

Aprendizado de máquina contra fraudes contábeis

Satoshi Kondo, Daisuke Miyakawa, Kengo Shiraki, Miki Suga, Teppei Usuki 13 de maio de 2020

Detectar e prevenir fraudes contábeis é uma preocupação para muitos formuladores de políticas ao redor do mundo. Esta coluna apresenta uma estrutura que incorpora técnicas de aprendizado de máquina para detectar e prever comportamento fraudulento das empresas ao relatar informações financeiras. A estrutura conta com um conjunto maior de informações firmes para obter melhor desempenho de detecção e, ao contrário das estruturas anteriores, fornece previsões para possíveis fraudes contábeis futuras.



18

que deve ser evitado do ponto de vista prático e devido a possíveis danos decorrentes de escândalos contábeis corporativos.

A deturpação de informações financeiras das empresas listadas, em particular, distorce a tomada de decisões de várias entidades econômicas envolvidas em transações financeiras, como

ações, negociação de títulos e empréstimos bancários, o que introduz uma alocação ineficiente de recursos. Nas relações comerciais reais, essas declarações incorretas também podem levar a riscos excessivos, dos quais nem o cliente nem o fornecedor estão cientes. Consequentemente, quando esse risco se torna aparente, pode ocorrer estagnação não intencional das atividades econômicas. Se a declaração incorreta for intencional ('fraude contábil') e ocorrer com frequência, poderão ocorrer consequências ainda mais graves, como quebra de mercado.

A pesquisa sobre mecanismos contábeis fraudulentos até agora tem sido principalmente teórica e conduzida em grande parte no campo contábil (por exemplo, Dechow et al. 1996). Uma vertente da literatura teórica concentra-se em possíveis 'razões' para uma empresa cometer fraudes contábeis. Os motivos discutidos para uma empresa falsificar intencionalmente as informações financeiras são, por exemplo, a tentativa de obter melhores termos de aquisição ao enfrentar um desempenho comercial lento. Esses estudos geralmente descrevem o comportamento ideal de uma empresa, dadas certas funções e restrições objetivas, especificando, portanto, os determinantes da fraude contábil.

A segunda vertente da literatura teórica busca identificar informações financeiras estatisticamente correlacionadas com a ocorrência de fraude contábil.

Um grande número de estudos analisou, por exemplo, 'provisões contábeis discricionárias'. Pensa-se que esses fatores estejam relacionados ao comportamento de ajuste dos lucros da administração e que estão fortemente correlacionados com fraudes contábeis.

Nos últimos anos, houve progresso na literatura empírica, que utiliza modelos estatísticos para detectar fraudes contábeis nas demonstrações financeiras e, assim, identifica as empresas que provavelmente se envolverão em comportamento fraudulento. Em estudos empíricos, Dechow et al. (2011) e Song et al. (2016) incorporam as variáveis independentes responsáveis pela fraude contábil nos modelos paramétricos para detectar um evento simultâneo. Esses modelos exibem um bom ajuste na amostra e também confirmam a conjectura teórica.

Embora a literatura empírica tenha feito um bom progresso na detecção de fraudes contábeis, ainda há um grande número de variáveis a serem consideradas. A inclusão de apenas algumas

A declaração incorreta de informações financeiras ('falsificação de demonstrações financeiras') é um evento econômico sério

UMA

UMA

Relacionado

- [Direcionamento de política eficaz com aprendizado de máquina](#)
Monica Andini, Emanuele Ciani, Guido de Blasio, Alessio D'Ignazio
- [Usando a Lei de Benford para detectar fraudes fiscais no comércio internacional](#)
Beata Javorcik Banu Demir

variáveis em um modelo, embora útil para testar hipóteses teóricas, negligencia uma grande quantidade de informações sobre empresas que podem trazer poder de previsão adicional.

Nosso objetivo em Kondo et al. (2019) é, portanto, incorporar um grande número de variáveis explicativas em um modelo que detecta fraudes contábeis usando métodos de aprendizado de máquina (um tipo de floresta aleatória desenvolvida em Chen et al. 2004). Conforme discutido em dois artigos recentes bem conhecidos (Perols 2011, Perols et al. 2017), detectar fraudes contábeis é como procurar uma agulha no palheiro. Nosso modelo utiliza a faixa mais ampla possível de informações corporativas, como indicadores financeiros construídos a partir de informações financeiras corporativas; variáveis relacionadas à governança corporativa, com foco nas informações dos acionistas, executivos e funcionários; e variáveis de transação bancária com base em informações bancárias como variáveis explicativas.

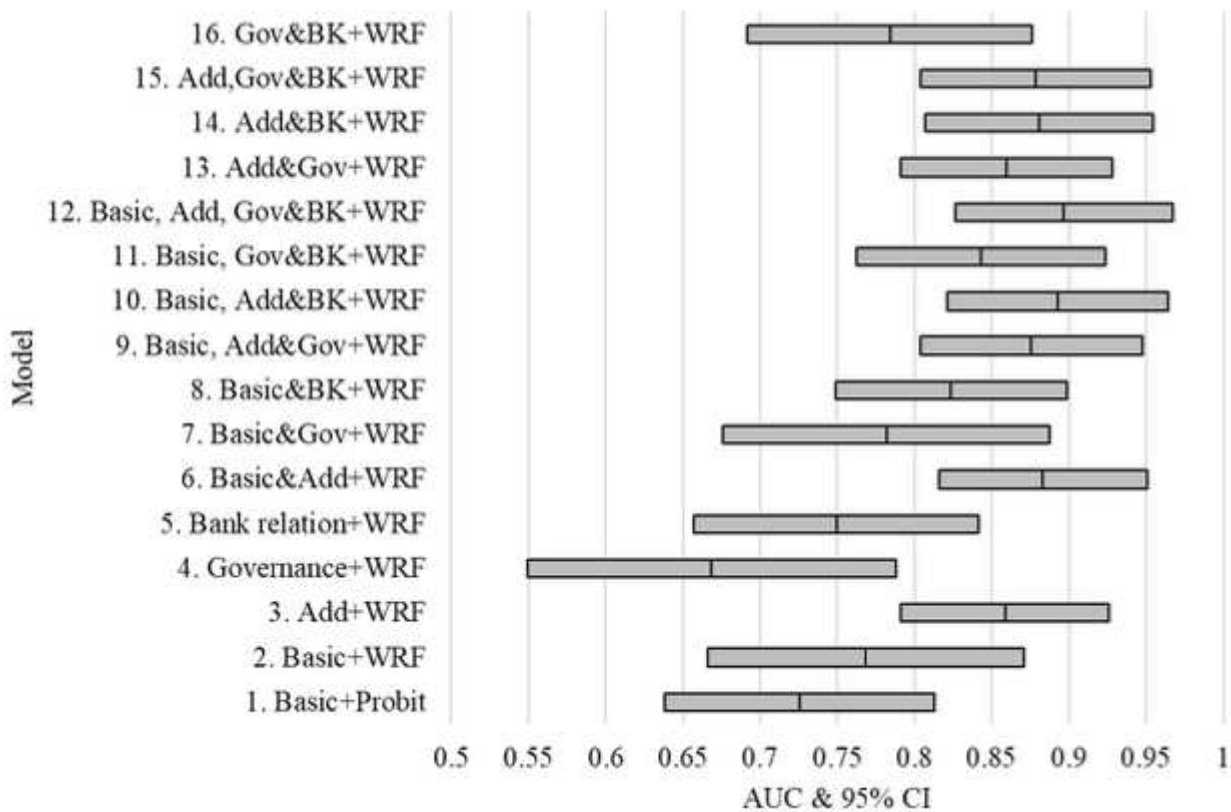
Um segundo objetivo do nosso modelo é abordar a ocorrência de futuras fraudes contábeis, ou 'previsão'. A maior parte da pesquisa existente concentra-se explicitamente na detecção de fraudes contábeis nas declarações atuais ('nowcasting'), que, diferentemente de um evento de falência, não são necessariamente facilmente observáveis ao público. Portanto, o 'nowcasting' tem um grande valor prático para a auditoria. Do ponto de vista dos negócios de auditoria, também seria prudente rastrear empresas especificamente onde é mais provável que ocorram erros de declaração. Assim, usamos nossa estrutura de análise baseada em aprendizado de máquina para prever fraudes contábeis.

Verificamos o desempenho do nosso modelo de previsão da seguinte forma.¹ Usando informações de fraude de conta dos dados de empresas abertas no Japão, comparamos o desempenho de diferentes modelos. O Modelo 1 é um modelo paramétrico que depende apenas de variáveis limitadas usadas em pesquisas existentes. O Modelo 2 é um modelo não paramétrico que usa aprendizado de máquina e depende apenas de variáveis limitadas e o modelo 3-16 são modelos não paramétricos que usam aprendizado de máquina com um conjunto de variáveis em expansão. O Modelo 12, especificamente, é um modelo completo, baseado em todas as variáveis descritas acima.

Primeiro, descobrimos que o uso de técnicas de aprendizado de máquina e o espaço de alta dimensão aumentam o desempenho da detecção (Figura 1). Como a expansão de variáveis no modelo 12 é possível apenas com o uso de aprendizado de máquina, não é fácil medir quanto a aplicação dos métodos de aprendizado de máquina versus a expansão do espaço variável realmente contribuiu para o melhor desempenho de detecção de fraude. Os resultados sugerem que há espaço para engenhosidade quando se trata dos métodos de construção de um modelo de detecção, mesmo quando se utiliza um grupo variável semelhante ao da pesquisa existente.

Figura 1 Resultados da avaliação de desempenho

Detection & Accounting Fraud Flag 2

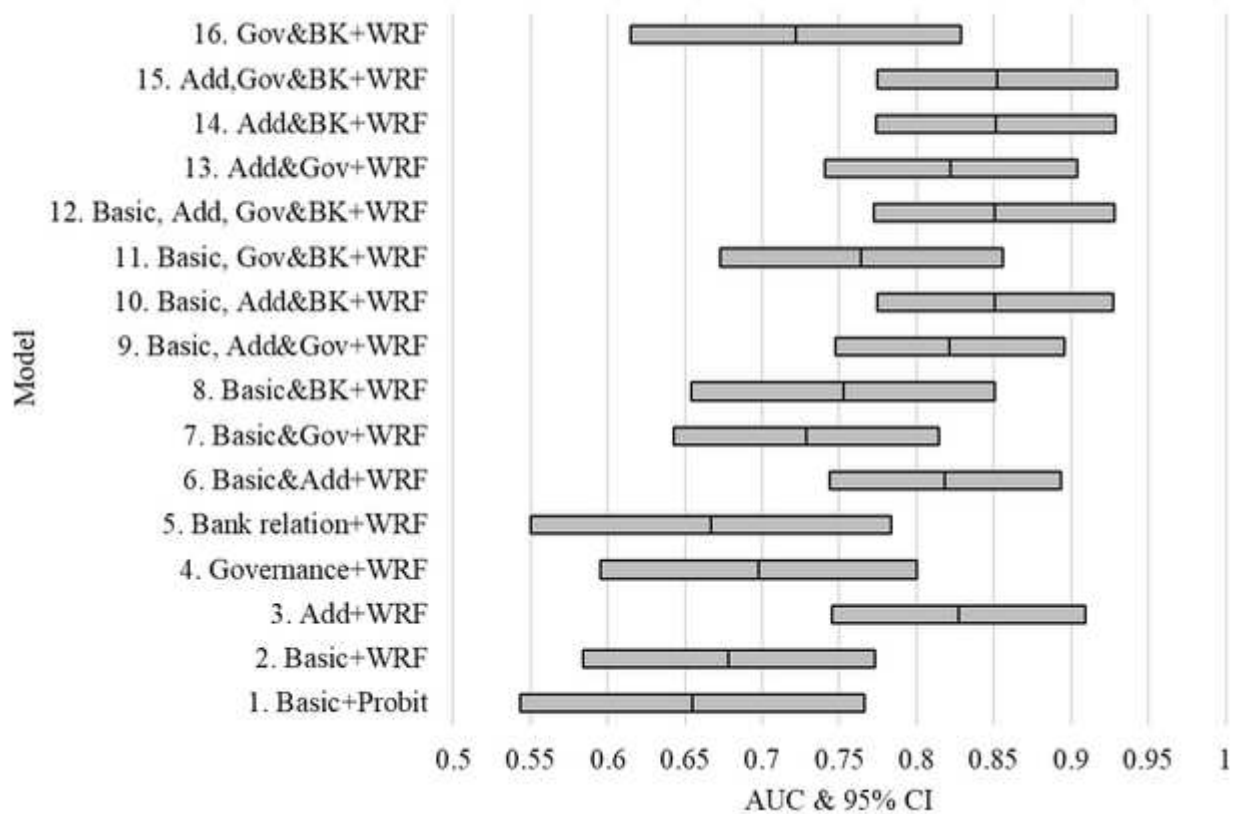


Nota : A Figura 1 mostra a AUC, um indicador de avaliação de desempenho, de todos os modelos de detecção treinados usando os Sinalizadores de fraude contábil 2 (eventos graves e auxiliares) e pontuados nos dados do teste, juntamente com seu intervalo de confiança de 95%. Veja Kondo et al. (2019) para mais detalhes.

Validamos o desempenho do nosso modelo em relação ao nosso segundo objetivo de previsão usando uma amostra de espera (Figura 2). Confirmamos que nosso modelo atinge um desempenho de previsão suficiente. Portanto, adicionamos à literatura mostrando que vale a pena usar um modelo de aprendizado de máquina com dados de alta dimensão para prever futuras fraudes contábeis.

Figura 2 Resultados da avaliação de desempenho com relação à previsão

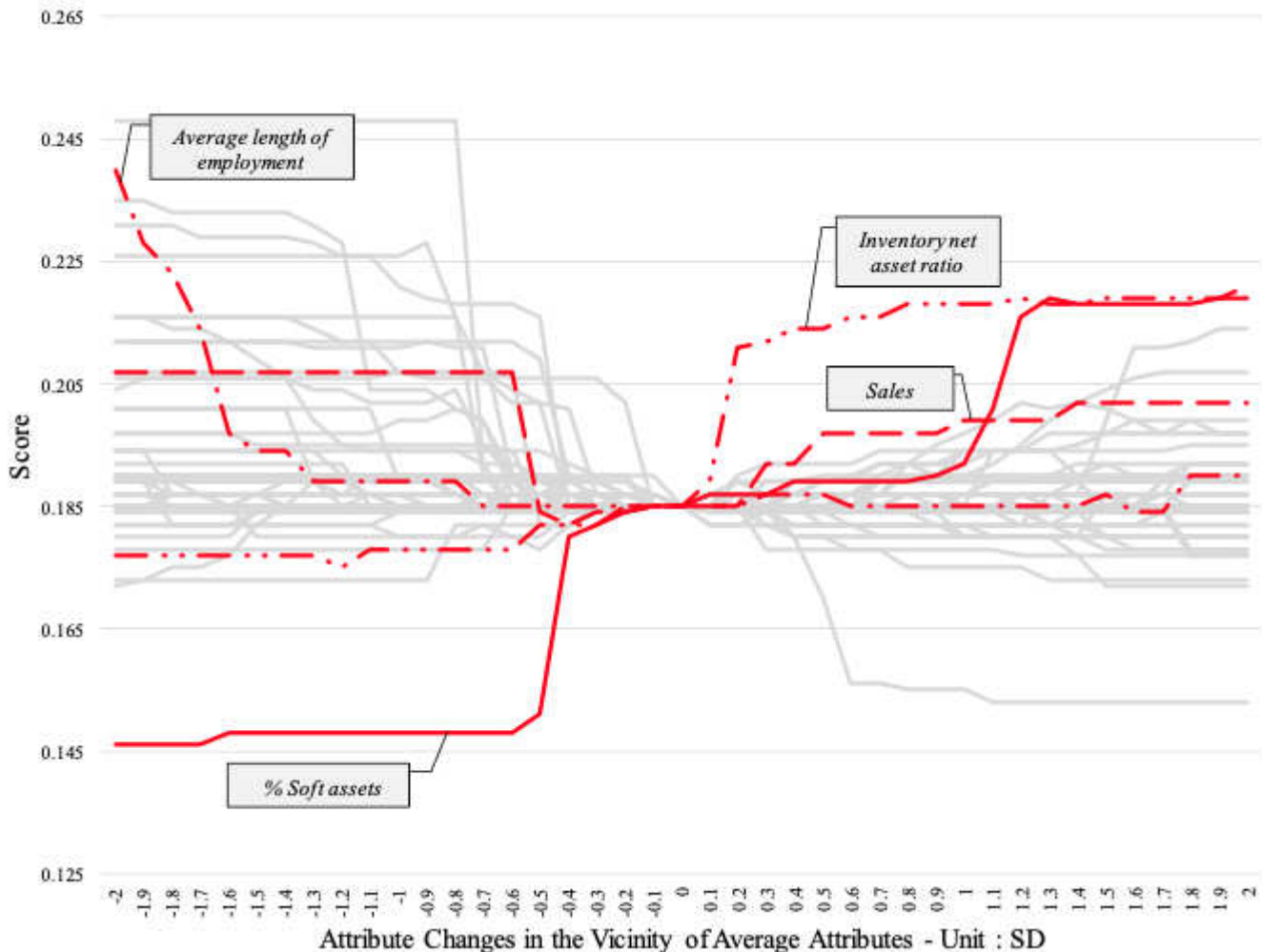
Prediction & Accounting Fraud Flag 2



Nota: A Figura 2 mostra a AUC, um indicador de avaliação de desempenho, de todos os modelos de previsão treinados usando o Accounting Fraud Flags 2 (eventos graves e auxiliares) e pontuados nos dados do teste, juntamente com seu intervalo de confiança de 95%. Veja Kondo et al. (2019) para mais detalhes.

Também mostramos que outras variáveis além daquelas utilizadas em pesquisas existentes podem ser úteis, pelo menos até certo ponto. Especificamente, a duração média do serviço dos funcionários e a porcentagem de ações em circulação mantidas pelos executivos da empresa são altas em termos de grau de importância. Demonstramos na Figura 3 como a pontuação estimada da previsão varia com essas variáveis. Os resultados sugerem que pode haver muitas outras variáveis que contribuem significativamente para a detecção e previsão de fraudes além das consideradas na literatura até o momento. Portanto, nossos resultados indicam que há espaço para esse modelo ser de uso mais prático se o espaço de recursos for expandido e, ao mesmo tempo, exigir mais pesquisas teóricas sobre mecanismos de fraude contábil.

Figura 3 Relação entre alterações de atributos nas proximidades dos atributos médios e alterações na pontuação



Nota : A Figura 3 ilustra mudanças na pontuação quando cada atributo é alterado para cima ou para baixo.

Deve-se notar que a construção do modelo baseado em aprendizado de máquina requer o estabelecimento de vários hiperparâmetros, especialmente na etapa de treinamento. Alguns exemplos incluem o número de árvores a serem construídas, o número mínimo de ramificações e o tipo de estatística mencionada na definição das regras de ramificação. Embora os resultados em geral não dependam desses parâmetros, é necessário salientar que os métodos de aprendizado de máquina não automatizam todas as tarefas envolvidas na construção do modelo.

Além disso, deve-se reconhecer que existem outros métodos de aprendizado de máquina para detectar e prever fraudes contábeis, além do utilizado em nosso artigo. De fato, alguns dos autores deste estudo alcançaram um bom desempenho usando diferentes métodos de aprendizado de máquina, o que sugere que um modelo mais robusto possa ser obtido no futuro combinando dezenas de modelos diferentes. Como é obviamente benéfico construir um modelo compacto e fácil de usar, os modelos futuros podem se basear em nosso estudo e limitar o número de variáveis usadas àquelas que consideramos relativamente importantes do ponto de vista da previsão. No entanto, deve-se observar que, mesmo que os resultados desejados sejam os mesmos, variáveis diferentes têm uma importância muito alta, dependendo se o objetivo é detectar ou prever fraudes. Além disso, mesmo ao executar a mesma previsão, configurações diferentes de sinalizadores de destino podem causar diferenças nos grupos de variáveis com uma classificação alta. Assim, a seleção de variáveis deve ser realizada de acordo com o objetivo do modelo ao construir um modelo tão compacto e fácil de usar.

Outra ressalva é que nosso modelo preditivo assume ambientes socioeconômicos e padrões contábeis constantes. Por exemplo, se houver um tipo completamente diferente de fraude em um novo negócio, o modelo atual poderá não ser capaz de detectar ou prever isso. Além disso, mesmo que as realidades de uma empresa permaneçam as mesmas, as variáveis financeiras podem assumir valores diferentes devido a mudanças nos padrões contábeis, o que pode afetar

os resultados de detecção e previsão de fraude. Portanto, é importante monitorar regularmente a eficácia do modelo.

Após a construção de um modelo preditivo satisfatório, a pontuação resultante deve ser analisada e as reações adequadas precisam ser decididas. Como existem empresas que atualmente cometem fraudes contábeis ou que provavelmente as cometerão no futuro próximo, uma equipe profissional com conhecimento especializado precisará lidar com esses casos imediatamente. É necessário um bom julgamento para estabelecer quando a probabilidade de fraude contábil é substancial o suficiente para justificar interferência. Relacionada a isso, a detecção de anomalias no nível da conta é um importante tema de pesquisa que complementa nosso modelo fraudulento no nível da empresa. Também devemos observar que há uma grande preocupação com relação à medida em que a equipe profissional local usa essas pontuações quando as pontuações do modelo são fornecidas. Portanto,

Há uma preocupação adicional de uma perspectiva política sobre como adotar essa tecnologia de previsão. Por exemplo, se os detalhes de um modelo para detectar contabilidade fraudulenta se tornarem amplamente conhecidos, as empresas que estiverem tentando algum tipo de fraude contábil poderão evitar a detecção. Na realidade, é improvável que os detalhes do modelo se tornem amplamente conhecidos e, portanto, não é fácil cometer fraudes sem ser detectado pelo modelo. Uma solução possível para esse problema seria melhorar continuamente os modelos, de modo que as empresas envolvidas em esforços maliciosos para contornar esses modelos achem mais difícil fazê-lo.

Outro tópico que precisa de mais discussão ao considerar medidas avançadas para evitar fraudes contábeis é a questão de se é possível suprimir eventos fraudulentos adotando uma abordagem de auditoria específica ou dispositivo de governança.

Nota do editor: A principal pesquisa em que esta coluna se baseia apareceu pela primeira vez como um [Documento](#) de [Discussão](#) do Instituto de Pesquisa de Economia, Comércio e Indústria (RIETI) do Japão.

Referências

Chen, C, A Liaw e L Breiman (2004), “Usando a floresta aleatória para aprender dados desequilibrados”, Relatório Técnico 666 Departamento de Estatística da Universidade da Califórnia em Berkley.

Dechow, PM, W Ge, CR Larson e RG Sloan (2011), “Prevendo distorções na contabilidade de materiais”, *Contemporary Accounting Research* 28: 17-82.

Dechow, PM, RG Sloan e AP Sweeney (1996), “Causas e conseqüências da manipulação de resultados: uma análise de empresas sujeitas a ações de execução por parte da SEC”, *Contemporary Accounting Research* 13: 1-36.

Kleinberg, J, H Lakkaraju, J Leskovec, J Ludwig e S Mullainathan (2018), “Decisões humanas e previsões de máquinas”, *Quarterly Journal of Economics* 133: 237-293.

Kondo, S, D Miyakawa, K Shiraki, M Suga e T Usuki (2019), “Usando o Machine Learning para detectar e prever fraudes contábeis”, RIETI Discussion Paper Series 19-E-103.

Perlos, J (2011), “Detecção de fraudes por demonstrações financeiras: uma análise de algoritmos estatísticos e de aprendizado de máquina”, *Auditoria: A Journal of Practice & Theory* 30: 19-50.

Perlos, J, B Bowen e C Zimmerman (2017), “Localizando agulhas no palheiro: usando o Data Analytics para melhorar a previsão de fraudes”, *Accounting Review* 92: 221-245.

Canção M, N Oshuro e A Shuto (2016), “Predicting Accounting Fraud: Evidence from Japan”, *The Japanese Accounting Review* 6: 17-63.

West, J e M Bhattacharya (2016), “Detecção inteligente de fraudes financeiras: uma revisão abrangente”, *Computer & Security* 57: 47-66.

Notas finais

1 Veja West e Bhattacharya (2016) para uma pesquisa sobre tendências recentes de desenvolvimento de modelos de previsão e previsão visando fraudes (cartões de crédito, valores mobiliários, seguros) no setor financeiro, incluindo fraudes contábeis.