and an Issue case indicates that the subsequent investigation has deemed the activity to be highly suspicious. The sample selection shows the number of alerts available our data set according to each criterion, applied in sequence. A more detailed description of our variables is available in Table 2.

Table 2: Predictor Details

Predictor Abbreviation	Details
COUNTRY-LEVEL	
Corruption Perception Index	Score of customer's residence country (CPI_R) and country of origin/destination of wire (CPI_W) according to Transparency International's Corruption Perception Index.
Financial Secrecy Index	Score of customer's residence country (FSI_R) and country of origin/destination of wire (FSI_W) according to Transparency International's Financial Secrecy Index.
Individualism Index	Score of customer's residence country (IDV_R) and country of origin/destination of wire (IDV_W) based on Hofstede's "Individualism" dimension of culture.
Masculinity Index	Score of customer's residence country (MAS_R) and country of origin/destination of wire (MAS_W) based on Hofstede's "Masculinity" dimension of culture.
Power-Distance Index	Score of customer's residence country (PDI_R) and country of origin/destination of wire (PDI_W) based on Hofstede's "Power-Distance" dimension of culture.
Uncertainty Avoidance Index	Score of customer's residence country (UAI_R) and country of origin/destination of wire (UAI_W) based on Hofstede's "Uncertainty Avoidance" dimension of culture.
ACCOUNT-LEVEL	
Customer Age	Age of customer associated with alert, at time of alert (CUS_AGE).
Account Age	Age of account associated with alert, at time of alert (ACC_AGE).
Customer Net Worth	Net Worth of customer associated with alert (NET_WRTH).
Alert Supplier Code	Code denoting source of alert, whether alert is generated by Business or Retail transactions (SUPP_CO).
TRANSACTION-LEVEL	
Amount Transfers In	Aggregate amount of incoming wire and electronic transfers over 180 days before alert (ΣTFI_{180}).
No. Transfers In	Number of incoming wire and electronic transfers over 180 days before alert ($\#TFI_{180}$).
Amount Transfers Out	Aggregate amount of outgoing wire and electronic transfers over 180 days before alert (ΣTFO_{180}).
No. Transfers Out	Number of outgoing wire and electronic transfers over 180 days before alert ($\#TFO_{180}$).
Amount Checks In	Aggregate amount of incoming checks over 180 days before alert (ΣCKI_{180}).
No. Checks In	Number of incoming checks over 180 days before alert ($\#CKI_{180}$).
Amount Checks Out	Aggregate amount of outgoing checks over 180 days before alert ((ΣCKO_{180})).
No. Checks Out	Number of outgoing checks over 180 days before alert ($\#CKO_{180}$).
Proprietary	
PROP Score	Risk score based on proprietary alert algorithm of financial institution.

Notes: The table reports the complete set of predictors used in our models along with their definitions and abbreviations for reference. The "Wire" variables refer only to the wire transactions on the day of an alert whereas the "Transfer" and "Check" variables refer to all relevant transactions appearing on accounts associated with an alert in the 180 day period preceding that alert.

Table 3: Country-level Models

		Combined			Corporate			People		
Model	**Balancing**	**TPR**	**FPR**	**AUC**	**TPR**	**FPR**	**AUC**	**TPR**	**FPR**	**AUC**
LR	**No Balancing**	0.70	0.43	0.722	0.89	0.50	0.845	0.59	0.42	0.670
	Under-sampling	0.76	0.49	0.727	0.90	0.51	0.850	0.61	0.43	0.664
	Hybrid-sampling	0.76	0.50	0.726	0.90	0.47	0.851	0.58	0.40	0.670
	Synthetic-sampling	0.71	0.43	0.723	0.92	0.53	0.861	0.60	0.41	0.659
RF	**No Balancing**	1.00	1.00	0.543	1.00	1.00	0.674	1.00	1.00	0.505
	Under-sampling	0.71	0.40	0.741	0.89	0.41	0.875	0.66	0.41	0.702
	Hybrid-sampling	0.65	0.31	0.726	0.88	0.34	0.878	0.66	0.41	0.695
	Synthetic-sampling	1.00	1.00	0.696	1.00	1.00	0.859	1.00	1.00	0.641
SVM	**No Balancing**	0.53	0.47	0.521	0.42	0.42	0.504	0.62	0.56	0.516
	Under-sampling	0.78	0.60	0.704	0.88	0.59	0.805	0.66	0.44	0.660
	Hybrid-sampling	0.68	0.50	0.662	0.88	0.59	0.807	0.68	0.47	0.636
	Synthetic-sampling	0.77	0.51	0.645	0.89	0.60	0.816	0.59	0.41	0.610
GBM	**No Balancing**	0.87	0.60	0.768	0.91	0.49	0.878	0.84	0.56	0.719
	Under-sampling	0.87	0.59	0.770	0.85	0.41	0.870	0.83	0.57	0.708
	Hybrid-sampling	0.74	0.40	0.771	0.90	0.43	0.881	0.83	0.55	0.716
	Synthetic-sampling	0.68	0.41	0.724	0.94	0.60	0.868	0.73	0.57	0.660

Notes: The table reports the performance of our Country-level model using logistic regression (LR), random forest (RF), support vector machine (SVM) and gradient boosting (GBM) in combination with no balancing, under-sampling, hybrid-sampling and synthetic-sampling, respectively. The performance is measured using True Positive Rate (TP Rate), False Positive Rate (FP Rate) and Area under the ROC Curve (AUC). The data sample comprises of 74,724 alerts (30,292 corporate-related and 43,954 people-related) with 1,183 Issue cases (524 corporate-related and 648 people-related). The model has 12 predictors.

Table 4: Cross-validation for Country-level Models with Hybrid-sampling.

Panel A: 5-Fold Cross-validation on AUC scores												
	Combined				**Corporate**				**People**			
Round	**LR**	**RF**	**SVM**	**GBM**	**LR**	**RF**	**SVM**	**GBM**	**LR**	**RF**	**SVM**	**GBM**
1	0.717	0.722	0.656	0.765	0.758	0.774	0.785	0.813	0.658	0.662	0.594	0.683
2	0.726	0.726	0.705	0.777	0.856	0.862	0.811	0.896	0.674	0.706	0.663	0.724
3	0.737	0.729	0.705	0.766	0.861	0.873	0.829	0.902	0.671	0.675	0.598	0.709
4	0.743	0.739	0.678	0.789	0.852	0.872	0.842	0.887	0.660	0.681	0.606	0.723
5	0.762	0.768	0.725	0.797	0.820	0.848	0.833	0.874	0.677	0.733	0.688	0.746
μ	0.737	0.737	0.694	0.779	0.829	0.846	0.820	0.874	0.668	0.691	0.630	0.717
σ	0.017	0.019	0.027	0.014	0.043	0.041	0.023	0.036	0.009	0.028	0.043	0.023
Panel B: 10-Fold Cross-validation on AUC scores												
	Combined				**Corporate**				**People**			
Round	**LR**	**RF**	**SVM**	**GBM**	**LR**	**RF**	**SVM**	**GBM**	**LR**	**RF**	**SVM**	**GBM**
1	0.769	0.757	0.721	0.799	0.870	0.902	0.865	0.915	0.655	0.692	0.571	0.708
2	0.777	0.770	0.760	0.814	0.798	0.811	0.780	0.864	0.664	0.697	0.570	0.718
3	0.719	0.722	0.674	0.761	0.823	0.818	0.804	0.870	0.651	0.677	0.595	0.702
4	0.746	0.754	0.720	0.812	0.820	0.824	0.823	0.844	0.689	0.672	0.622	0.685
5	0.710	0.693	0.692	0.762	0.810	0.819	0.810	0.867	0.670	0.657	0.570	0.746
6	0.754	0.726	0.714	0.763	0.870	0.867	0.818	0.907	0.686	0.761	0.642	0.774
7	0.743	0.736	0.696	0.769	0.866	0.888	0.847	0.917	0.691	0.735	0.683	0.766
8	0.726	0.740	0.713	0.786	0.807	0.850	0.819	0.869	0.653	0.651	0.550	0.669
9	0.724	0.736	0.738	0.769	0.847	0.865	0.804	0.877	0.648	0.660	0.572	0.670
10	0.729	0.728	0.702	0.775	0.807	0.854	0.788	0.894	0.700	0.738	0.663	0.750
μ	0.740	0.736	0.713	0.781	0.832	0.850	0.816	0.882	0.671	0.694	0.604	0.719
σ	0.022	0.021	0.024	0.021	0.029	0.031	0.025	0.025	0.019	0.038	0.046	0.039

Notes: The table reports the AUCs for 5-fold and 10-fold cross-validation for the hybrid-sampled Country-level model with logistic regression (LR), random forest (RF), support vector machine (SVM) and gradient boosting (GBM). The data sample comprises of 82,964 alerts (30,303 corporate-related and 44,529 people-related) with 1,240 Issue cases (524 corporate-related and 659 people-related). The model has 12 predictors.

Table 5: Country-level Predictor Importance for Country-level Model with Hybrid-sampling

	Combined				Corporate				People			
Predictor	**LR**	**RF**	**GBM**	**Ave.**	**LR**	**RF**	**GBM**	**Ave.**	**LR**	**RF**	**GBM**	**Ave.**
$\mathbf{CPI_R}$	*	5	5	5	***	5	4	4	***	3	3	3
$\mathbf{FSI_R}$	***	6	3	4	·	8	6	8	***	1	2	2
$\mathbf{IDV_R}$	***	1	1	1	***	1	1	1	***	2	4	4
$\mathbf{MAS_R}$	***	9	6	8		9	11	9	***	9	8	8
$\mathbf{PDI_R}$	***	4	7	6	***	4	7	5	***	6	6	5
$\mathbf{UAI_R}$	***	8	10	9	***	7	5	6	***	8	7	7
$\mathbf{CPI_W}$	***	3	4	3	***	3	3	3	***	4	1	1
$\mathbf{FSI_W}$		12	8	11	***	10	12	12	*	12	11	12
$\mathbf{IDV_W}$		2	2	2	***	2	2	2	***	7	12	11
$\mathbf{MAS_W}$	***	10	11	10		11	10	10		5	10	9
$\mathbf{PDI_W}$	***	7	9	7	*	6	8	7	***	11	9	10
$\mathbf{UAI_W}$	***	11	12	12	***	12	9	11	***	10	5	6

Notes: The table reports the importance of the Country-level predictors by ranking for the Hybrid-sampled Country-level model applied to the full sample (combined) and its partitions (Corporate & People accounts). Estimates of importance are obtained from the logistic regression (LR), random forest (RF), gradient boosted model (GBM) algorithms. A weighted average of RF and GBM (Ave.) is included. For LR, ***,**,* and · denote 0.1%, 1%, 5% and 10% levels of significance. RF and GBM are both tree-based algorithms and so their estimates are based on the mean decrease in the Gini index of each node across all trees. The Gini index measures node impurity. The data sample comprises of 82,964 alerts (30,303 corporate-related and 44,529 people-related) with 1,240 Issue cases (524 corporate-related and 659 people-related). The model has 12 predictors.

Table 6: Country, Account & Transaction-level Models

		Combined			Corporate			People		
Model	**Balancing**	**TPR**	**FPR**	**AUC**	**TPR**	**FPR**	**AUC**	**TPR**	**FPR**	**AUC**
LR	**No Balancing**	0.72	0.44	0.740	0.84	0.41	0.836	0.73	0.46	0.714
	Under-sampling	0.73	0.41	0.747	0.90	0.46	0.849	0.74	0.48	0.706
	Hybrid-sampling	0.72	0.42	0.747	0.90	0.54	0.846	0.69	0.43	0.712
	Synthetic-sampling	0.71	0.42	0.740	0.87	0.52	0.831	0.74	0.49	0.685
RF	**No Balancing**	0.96	0.53	0.908	0.92	0.28	0.930	1.00	1.00	0.842
	Under-sampling	0.93	0.48	0.895	0.97	0.55	0.932	0.91	0.59	0.835
	Hybrid-sampling	0.94	0.47	0.911	0.96	0.41	0.938	0.88	0.49	0.848
	Synthetic-sampling	1.00	1.00	0.772	0.89	0.42	0.877	1.00	1.00	0.638
SVM	**No Balancing**	0.88	0.51	0.835	0.91	0.59	0.881	0.73	0.48	0.737
	Under-sampling	0.89	0.51	0.801	0.95	0.54	0.880	0.79	0.57	0.739
	Hybrid-sampling	0.86	0.43	0.845	0.90	0.53	0.886	0.72	0.53	0.722
	Synthetic-sampling	0.65	0.41	0.723	0.86	0.52	0.847	0.55	0.40	0.603
GBM	**No Balancing**	0.88	0.47	0.842	0.93	0.59	0.880	0.84	0.57	0.769
	Under-sampling	0.93	0.53	0.853	0.95	0.40	0.916	0.82	0.40	0.799
	Hybrid-sampling	0.87	0.41	0.863	0.93	0.40	0.921	0.83	0.47	0.799
	Synthetic-sampling	0.67	0.40	0.717	0.88	0.40	0.846	0.76	0.55	0.652

Notes: The table reports the performance of our Country, Account & Transaction-level model using logistic regression (LR), random forest (RF), support vector machine (SVM) and gradient boosting (GBM) in combination with no balancing, under-sampling, hybrid-sampling and synthetic-sampling, respectively. The performance is measured using True Positive Rate (TP Rate), False Positive Rate (FP Rate) and Area under the ROC Curve (AUC). The data sample comprises of 74,246 alerts (30,292 corporate-related and 43,954 people-related) with 1,182 Issue cases (524 corporate-related and 648 people-related). The model has 24 predictors.

Table 7: Country, Account & Transaction-level Models with PROP Score

		Combined			Corporate			People		
Model	**Balancing**	**TPR**	**FPR**	**AUC**	**TPR**	**FPR**	**AUC**	**TPR**	**FPR**	**AUC**
LR	**No Balancing**	0.77	0.48	0.754	0.82	0.40	0.840	0.79	0.47	0.733
	Under-sampling	0.78	0.45	0.763	0.91	0.49	0.848	0.79	0.47	0.734
	Hybrid-sampling	0.77	0.44	0.764	0.90	0.58	0.851	0.76	0.46	0.733
	Synthetic-sampling	0.79	0.47	0.756	0.86	0.47	0.834	0.83	0.56	0.723
RF	**No Balancing**	0.89	0.40	0.894	0.95	0.33	0.946	1.00	1.00	0.845
	Under-sampling	0.87	0.43	0.878	0.95	0.40	0.943	0.91	0.57	0.846
	Hybrid-sampling	0.92	0.44	0.901	0.97	0.44	0.952	0.83	0.41	0.855
	Synthetic-sampling	0.90	0.58	0.790	0.88	0.43	0.873	1.00	1.00	0.690
SVM	**No Balancing**	0.87	0.50	0.828	0.89	0.54	0.883	0.78	0.59	0.742
	Under-sampling	0.88	0.57	0.789	0.94	0.51	0.896	0.81	0.57	0.753
	Hybrid-sampling	0.88	0.53	0.833	0.91	0.57	0.896	0.75	0.54	0.731
	Synthetic-sampling	0.78	0.56	0.745	0.83	0.41	0.848	0.65	0.49	0.667
GBM	**No Balancing**	0.89	0.56	0.829	0.92	0.58	0.886	0.91	0.57	0.818
	Under-sampling	0.87	0.43	0.850	0.94	0.40	0.922	0.91	0.51	0.828
	Hybrid-sampling	0.87	0.42	0.855	0.96	0.55	0.926	0.83	0.40	0.828
	Synthetic-sampling	0.74	0.48	0.709	0.88	0.48	0.840	0.80	0.58	0.689

Notes: The table reports the performance of our Country, Account & Transaction-level model, with the PROP score variable included, using logistic regression (LR), random forest (RF), support vector machine (SVM) and gradient boosting (GBM) in combination with no balancing, under-sampling, hybrid-sampling and synthetic-sampling, respectively. The performance is measured using True Positive Rate (TP Rate), False Positive Rate (FP Rate) and Area under the ROC Curve (AUC). The data sample comprises of 74,724 alerts (30,292 corporate-related and 43,954 people-related) with 1,182 Issue cases (524 corporate-related and 648 people-related). The model has 25 predictors.

Table 8: Country-level Predictor Importance for Country, Account & Transaction-level Model with Hybrid-sampling

	Combined				Corporate				People			
Predictor	**LR**	**RF**	**GBM**	**Ave.**	**LR**	**RF**	**GBM**	**Ave.**	**LR**	**RF**	**GBM**	**Ave.**
$\mathbf{CPI_R}$		12	18	17	***	12	18	12	***	14	13	13
$\mathbf{FSI_R}$	***	8	6	7		14	12	14	***	9	9	9
$\mathbf{IDV_R}$		5	1	1	***	2	1	1	***	7	1	2
$\mathbf{MAS_R}$	***	15	11	13	***	15	19	16	***	12	17	16
$\mathbf{PDI_R}$	***	11	12	11	***	9	10	10	•	16	19	17
$\mathbf{UAI_R}$	***	13	16	15	***	5	4	4	•	15	14	15
$\mathbf{CPI_W}$	***	17	7	10	***	10	14	11	***	17	4	7
$\mathbf{FSI_W}$	***	21	17	20		16	20	20	**	23	22	24
$\mathbf{IDV_W}$	***	10	4	5		4	2	3	***	20	24	22
$\mathbf{MAS_W}$	*	18	21	19		18	17	19		18	23	21
$\mathbf{PDI_W}$	***	22	20	21		13	13	13	***	24	21	23
$\mathbf{UAI_W}$	***	23	23	23	***	21	23	21		21	15	18

Notes: The table reports the importance of the Country-level predictors by absolute ranking for the Hybrid-sampled Country, Account & Transaction-level model applied to the full sample (combined) and its partitions (Corporate & People accounts). Estimates of importance are obtained from the logistic regression (LR), random forest (RF), gradient boosted model (GBM) algorithms. A weighted average of RF and GBM (Ave.) is included. For LR, ***,**,* and • denote 0.1%, 1%, 5% and 10% levels of significance. RF and GBM are both tree-based algorithms and so their estimates are based on the mean decrease in the Gini index of each node across all trees. The Gini index measures node impurity. The data sample comprises of 74,246 alerts (30,292 corporate-related and 43,954 people-related) with 1,182 Issue cases (524 corporate-related and 648 people-related). The model has 24 predictors.

Table 9: Country-level Predictor Importance for Country, Account & Transaction-level Model with Hybrid-sampling and PROP Score included

	Combined				Corporate				People			
Predictor	**LR**	**RF**	**GBM**	**Ave.**	**LR**	**RF**	**GBM**	**Ave.**	**LR**	**RF**	**GBM**	**Ave.**
PROP	***	5	3	3	***	7	10	9	***	4	1	1
$\mathbf{CPI_R}$		12	11	12	***	14	12	13	***	10	6	8
$\mathbf{FSI_R}$	***	9	6	8		15	14	15	***	8	4	7
$\mathbf{IDV_R}$		8	4	5	***	3	2	2	***	13	12	13
$\mathbf{MAS_R}$	***	19	17	17	***	17	18	19	***	15	13	16
$\mathbf{PDI_R}$	***	14	20	18	***	11	8	10	·	14	16	15
$\mathbf{UAI_R}$	***	13	14	13	***	12	11	11	·	17	17	17
$\mathbf{CPI_W}$	***	18	10	11	***	13	6	6	***	18	5	10
$\mathbf{FSI_W}$	***	24	19	21		16	17	17	**	24	20	24
$\mathbf{IDV_W}$	***	11	1	2		2	1	1	***	21	24	21
$\mathbf{MAS_W}$	*	16	22	20		18	16	16		20	25	20
$\mathbf{PDI_W}$	***	21	21	22		8	23	14	***	25	19	25
$\mathbf{UAI_W}$	***	23	23	23	***	22	20	22		23	21	22

Notes: The table reports the importance of the Country-level predictors by absolute ranking for the Hybrid-sampled Country, Account & Transaction-level model applied to the full sample (combined) and its partitions (Corporate & People accounts) with the PROP score variable included. Estimates of importance are obtained from the logistic regression (LR), random forest (RF), gradient boosted model (GBM) algorithms. A weighted average of RF and GBM (Ave.) is included. For LR, ***,**,* and · denote 0.1%, 1%, 5% and 10% levels of significance. RF and GBM are both tree-based algorithms and so their estimates are based on the mean decrease in the Gini index of each node across all trees. The Gini index measures node impurity. The data sample comprises of 74,724 alerts (30,292 corporate-related and 43,954 people-related) with 1,182 Issue cases (524 corporate-related and 648 people-related). the model has 25 predictors.