Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Uma proposta de análise on-line e identificação de transações financeiras fraudulentas com visual analytics

Rodrigo Araujo Lima Torres

Dissertação apresentada como requisito parcial para conclusão do
Mestrado Profissional em Computação Aplicada

Orientador
Prof. Dr. Marcelo Ladeira

Brasília
2020

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

Torres, Rodrigo Araujo Lima
TT693p Uma proposta de análise on-line e identificação de transações financeiras fraudulentas com visual analytics / Rodrigo Araujo Lima Torres; orientador Marcelo Ladeira. -- Brasília, 2020.
54 p.

Dissertação (Mestrado - Mestrado Profissional em Computação Aplicada) -- Universidade de Brasília, 2020.

1. Aprendizado de Máquina. 2. Detecção de Fraudes Bancárias. 3. Mineração de Dados. 4. Visual Analytics. 5. Identificação de Outliers. I. Ladeira, Marcelo, orient. II. Título.

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Uma proposta de análise on-line e identificação de transações financeiras fraudulentas com visual analytics

Rodrigo Araujo Lima Torres

Dissertação apresentada como requisito parcial para conclusão do
Mestrado Profissional em Computação Aplicada

Prof. Dr. Marcelo Ladeira (Orientador)
CIC/UnB

Prof. Dr. Adriano Lorena Inácio de Oliveira
Universidade Federal de Pernambuco

Prof. Dr. João Carlos Felix Souza
Universidade de Brasília

Prof. Dr. Luís Paulo F. Garcia
Universidade de Brasília

Prof. Dr. Marcelo Ladeira
Coordenador do Programa de Pós-graduação em Computação Aplicada

Brasília, 23 de Novembro de 2020

Dedicatória

Eu dedico essa dissertação à minha esposa, Nayara Lima Torres, que tanto me apoiou e me incentivou durante as dificuldades enfrentadas ao longo do período do mestrado. A Deus, que me iluminou, me deu coragem e me permitiu perseverar. Ao meu orientador pelos ensinamentos e que me mostrou o caminho do sucesso, seguindo sempre com fé, força e foco. Por fim, ao meu empregador que me deu a oportunidade de desenvolver o tema desta pesquisa, fornecendo informações e ferramentas. Obrigado!

Agradecimentos

Ao Banco X que apoiou esse trabalho e concordou em fornecer dados reais, respeitando todas as conformidades, a privacidade do cliente e as normas de segurança. Os dados utilizados foram devidamente descaracterizados e nenhum dado pessoal de cliente foi divulgado, respeitando a Lei Geral de Proteção de Dados, e foram cruciais para o desenvolvimento deste trabalho.

Ao programa PPCA da Universidade de Brasília, a Deus e à minha família com seu apoio e encorajamento.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), por meio do Acesso ao Portal de Periódicos.

Resumo

As instituições financeiras lidam com centenas de milhares de transações bancárias por dia e precisam garantir segurança e qualidade para seus clientes. Atualmente, pesquisas em padrões predefinidos são insuficientes para identificar fraudes devido à evolução contínua dos métodos fraudulentos usados pelos criminosos.

Os sistemas utilizados para este fim baseiam-se na aplicação de alguns métodos de Inteligência Artificial, que negligenciam a análise de processos humanos e fazem pouco uso das técnicas de *Visual Analytics* (VA). O domínio de detecção de fraudes envolve aspectos orientados ao tempo e multivariados para identificar transações anômalas, tornando a detecção de fraudes uma tarefa difícil.

Este trabalho descreve a criação de um modelo para cada cliente, com base em seu comportamento de movimentações financeiras, utilizando técnicas de identificação de *outliers*, que sinalizam possíveis fraudes. Os resultados são apresentados em consoles de VA, que permitem exploração visual, descoberta e integração da análise humana, reduzindo a taxa de falsos positivos na identificação de transações financeiras fraudulentas. Aplicamos essa abordagem a uma instituição financeira brasileira, com um volume diário de mais de 30 milhões de transações e movimentações bancárias.

Este projeto envolve uma abordagem híbrida composta pelo: **(1)** uso de algoritmos não supervisionados para de detecção *outliers*; e **(2)** uso de VA para apoiar a análise humana em tempo real com o objetivo de reduzir a incidência de falsos positivos.

Informações fraudulentas potenciais são apresentadas usando técnicas de VA que permitem aos especialistas avaliar transações suspeitas sem aumento dos tempos normais de processamento. Os resultados obtidos sinalizam que nossa abordagem pode superar o desempenho do método de detecção de fraudes hoje utilizado na instituição brasileira.

Palavras-chave: Aprendizado de Máquina, Detecção de Fraudes Bancárias, Mineração de Dados, Visual Analytics, Identificação de Outliers.

Abstract

Financial institutions handle with hundreds of thousands of wire transactions per day and need to ensure security and quality for their customers. Searching on predefined patterns is insufficient to identify frauds due to continuous evolution of fraudulent methods used by criminals.

Systems used for this purpose are based on the application of some methods of Artificial Intelligence, neglect human process analysis and make little use of Visual Analytics (VA) techniques. Frauds detection domain involves time-oriented and multivariate aspects to identify anomalous transactions making fraud detection a difficult task.

This work describes the creation of a model for each client, based on their behavior of financial transactions, using outlier identification techniques, which indicate possible fraud. The results are presented on VA consoles, which allow visual exploration, discovery and integration of human analysis, reducing the rate of false positives in the identification of fraudulent financial transactions. We apply this approach to a Brazilian financial institution, with a daily volume of over 30 million bank transactions and transactions.

Our framework includes a hybrid approach composed of: **(1)** use of unsupervised outlier detection algorithms; and **(2)**u se of VA to support the real time human analysis with the aim of reducing the incidence of false positives.

Potential fraudulent information are presented using VA techniques allowing specialists to evaluate suspicious transactions with no increase of the normal processing times. The results obtained sign evidence that our approach can overcome the performance of the fraud detection method today used at the Brazilian institution.

Keywords: Banking Frauds Detection, Data Mining, Machine Learning, Visual Analytics, Outliers Identification.

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

Definição do Problema

Bancos brasileiros e outras instituições financeiras ao redor do mundo sofrem impactos de fraudes em transações monetárias. A Confederação Nacional de Dirigentes Lojistas (CNDL) e o Serviço de Proteção ao Crédito (SPC Brasil) realizaram uma pesquisa[1] mostrando que 46% dos brasileiros sofreram algum tipo de golpe na internet entre 2018 e 2019. As perdas estimadas equivalem a R$ 1,8 bilhão (cerca de US$ 400 milhões), impactando 12,1 milhões de pessoas.

O banco brasileiro que apoiou esse trabalho, a partir de agora chamado Banco X, tem 4,4 milhões de clientes em todo o país e está presente em todos os estados brasileiros e no Distrito Federal. Registra mais de 30 milhões de transações por dia. Embora a maioria dessas transações sejam legítimas, 5 a 10 clientes relatam sofrer fraudes financeiras todos os meses. Atualmente, 73% das transações são feitas em canais digitais, onde 53% são por aplicativos móveis. O Banco X informou que 1.388 clientes foram vítimas de algum tipo de fraude de janeiro a julho de 2020. Nesse período, o Banco X registrou prejuízos de R$ 32,27 milhões (cerca de US$ 7 milhões) que não puderam ser recuperados. A tabela 1.1 descreve o impacto financeiro por tipo de transação e o número de clientes que sofreram algum tipo de fraude.

O Banco X e outras instituições financeiras têm dificuldade em processar, em tempo real (intervalo de tempo inferior ao prazo máximo definido pelo Banco Central do Brasil, para que uma transação seja processada, de acordo com o seu tipo, valor e canal onde foi iniciada), todo o volume de dados de transações e identificar, em tempo hábil, aquelas que são fraudes. As tarefas de monitoramento de transações bancárias envolvem dados de natureza complexa [1], com base em atributos temporais, que requerem meios sofisticados para análise e exploração.

[1] `https://site.cndl.org.br/mais-de-12-milhoes-de-consumidores-sofreram-alguma-fraude-\financeira-nos-ultimos-12-meses-aponta-pesquisa-cndlspc-brasil/`

Tabela 1.1: Impactos financeiros de fraudes por tipo de transações

Tipo da Transação	**Quantidade**	**Perdas Financeiras - R$**
Boletos de Cobrança	632	5.756.276,90
Cheques	1	15.000,00
DOC	17	44,044.80
TEC (Intercredis)	73	613.938,96
TED	678	25.841.114,92
Total	**1.401**	**32.270.375,58**

Quantidade de fraudes e valor total dos prejuízos por tipo de transação
Fonte: Banco X

Os impactos das fraudes bancárias repercutem sobre os clientes e instituições financeiras, rompem o equilíbrio econômico, causam impactos à imagem e geram altos custos operacionais às organizações. Atualmente, a área de Prevenção e Combate à Fraude do Banco X possui 20 profissionais dedicados à identificação de anomalias e atuação nos casos em que haja confirmação de movimentações fraudulentas. Atualmente, estima-se um volume médio entre 5 a 10 ocorrências deste tipo por dia.

Deste volume de ocorrências, 97% das fraudes são identificadas pelo próprio cliente, que sinaliza ao Banco X algum tipo de movimentação indevida em sua conta bancária. A maior desafio da área de prevenção e combate à fraude da instituição é conseguir analisar e identificar, dentre as mais de 30 milhões de transações diárias, àquelas que são fraudulentas, e atuar de forma tempestiva evitando perdas e prejuízos financeiros. Trata-se de uma barreira que impõe altos custos operacionais e que afeta a confiança e segurança percebida por parte dos clientes.

As fraudes são um risco inerente ao negócio bancário onde, dentre as diversas publicações de soluções para se resolver este problema, observa-se um aumento, desde 2011, no número de artigos sobre a adoção da abordagem de VA, como forma de identificar fraudes bancárias, em tempo real, e alternativa capaz de processar o volume diário de transações bancárias. Segundo Kielman [2], o processo de detecção de fraudes em transações bancárias é definido como sendo um problema de VA.

A Figura 1.1 ilustra a distribuições das fraudes de acordo com o tipo de movimentação. Observa-se que a média dos valores fraudados por meio de cheques é superior aos demais tipos de movimentação. Porém, 98,5% do total de transações de TED (Transferência Eletrônica Disponível) possuem valor de até R$ 219.230,00 reais. Boletos de cobrança e TED, possuem transações cujos valores geram os maiores impactos financeiros, com casos de fraudes de até R$ 400.000,00 e R$ 495.000,00, respectivamente, bem superiores aos valores envolvidos nos demais tipos de transações.

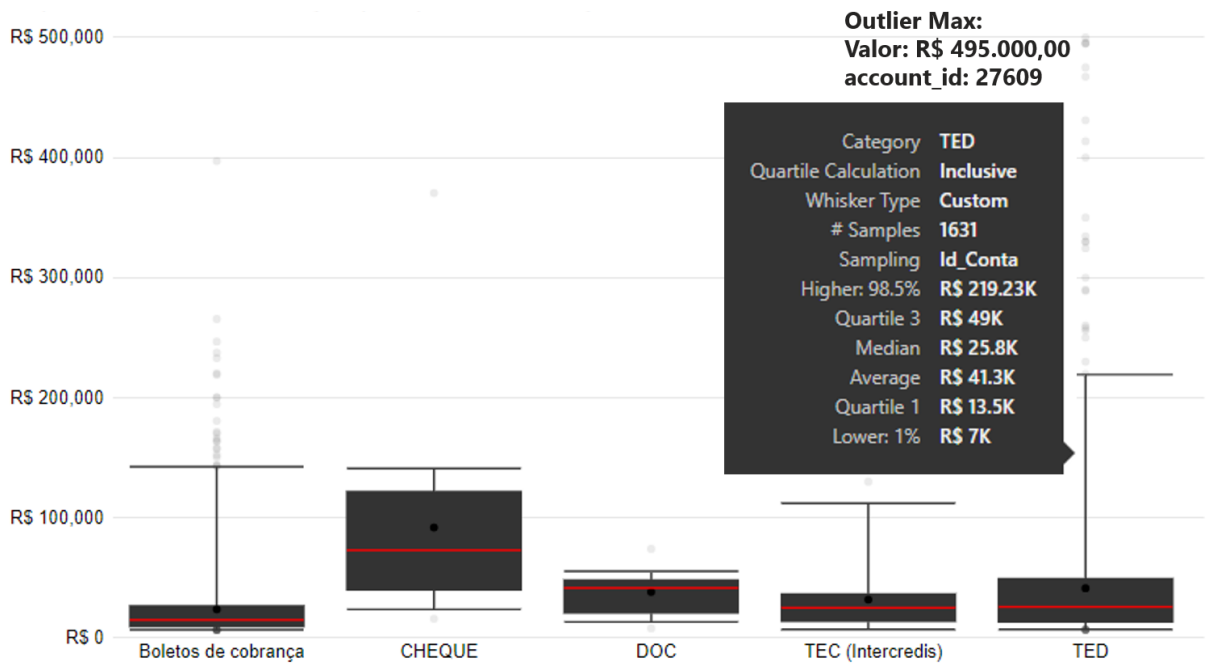

Figura 1.1: Distribuição das fraude por valor e tipo de movimentação

As soluções de detecção de fraudes comerciais no mercado geralmente envolvem técnicas de mineração de dados, negligenciam a análise de processos humanos e fazem pouco uso das técnicas de VA (*Visual Analytics*). Kielman et al. [2] descrevem a detecção de fraudes como um problema de VA aberto que requer exploração visual, descoberta e análise. Olhando para os cenários de fraude financeira, West et al. [3] e Hoogs et al. [4] concluíram em seu trabalho que técnicas supervisionadas, como regressão logística, *Support Vector Machine* e *Random Forest*, demonstraram uma baixa precisão quando usadas para classificar eventos fraudulentos ou produzem muitos falsos positivos.

De forma a reduzir a complexidade temporal e evitar influências dos padrões de transações entre as contas, a abordagem deste trabalho é diferente. O SDF (Sistema de Detecção de Fraudes) proposto, constrói um modelo para cada conta do cliente com base em seu comportamento financeiro. Através do SDF é possível identificar *outliers* de transações, e técnicas de VA são utilizadas para reduzir a taxa de falso positivo, por meio das seguintes fases: Geração de perfil com base no comportamento do cliente; Análise de Transações em tempo real; Sinalizar transações *online* suspeitas, usando VA; e Avaliação e Intervenção Humana.

Os dados utilizados no projeto são reais, do Banco X, que foram construídos em colaboração com especialistas da instituição que atuam na área de prevenção e combate a fraudes. Com o SDF em execução, os especialistas podem agir sem causar impactos ou atrasos nas transações *online* legítimas. A mesma equipe pode analisar um conjunto

maior de relatórios e a análise de gráficos de VA ajuda a identificar um maior número de ocorrências de fraude.

Outra contribuição do trabalho é a capacidade de lidar com um enorme volume diário de transações financeiras, permitindo descobrir e sinalizar, em tempo real, possíveis fraudes sem gerar impactos ou atrasos nos tempos de processamento de transações legítimas.

Os conjuntos de dados utilizados neste trabalho são muito sensíveis. É imperativo que se cumpra os regulamentos de privacidade e segurança definidos pela Resolução CMN (Conselho Monetário Nacional) 4.658/2018, emitida pelo Banco Central do Brasil, a qual dispõe sobre a política de segurança cibernética, assim como, Leis 13.709/2018 e 13.853/2019, que entram em vigor em agosto de 2020, conhecidas como LGPD (Lei Geral de Proteção de Dados Pessoais). Assim, não é permitido fornecer o conjunto de dados e códigos utilizados, para evitar o uso indevido dessas informações pelos fraudadores.

Entretanto, as fases do framework utilizado para construção do SDF são detalhadas nos demais capítulos.

1.1 Hipóteses do Projeto

Huang [10] afirma em seus trabalhos que além dos custos associados aos falso positivos (em caso de sinalização indevida de uma transação legítima), existem àqueles relativos aos falso negativos (fraudes não detectadas), que podem ser ainda maiores que os primeiros. Neste sentido, este trabalho visa verificar a hipótese de que a integração da análise humana ao processo de investigação das transações suspeitas de fraude melhora o desempenho dos modelos e permite reduzir as incidências de falsos positivos.

Para este projeto foram utilizados 6 meses de dados de 1.388 clientes que sofreram algum tipo de fraude entre janeiro e julho de 2020. Este conjunto de dados possui 2 classes de transações: legítima e fraudulenta. A quantidade de registros que representam transações fraudulentas equivalem a 0,0001% da amostra.

A distribuição desbalanceada do conjunto de dados provou ser um grande desafio a ser resolvido. Alguns autores como West et al. [3] afirmam a baixa precisão dos algoritmos supervisionados para a detecção de fraudes está relacionada à imensa disproporcionalidade entre as transações legítimas e fraudulentas. Neste sentido, pretende-se verificar a hipotese de que o uso de algoritmos não supervisionados, para sinalizar as ocorrências de transações fraudulentas, possui melhor desempenho do que os supervisionados, se valendo de funções de busca de padrões e identificação de *outliers*, considerando-se a interdependência das séries temporais de transações de cada cliente.

1.2 Objetivos

O objetivo geral da pesquisa é implementar uma prova de conceito capaz de sinalizar, em tempo real, as transações com possibilidade de serem fraudulentas, utilizando algoritmos não supervisionados de identificação de *outliers*.

Como objetivos específicos tem-se a criação de um framework de detecção de eventos e um *dashboard* interativo utilizando técnicas de VA, como alternativa capaz de:

1. **Integrar a análise humana ao processo** de investigação das transações financeiras suspeitas, como uma abordagem visando reduzir incidências de falsos positivos;
2. **Técnicas de VA** para disponibilizar painéis interativos que permitam exploração dos dados e averiguação das ocorrências de possíveis fraudes e anomalias; e
3. **Permitir a atuação dos profissionais**, da área de Prevenção e Combate à Fraude, sem que hajam impactos ou atrasos nas transações *online* legítimas.

A meta é de reduzir em 10% as perdas anuais com as fraudes, o que representa aproximadamente 1,2 milhões de reais por ano. Uma vez que a triagem (filtros) das transações classificadas como possíveis fraudes, será realizada de forma automatizada, o FDS escala a produtividade do time, permitindo que mais transações sejam analisadas pelos especialistas da área de prevenção e combate à fraudes.

O Capítulo 2 contém os trabalhos correlatos ao tema desta dissertação. Nos Capítulo 3 e 4 estão descritos a metodologia utilizada e os experimentos realizados e a estrutura do sistema de detecção de fraudes. No Capítulo 5 tem-se os resultados obtidos com os experimentos e é apresentado o painel do VA. Por fim, o Capítulo 6 contém conclusões e trabalhos futuros.

Capítulo 2

Revisão da Literatura

Este capítulo apresenta estudos correlatos relacionados ao tema de fraudes em transações bancárias e uso de VA (*Visual Analytics*), como abordagem gráfica para sinalizar possíveis suspeitas de fraudes através de painéis interativos. Inicialmente, foi utilizado a *string (TS=(Financ* AND Fraud* AND Detect*))* que retornou 661 artigos relacionados ao domínio de fraudes financeiras.

Para a revisão sistemática da literatura, aplicou-se técnicas bibliométricas quantitativas de gráfico de rede e o *bibliographic-coopling*, visando identificar as tendências do tema. A TEMAC (Teoria do Enfoque Meta-Analítico Consolidado) foi aplicada seguindo os passos descritos por Ari Mariano Rocha [5], visando o filtro e identificação dos artigos de maior impacto.

Para identificar os estudos mais impactantes no domínio de fraudes, e usando o VA como abordagem para sinalizar transações suspeitas, usamos uma nuvem de palavras para avaliar termos de títulos, palavras-chave e resumo dos artigos e encontrar os termos mais utilizados. O *bibliographic-coopling* com intuito de identificar tendências e selecionar os artigos mais relevantes nas bases do *Web of Science*, Google Acadêmico e Scopus. O *VOS Viewer* foi utilizado nas análises do número de citações, clusters, força das conexões e peso dos artigos. A Figura 2.1 ilustra os clusters dos artigos de maior impacto.

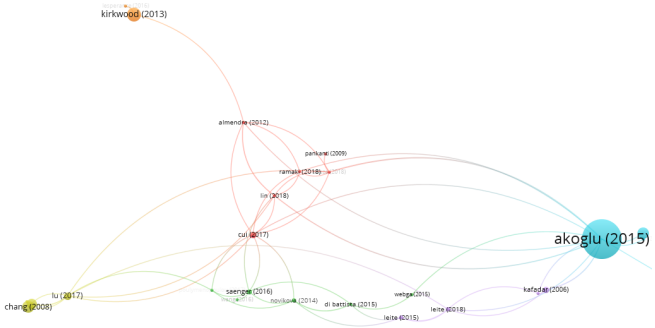

Figura 2.1: Cluster dos artigos de maior impacto em relação ao tema

Os artigos selecionados como os mais relevantes são comentados a seguir.

Desde 2011, o número de publicações relacionadas à abordagem de VA tem crescido para identificar fraudes bancárias em tempo real, e como alternativa capaz de processar o volume de transações envolvidas nesse cenário.

Em 2017 Lu et al. [6] apresentam o estado da arte do uso do VA para identificação de transações suspeitas. Nesta época, o VA começa a ganhar espaço como uma técnica capaz de suportar tarefas de análise preditiva em um *pipeline* de PVA (*Predictive Visual Analytics*), proposto por Lu et al. [6].

Chandola et al. [7] definem, como detecção de *outliers*, o problema de encontrar padrões que não estão em conformidade com o comportamento esperado, conforme ilustrado na Figura 2.2. Em estatística, uma anomalia (também conhecida como *outlier*) é uma observação ou evento que se desvia tanto de outros eventos para levantar suspeitas de que foi gerado por um meio diferente.

A principal contribuição para este trabalho foi a demonstração de modelos eficazes de gráficos de dispersão para identificação de anomalias utilizando técnicas multi-classe e com apenas uma única classe, essencial no contexto deste trabalho, onde observou-se conjuntos de dados de clientes e tipos de transações que possuíam apenas uma classe, já que não foram vítimas de nenhum tipo de fraude. A partir deste trabalho, tomou-se a decisão de utilizar o *One-Class Classifier*. A Figura 2.2 é original do trabalho do autor [7], e por isso está em inglês.

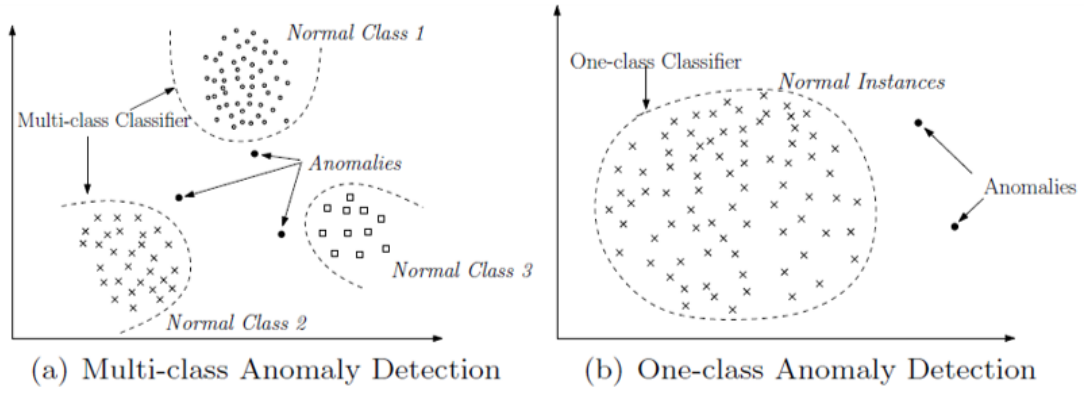

Figura 2.2: *Using classification for anomaly detection.*

Os sistemas tradicionais de mercado, utilizados para detecção de fraudes financeiras, utilizam algoritmos supervisionados que se baseiam no processo de aprendizado das fraudes conhecidas. Estes sistemas buscam identificar comportamentos similares, ou indicam possíveis novas fraudes através da detecção de comportamentos "anormais" [8].

Apesar do número elevado de sistemas e pesquisas que se propõem a detectar fraudes utilizando algoritmos supervisionados, esses sistemas podem sinalizar indevidamente as

ocorrências, gerando custos de falso positivo (em caso de sinalização indevida de uma transação legítima), que são ainda maiores quando ocorrem falso negativos (fraudes não detectadas), conforme citado por Huang [9].

Relativo à baixa precisão dos algoritmos supervisionados para a detecção de fraudes, West et al. [3] afirma que isto se deve à imensa desproporcionalidade entre as transações legítimas e fraudulentas. Esta afirmação é confirmada por Bolton e Hand [10] que demonstram, em suas publicações, que os modelos comumente utilizados possuem um melhor desempenho na classificação de transações legítimas, e uma baixa acurácia na identificação de transações fraudulentas.

Hoogs et al [4] confirma esta situação em seus estudos, utilizando técnicas supervisionadas em bases desbalanceadas. O resultado dos trabalhos destes quatro autores inspirou este estudo no sentido de utilizar algoritmos não supervisionados para identificação das possíveis fraudes. O modelo foi validado em conjunto com especialistas utilizando o cálculo do *F1 measure*, que é a média harmônica entre precisão e revocação, a partir da matriz de confusão gerada com as sinalizações.

Carminati et al. [11] identificam fraudes como um problema complexo porque não há uma variável específica a ser explicada, neste domínio. Sua contribuição neste trabalho foi descrever o uso de algoritmos não supervisionados para identificar *outliers*, que representam fraudes bancárias, em um conjunto de transações financeiras, sem a necessidade de treinamento prévio [12], ou rotulagem de dados, buscando padrões. A identificação dos *outliers* ocorre por meio da geração de perfis baseados nos históricos de transações dos clientes.

Considerando a complexidade dos dados de transações bancárias, Aigner et al. [1] apresenta formas de utilizar visualizações gráficas para se estabelecer relações entre os dados, com base em suas dependências temporais, o que no contexto deste trabalho representaria estabelecer relações entre diferentes tipos de transações ao longo do tempo.

Em uma abordagem similar à de Carminati et al. [11], Roger Leite et al. [13] descreve o envolvimento da análise humana na avaliação das suspeitas de fraudes. Estes autores propuseram alguns modelos semi-supervisionados que inspiraram nossa estrutura de SDF (Sistema de Detecção de Fraudes), sendo capazes de fornecer um método *online* para avaliar se uma transação diverge dos padrões históricos de transferências e pagamentos do cliente, calculando pontuações para cada conta. Eles também apresentam interfaces gráficas para análises humanas e apoio a decisões, buscando a redução de falsos positivos.

O SDF é composto por um *framework* envolvendo: descrição dos passos necessários para a criação dos perfis; e as etapas para análise de *scores online*. Nele estão sendo combinados técnicas de IA (Inteligência Artificial), visualização, reconhecimento de padrões e mineração de dados para dar suporte à análise regulatória, havendo sido acrescentados meios para uma exploração interativa dos dados visualizados.

Markus Goldstein e Andreas Dengel [12] contribuem com a comparação de algoritmos como os HBOS, XBOS e outros modelos não supervisionados, em relação ao tempo de execução e precisão para cada um deles detectar anomalias. HBOS é um algoritmo, baseado em histograma, que permite a detecção de *outliers*, com tempo de execução linear, assumindo a independência dos atributos.

No domínio de fraudes bancárias, o HBOS permite detectar situações em que determinadas transações ocorrem muito raramente dentro do conjunto de dados (este é outro motivo da importânica de se criar um modelo para cada cliente), e onde os atributos diferem significativamente das instâncias normais.

Para a utilização do HBOS foram adotados as seguintes premissas:
(1) Histograma univariado para cada atributo;
(2) Dados categóricos: contagem simples;
(3) Dados numéricos: Tamanho estático do *bin*, com k *bins* de tamanho igual; Tamanho dinâmico do *bin*, com $\frac{N}{k}$ instâncias por *bin*;
(4) A frequência (quantidade relativa) de amostras em um *bin* é usada como estimativa de densidade;
(5) Os histogramas são normalizados para [0,1] para cada atributo único;

Assim, o HBOS para cada instância p é calculado como um produto do inverso da densidade estimada, dada pela seguinte fórmula:

$$HBOS(p) = \sum_{i=1}^{d} log(\frac{1}{hist_i(p)}) \tag{2.1}$$

Devido à precisão do ponto flutuante, o produto é substituído pela soma dos logaritmos (não muda a ordem das pontuações), usando $log(a.b) = log(a)+log(b)$. Então, assume-se a independência de atributos semelhantes ao Naive Bayes, e diferentes técnicas de histograma (categórica, estática, dinâmica) podem ser combinadas.

XBOS é um algoritmo de detecção de anomalias baseado em *clusters*. o XBOS é relativamente mais lento que o HBOS, mas tem uma maior precisão e desempenho superior a outros modelos.

O XBOS possui dois estágios: no primeiro é utilizado o *k-means* para criação dos *clusters*; no segundo estágio são calculadas as interações cruzadas entre os *clusters*. Em decorrência do segundo estágio é possível observar na Figura 2.3 que existe uma maior interação entre os *clusters* C1 e C2. Neste modelo a força da interação está relacionado ao tamanho e proximidade entre os *clusters*.

Na Figura 2.3, C1 é o maior *cluster*. Assim, os pontos de dados pertencentes ao pequeno *cluster* (C2) são classificados como 'não tão anômalos', em comparação com C3,

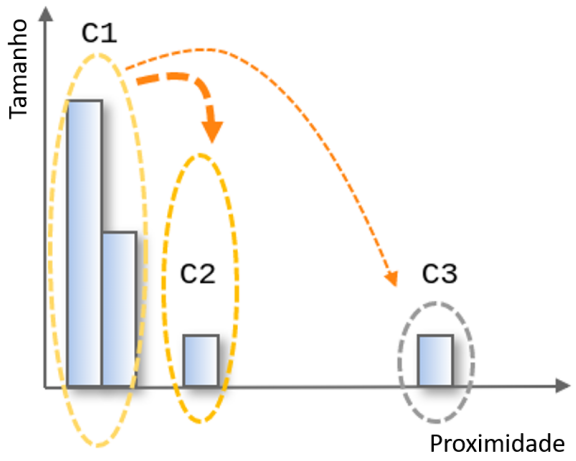

Figura 2.3: XBOS: representação da interação entre *clusters*

por possuírem uma maior proximidade e interação com C1.

Diversos autores tratam de técnicas de identificação de *outliers* em dados multidimensionais. Chandola et al. [7] e Zimek et al. [14] se concentram na identificação de *outliers* em altas dimensões, e Schubert et al. [15] lidam com técnicas de identificação de *outliers* locais.

Por outro lado, os conjuntos de dados nem sempre podem ser tratados de forma independente, devendo serem consideradas as interdependências, durante o processo de detecção de anomalias (ver Figura 2.4). Akoglu et al. [16] foi um dos artigos mais citados, e apresenta estudos de uso de gráficos eficazes para capturar as interdependências dos objetos de dados, permitindo a realização de análises preditivas e detecção de uma anomalias. São apresentados métodos onde, dado um conjunto de transações suspeitas, pretende-se encontrar outras pessoas que tenham fortes conexões com todos ou a maioria dos suspeitos existentes. Os pontos marcados com um círculo vermelho na Figura 2.4, sinalizam os relacionamentos existentes entre as vítimas e os suspeitos de ocasionarem as fraudes.

As técnicas de VA apresentadas nos estudos da Figura 2.4 permitem identificar os seguintes aspectos:

1. **Representatividade poderosa**: Representações gráficas facilitam a apresentação de conjuntos de dados, permitindo a incorporação de atributos/tipos de nós e arestas.

2. **Natureza relacional dos domínios de problema**: A natureza das anomalias são relacionais. No domínio da fraudes, por exemplo, pode-se aventar dois tipos de cenários: (1) fraudes oportunistas (se alguém comete fraude, é provável que seus

conhecidos também o façam), e (2) fraudes organizadas realizadas pela colaboração de um grupo.

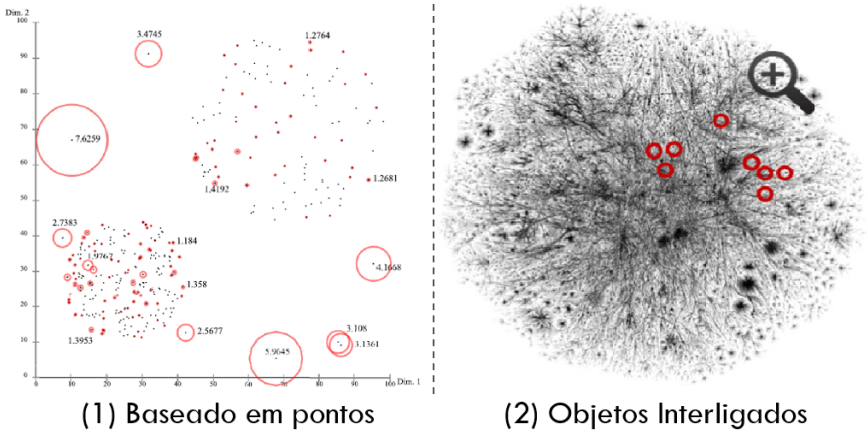

Figura 2.4: Identificação de *outliers* através de pontos e interdependência dos dados

Um dos algoritmos citados por [7] [14] [15], e utilizado neste trabalho para a identificação de *outliers* é o de cálculo de distância (Di) de Cook [17], que indica o grau de distorção que uma determinada transação gera no modelo. Estes autores apresentam como esse modelo pode contribuir para a detecção de fraudes identificando pontos que influenciam negativamente o modelo de regressão de mínimos quadrados.

Este algoritmo [17] identifica as transações que distorcem o resultado e a precisão do modelo de regressão de mínimos quadrados, através da remoção de través da remoção de Y_i (atributos de uma determinada transação, baseado no *timestamp*) do conjunto de observações. Trata-se de uma técnica de análise de resíduos que permite identificar *outliers* influentes em um conjunto de variáveis preditoras.

Tecnicamente, o Di de Cook é calculado removendo-se a i-ésima observação do modelo e recalculo a regressão. Ele resume o quanto todos os valores no modelo de regressão mudam quando a observação é removida. Assim, o autor [17] define os seguintes passos para cálculo da **Distância de Cook**:

(1) Identificação das variáveis resposta (X_i), e cálculo de Y_i por meio da **regressão linear múltipla** de cada transação existente no dataset, sendo n o número de atributos;

$$Y_i = \beta_0 + \sum_{i=1}^{n} X_i \beta_i \tag{2.2}$$

(2) Obter o estimador $\widehat{Y}_i$ removendo-se a i-ésima observação/transação (Y_i) do modelo e recalculando a regressão;

(3) Cálculo de e_i, como a diferença entre os resíduos de Y_i observado e $\widehat{Y}_i$ estimado;

$$e_i = Y_i - \widehat{Y}_i \tag{2.3}$$

(4) Obtem-se o erro médio quadrado σ^2 do modelo de regressão; e
(5) Calcula-se Di de Cook por meio da fórmula, considerando $\boldsymbol{P}$ como o número de variáveis reposta utilizada na regressão:

$$D_i = \frac{\sum_{j=1}^{n}(\widehat{Y}_j - \widehat{Y}_{j(i)})^2}{P\sigma^2} \tag{2.4}$$

, ou representação equivalente expressando o uso da Matriz H, onde $\boldsymbol{h}$ii deriva diretamente do cálculo dos valores ajustados por meio da matriz chapéu, em que $\widehat{Y} = \mathbf{Hy}$; ou seja, a alavancagem é um elemento da diagonal da Matriz H. Esta derivada parcial descreve o grau pelo qual o i-ésimo valor medido influencia o i-ésimo valor ajustado.

$$D_i = \frac{e_i^2}{P\sigma^2}\left[\frac{h_{ii}}{(1-h_{ii})^2}\right] \tag{2.5}$$

, onde

$$h_{ii} = \frac{\partial \widehat{Y}_i}{\partial Y_i} \tag{2.6}$$

Wouter et al. [18] apresenta a área sob a curva (AUC) como medida de desempenho, onde o autor [12] compara a *performance* de diversos algoritmos através desta métrica.

Fei et al. [19] traz o uso do *Isolation Forest* no contexto de detecção de fraudes, que é um algoritmo de aprendizagem não supervisionado que trabalha com o princípio de isolar anomalias. Para isolar um ponto de dados, no *Isolation Forest*, o algoritmo gera recursivamente partições na amostra selecionando aleatoriamente um atributo e, em seguida, selecionando aleatoriamente um valor dividido para o atributo, entre os valores mínimo e máximo permitidos para aquele atributo.

Do ponto de vista matemático, o particionamento recursivo pode ser representado por uma estrutura de árvore chamada *Isolation Tree*, enquanto o número de partições necessárias para isolar um ponto pode ser interpretado como o comprimento do caminho, dentro da árvore, para chegar a um nó de terminação. Por exemplo, o comprimento do caminho do ponto X_i é maior do que o comprimento do caminho de X_j na Figura 2.5.

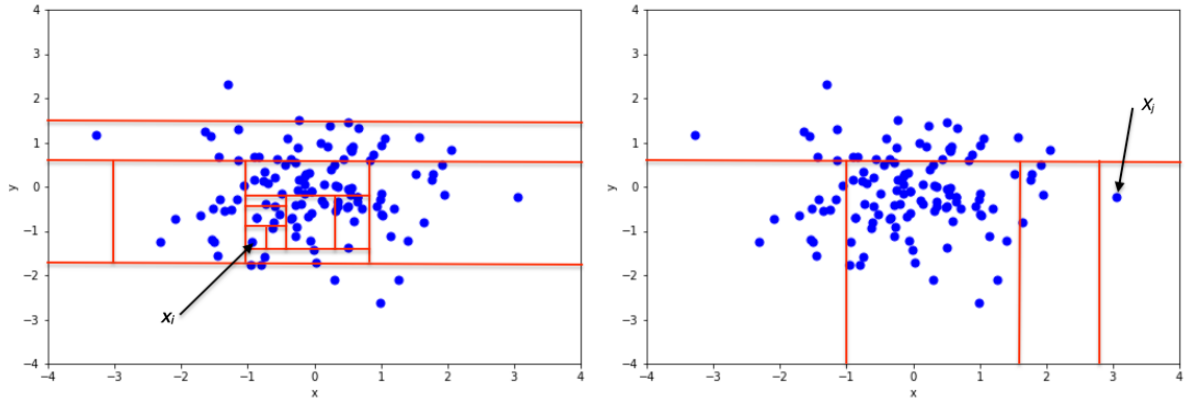

Figura 2.5: Isolamento de pontos X_i (não anômalo) e X_j (anômalo) - Gaussiana 2D

Kuhn Max [20] traz o algoritmo do *One-Class SVM*, empregado em cenários onde há um desequilíbrio acentuado em conjuntos de dados com apenas uma classe (Não Fraude). O SVM constrói um hiperplano, em um espaço dimensional alto ou infinito, que pode ser usado para identificação de *outliers*. Uma boa separação é conseguida pelo hiperplano que possui a maior distância aos pontos de dados de treinamento mais próximos de qualquer classe (a chamada margem funcional), pois em geral quanto maior a margem, menor o erro de generalização do classificador.

Este algoritmo se baseia na estimativa da densidade dos pontos de dados. Este método assume que a distribuição segue os modelos Gaussiano e de Poisson. Em seguida, os testes de discordância podem ser usados para testar as novas transações. Esses métodos são robustos e permitem escalar as variâncias causados pelos desbalanceamentos.

A Figura 2.6 representa as fases para a condução dos experimentos baseado no modelo de referência *CRoss Industry Standard Process for Data Mining* (CRISP-DM [21]). No capítulo 3 são detalhadas as fases do *framework* que compõem o SDF, a arquitetura proposta e *dashboard* do VA, o detalhamento dos experimentos, e o processo de validação do resultado dos algoritmos com especialistas.

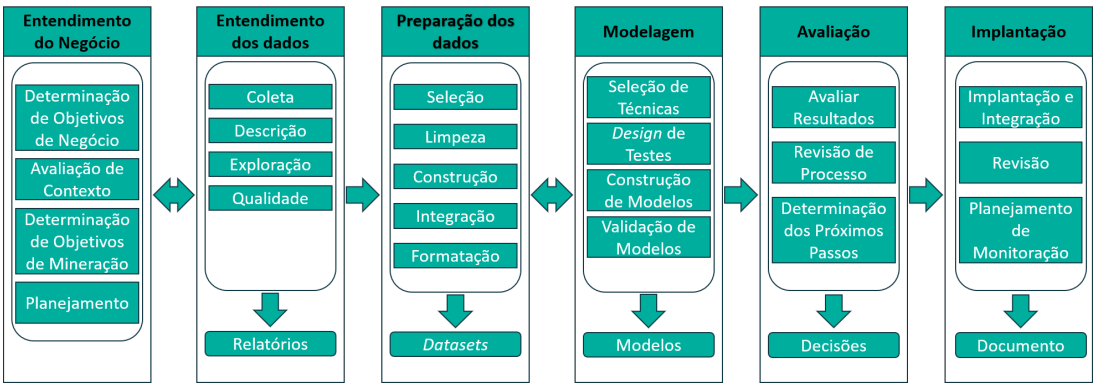

Figura 2.6: Fases da mineração de dados com base no modelo de referência CRISP-DM.

Capítulo 3

Metodologia Adotada

A metodologia experimental utilizada neste projeto compreendeu as seguintes etapas:

1. Definição do Problema;
2. Revisão da literatura e trabalhos relacionados através do Web of Science, Scopus e Google Acadêmico. Adoção da TEMAC (Teoria do Enfoque Meta Analítico Consolidado) para identificação dos artigos de maior impacto em relação ao tema;
3. Adotado o CRISP-DM [21] como modelo de referência;
4. Arquitetura Proposta utilizando VA;
5. Realização de Experimentos;
6. Avaliação da precisão dos algoritmos com especialistas;
7. Criação de *Dashboard* dos algoritmos selecionados através do VA.

3.1 Definição do Problema

Neste projeto foi utilizado um conjunto de dados anonimizado do Banco X. O conjunto de dados utilizado contém as transações de pagamentos de boletos, cheques e transferências entre contas, executadas ou recebidas pelos clientes durante os meses de janeiro a julho de 2020. O conjunto de dados, cujos atributos foram detalhados na Tabela 3.1, possui uma natureza multivariada e de tipagem diversa, como: variáveis categóricas, numéricas, contínuas e séries temporais. Os aspectos multivariados são um indicativo da complexidade [11] da natureza de dados, que requerem uma combinação sofisticada de diferentes técnicas para serem abordadas.

Por meio de um processo de ETL (*Extract, Transform and Load*), foi realizada a extração de um conjunto de dados contendo 14.021.371 registros de transações bancárias

de 1.388 clientes, que sofreram algum tipo de fraude entre janeiro e julho de 2020. Este conjunto de dados tem 2 classes de transações: **legítima** e **fraudulenta**. O total de fraudes registradas de janeiro a julho de 2020 (1.401), representam 0,0001% da amostra.

A distribuição esparsa e desequilibrada do conjunto de dados provou ser um grande desafio a ser resolvido. Assim, decidiu-se trabalhar sob a hipótese de atributos independentes e não correlacionados com base nos estudos Carminati et al. [11]. Adicionalmente, West et al. [3] afirmam que é possível identificar técnicas mais apropriadas para a detecção de fraudes, ao dividir os dados pelo tipo de transação (pagamentos, transferências de contas, etc.). Essa aproximação permite uma visualização e interpretação muito mais fácil de modelos e resultados, além de uma complexidade temporal e espacial reduzida. A janela de dados de 6 meses é atualizada uma vez por semana. A conta de cada cliente é tratada separadamente para evitar influências de transações de uma para outra, o que permite identificar padrões de comportamento para cada conta.

Utilizou-se análise preditiva, mineração de dados e *Machine Learning* (ML) para identificar *outliers* de transações para o conjunto de dados de cada cliente. Esses processos mitigam o sub-treinamento devido à falta de dados históricos, construindo perfis bem treinados e calibrando modelos para a evolução do comportamento financeiro dos usuários no futuro. Para esta calibração futura, desliza-se a janela de dados de 6 meses ao longo do tempo, analisando novas transações. Foram realizadas análises de comportamento confrontando transações com as informações contidas no mecanismo de negócios que segue normas e regras do Banco Central do Brasil.

No processo de transformação dos dados, estes foram preparados para tratar *time stamps*, realizar conversões de *strings*, normalizar dados (distribuição gaussiana) e para seleção de recursos que têm pouca correlação usando PCA (Análise de Componentes Principais) e Eliminação de Recursos Recursivos (RFE). Esses métodos permitem extrair variáveis importantes de um conjunto de dados altamente dimensional. Com menos variáveis, a visualização se torna mais significativa.

Especialistas sugeriram implementar filtros de regras de negócios que reduzam falsos positivos. Além disso, exclui-se transação abaixo de R$ 100 reais ($\approx$ US$ 20) uma vez que o custo envolvido na avaliação desses casos é maior do que o valor da fraude, e não há registros de fraudes furtivas envolvendo valores abaixo desse valor. Para evitar distorções no modelo do cliente, cria-se um atributo onde especialistas podem definir transações confirmadas como fraudulentas, de modo que nosso modelo a ignora durante a análise. Finalmente, implementou-se uma regra para sinalizar sequência de transações do mesmo cliente, destino, tipo e categoria com as diferenças entre os *time stamps* inferiores a 30 minutos. Transações com as mesmas características que ocorrem muito perto no tempo não são usuais e o fraudador tende a seguir esse padrão com a intenção de

maximizar seus lucros, especialmente se eles tiverem acesso à conta do cliente. Por fim, criou-se um atributo onde os especialistas podem "marcar" transações confirmadas como fraudulentas para que os modelos não as considerem durante a detecção de anomalias, evitando distorções no modelo de comportamento das transações dos clientes.

3.2 Entendimento dos Dados

Nessa seção serão tratadas atividades de coleta e exploração dos dados, visando obter os primeiros *insights*. Existe uma ligação importante entre as fases de entendimento do negócio e de entendimento dos dados. Este projeto foi guiado por uma análise aprofundada de um conjunto de dados do mundo real, que é primordial para se melhorar o processo de detecção de fraudes existente, e para fornecer *insights* para a tomada de decisões.

Os dados utilizados foram extraídos da base de conta corrente, de cadastros de clientes e de lançamentos de transações do Banco X, coletados entre janeiro de 2020 e julho de 2020. O conjunto de dados foi anonimizados removendo atributos que permitissem a identificação dos clientes, e mantendo aqueles que pudessem garantir que as análises fossem realizadas utilizando valores reais de movimentações bancárias.

A Tabela 3.1 resume os atributos (descaracterizados) dos cinco principais tipos de transações. Cada cliente possui até cinco perfis de comportamento, de acordo com o tipo de movimentação financeira (Boletos de Cobrança, Cheques, DOC, TEC e TED).

Esses dados podem ser categóricos, numéricos, contínuos ou temporais. Os aspectos multivariados são um indicativo de natureza complexa de dados [1], que requerem uma combinação sofisticada de diferentes técnicas para serem abordadas. Os dados foram preparados para tratar *time stamps*, conversões de *strings* e normalização de dados (distribuição gaussiana).

Para a solução SDF foram utilizados diferentes *datasets*: a base de transações históricas, contendo os *scores* calculados para os comportamentos de crédito e débito, conforme o tipo de movimentação de cada cliente; a base contendo as transações diárias *online*; e a base de registros das sinalizações de possíveis fraudes, que contém o resultado da avaliação do grau de influência que a nova transação causa na regressão linear histórica e demais desvios nos resíduos calculados para o conjunto de transações analisadas.

O atributo **FRAUDES** da Tabela 3.1 possui um papel fundamental no processo de sinalização das transações classificadas como suspeitas. Aquelas que se enquadrem nas regras definidas na etapa B da Figura 3.2, referente à arquitetura do SDF, recebem uma *flag* indicando que seus dados devem ser exibidos nos painéis do VA.

O Banco X classifica as fraudes como sendo dos seguintes tipos:

Tabela 3.1: Atributos e Descrição dos campos do *dataset*

Atributos	**Descrição do Campo**
DATA_TRANS	data e *time stamps*
ID_TRANS	identificação única da transação
COD_TRANS	código para identificar tipos de transações
TIPO_TRANS	descrição do tipo da transação (DOC, TED, TEC, BOLETO)
TIPO_CLT	define se é uma transação de débito ou crédito
VALOR_TRANS	valor monetário das transações
CENTRAL	número da central
COD_COOP	número da cooperativa
COD_PAC	número do PAC onde foi realizada a transação
RISCO_PAC	grau de risco do terminal utilizado na transação
CONTA	número da conta
BANCO_ENVOLVIDO	código do banco envolvido
AGENCIA_ENVOLVIDA	número da agência envolvida
CONTA_ENVOLVIDA	número da conta envolvida na transferência ou pagamento
FAV_PF_PJ	número do CPF ou CNPJ do titular da favorecido
NOME_CLIENTE	nome do titular da conta
RISCO	grau de risco do titular da conta
TIT_PF_PJ	número do CPF ou CNPJ do titular da conta
RENDA	valor declarado da renda do cliente
DATA_NASC	data de nascimento do cliente
UF	cidade do endereço do cliente
DIF_TRANS	identificação do intervalo de tempo entre sucessivas transações
LIMITE_OPERACAO	registra suspeitas com base nas regras
FRAUDES	registro das sinalizações de fraudes
TIPO_FRAUDE	classificação do tipo da fraude
MODALIDADE	modalidades de atuação utilizadas pelos fraudadores

Documental: casos em que há o uso de documentos falsos para abertura de contas, emissão de boletos ou criação de contratos com o intuito de realizar movimentações indevidas ou obtenção de empréstimos;

Eletrônica: são aquelas relacionadas aos acessos indevidos ocorridos em canais como Internet Banking, ATM ou Mobile, estando a vítima coagida ou não;

Golpe e Estelionato: estão relacionados à golpes envolvendo processos de engenharia social e outros meios, que levam a vítima a realizar transferências para contas bancárias do criminoso e pagamentos de contas utilizando cheques, boletos falsos e cartões; e

Pulverização e Lavagem de Dinheiro: trata-se de formas de distribuir, legalizar e sacar os recursos. São utilizadas técnicas para evitar a declaração, rastreamento e recolhimento de impostos.

Correlacionando as modalidades de atuação utilizadas pelos fraudadores, com as classificações dos tipos de fraudes, indicadas pelos especialistas da área de prevenção e combate à fraudes, é possível entender as formas de atuação dos fraudadores, conforme ilustrado na Figura 3.1.

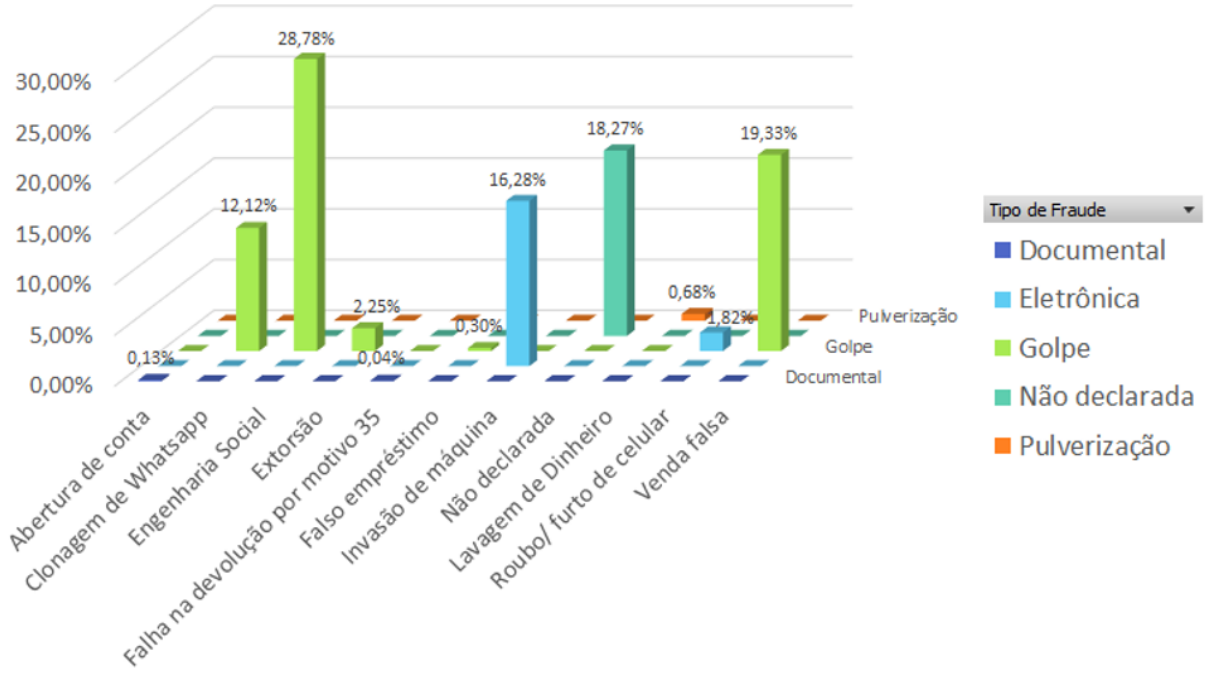

Figura 3.1: Percentual de ocorrências conforme modalidades e tipo das fraudes.

Através da Figura 3.1 observa-se que os golpes aplicados por meio de engenharia social possuem o maior percentual de ocorrência de fraudes (28,78%). Os golpes são aplicados através de diversas modalidades, gerando os maiores impactos e respondendo por 62,48% dos casos. As fraudes eletrônicas, provenientes de invasões de máquinas e roubo de aparelhos de celular, representam juntos 18,62% do total.

Este entendimento, nos permitiu identificar os comportamentos mais comuns dos fraudadores ao aplicarem os diversos tipos de fraudes, embasando a criação das regras gerais dos modelos utilizados para gerar os perfis de análise de comportamento.

3.3 Arquitetura do SDF utilizando VA

O Sistema de Detecção de Fraudes (SDF) do Banco X é composto por um *framework* que possui modelo auto adaptativo baseado em perfil, projetado para detecção de anomalias, que computa transações financeiras individuais de clientes criadas com base em transações históricas. Conforme descrito na Figura 3.1, o SDF utiliza um processo de ETL para extração dos registros de transações bancárias.

Para tratar a distribuição esparsa e desequilibrada do conjunto de dados, o SDF considera que os atributos do *dataset* são independentes e não correlacionados. Os dados dos clientes são divididos pelo tipo de transação (pagamentos, transferências de contas etc.). O sistema utiliza uma janela delizante de 6 meses de dados, que é atualizada uma vez por semana. A conta de cada cliente é tratada individualmente de forma a identificar padrões de comportamento para cada conta.

Para calibração do modelo, desliza-se a janela de dados de 6 meses ao longo do tempo, analisando novas transações. As novas contas iniciam com limites de créditos e parâmetros pré definidos contidas no mecanismo de negócios que seguem normas e regras do Banco Central do Brasil. A janela deslizando permite que o modelo vá se ajustando conforme o decorrer das movimentações do cliente.

Este sistema foi inspirado nos modelos de Carminati et al. [11] e Roger Leite et al. [13], e como resultado é capaz de:

1. Abordagem híbrida: uso de algoritmos não-supervisionados de detecção de *outliers*; análise de perfil de comportamento e de movimentações financeiras; e integração da análise humana ao processo de investigação das transações financeiras suspeitas, com o intuito de reduzir a incidência de falsos positivos;

2. Análise *online* das novas transações, identificando-se o quanto que elas destoam dos comportamentos padrões de movimentação financeira dos clientes;

3. Uso de técnicas de VA como solução para sinalizar e identificar as transações suspeitas, de forma gráfica; e

4. Permitir que os funcionários atuem sem gerar atrasos nas transações *online*, legítimas.

A Figura 3.2 apresenta o fluxo do sistema de processamento de transações do SDF. As seis primeiras colunas representam sistemas que já estão em produção diária. A última coluna representa o SDF que foi estruturado nas quatro fases que seguem:

Geração de perfil com base no comportamento das transações do cliente. Para cada conta do cliente, o SDF gera perfis [11] com base no histórico de transações da conta (A, Figura 3.2). A geração de perfil é um processo de lote não vinculado às outras fases. Ele constrói as informações com uma janela deslizante de dados de 6 meses, e é atualizado com um atraso de uma semana.

Análise de Transações *online*. Representa a análise em tempo real para detecção de fraudes, levando em conta transações históricas e *online* (B, Figura 3.2). Os resíduos de regressão são calculados buscando identificar o quão negativamente a nova transação influencia o modelo. É verificado se a nova transação se distancia dos *clusters*, representando um *outlier*. Finalmente, validou-se as regras de negócios relacionadas ao comportamento considerando o tipo conhecido de fraudes (por exemplo: destino, tipo, categoria e carimbo de tempo inferior a 30 minutos).

Sinalizar Transações *online* Suspeitas. Transações que afetam negativamente a regressão ou violam regras de negócios e comportamentos, são sinalizadas como possíveis fraudes. Dashboards usados por especialistas (C, Figura 3.2).

Avaliação e Intervenção Humana. Especialistas da Área de Prevenção e Combate a Fraudes exploram dados de transações com sinalização suspeita, aplicando filtros em painéis VA. Os especialistas podem rejeitar ou aprovar uma transação suspeita através do controle (ALC/CTR) e camadas de autorização (D, Figura 3.2). Um atraso, em transações suspeitas (por exemplo: 30 minutos), é aplicado para permitir que os especialistas executem análises mais profundas.

A Figura 3.2 apresenta a estrutura SDF integrada ao sistema de processamento transacional atual do Banco X. Cada camada representa sistemas existentes que já estão em produção diária, onde: **(1) CTE:** O Controle de Execução é responsável por fazer chamadas para outras camadas; **(2) AUD:** Registra todas as transações para fins de auditoria; **(3) CTA:** Autenticar usuários/clientes; **(4) CTR:** Controle de Transação (desempenho da equipe hoje); **(5) CLO:** Controle dos limites operacionais (regras de negócios); **(6) ALC:** Controle de elevação/autorizações; **(7) SDF:** Sistema de detecção de fraudes. A última coluna representa o SDF, integrado ao sistema de produção diária.

A Figura 4.1 no Capítulo 4 ilustra a representação de cada camada no contexto da arquitetura construída durante a execução dos experimentos.

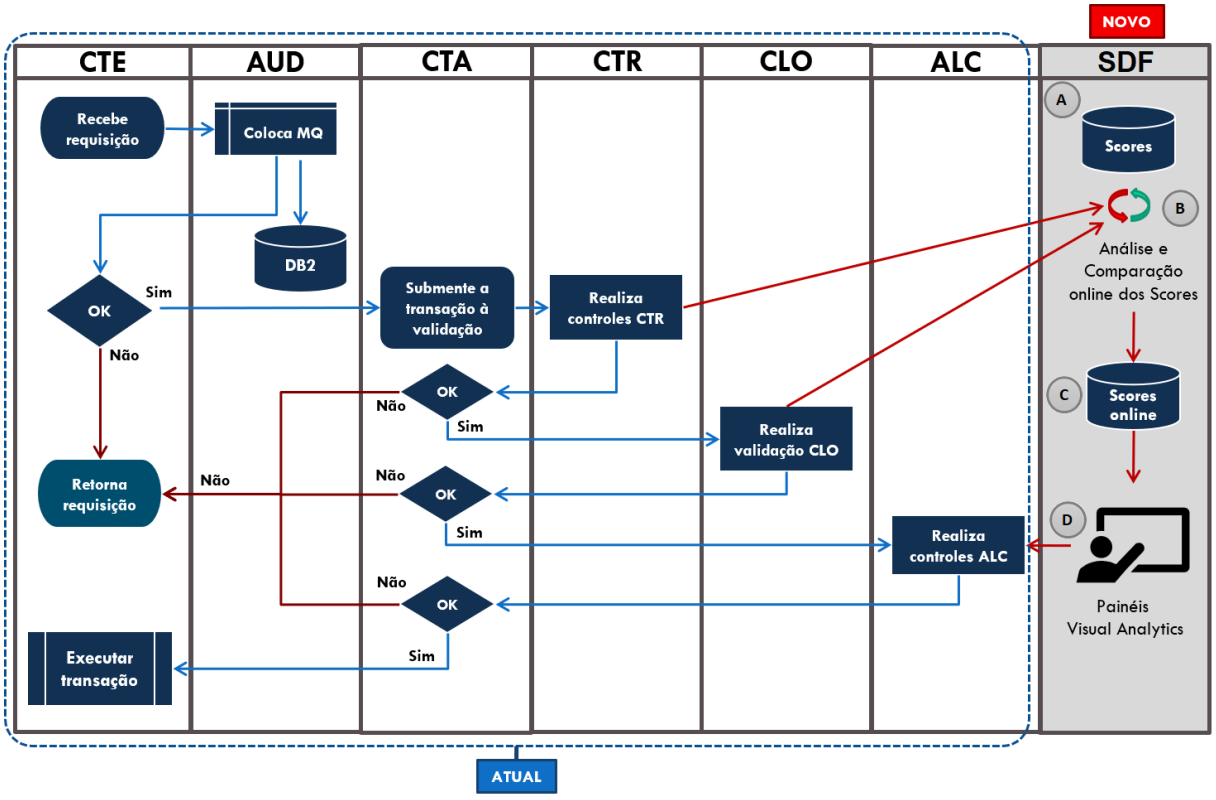

Figura 3.2: Fluxo do sistema de processamento de transações.

Capítulo 4

Experimentos Realizados

Para implementar o SDF em uma estrutura de detecção de fraudes totalmente integrada e em tempo real, levou-se em consideração as definições de arquitetura existentes do Banco X. A premissa foi utilizar soluções já adquiridas e com contratos de suporte ativo na instituição.

4.1 Visão Geral da Arquitetura

A letra A da Figura 4.1 representa o processo de ETL onde foi utilizado a solução IBM Data Stage para carregar as informações de histórico das transações para a área de *stage*, que conta com um *cluster* de 4 servidores, que rodam em cima do Suze ZLinux *Enterprise* V. 12, com 16 núcleos, e 32 GB de RAM, cada.

A letra B da Figura 4.1 representa o processo de modelagem, avaliação de intervalo de transações, análise de padrões de movimentação das contas dos clientes e aplicação das regras de negócios, gerando os perfis para os clientes para tratamento e segmentação dos dados por tipo de transação, cliente e janela deslizante dos últimos 6 meses. Neste processo foi utilizado o SAS *Enterprise Guide* 7.1 (64 bits), que foi instalado em 2 servidores de *cluster* com 32 núcleos e 96 GB de RAM.

Os dados dos perfis são utilizados para a análise conjunta com os dados das transações *online*, avaliando o quão negativamente estas afetam o modelo de regressão. Além disso, são executados algoritmos de identificação de *outliers*, utilizando o PySpark Streaming e o Kafka. Estes dois componentes foram instalados em 4 servidores Red Hat Enterprise Linux 64 bits, com 16 núcleos e 64 GB de RAM, cada. O SAS VA, representado pela letra C da Figura 4.1 é executado em um único servidor com 32 núcleos e 96 GB de RAM. A letra D da Figura 4.1 representa a atuação dos especialistas através do ALC e CTR. No total, foram alocados 224 núcleos e 672 GB de RAM para este ambiente.

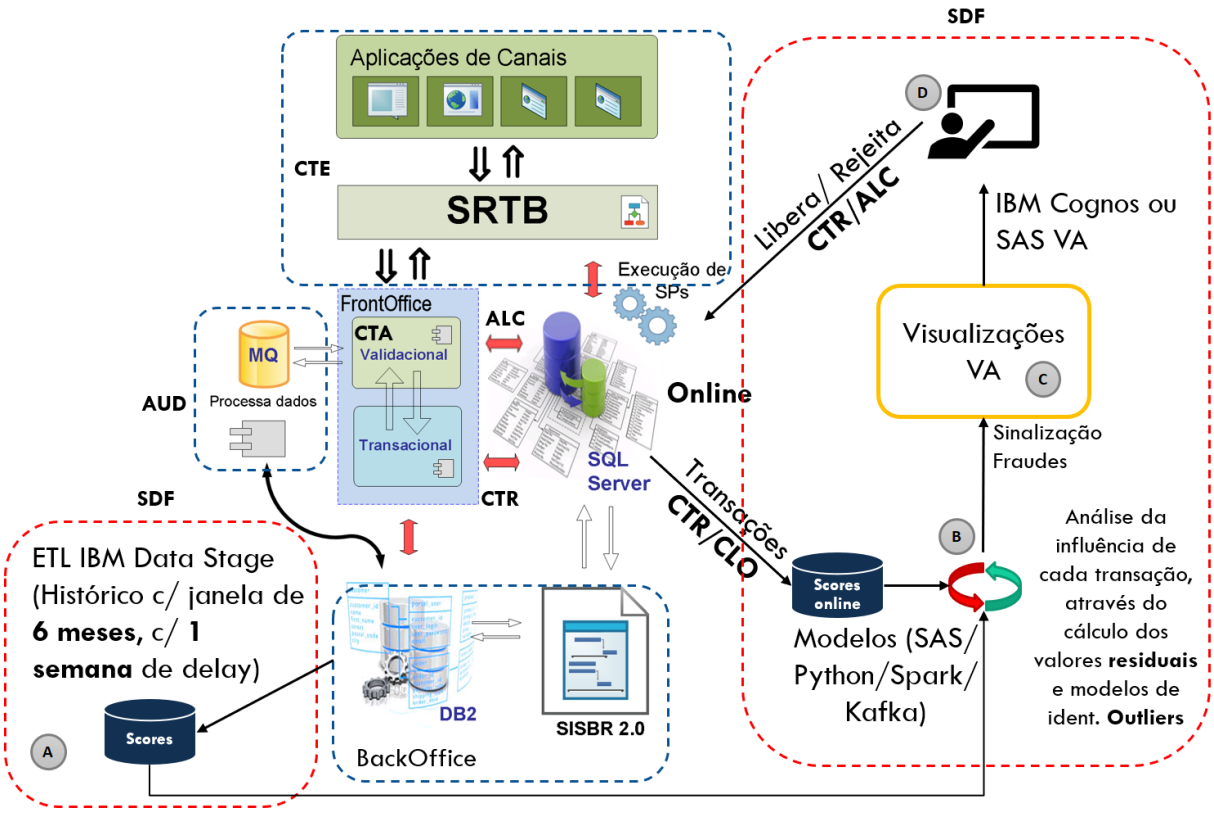

Figura 4.1: Visão geral da arquitetura implementada para os experimentos do SDF.

4.2 Detalhamento dos Experimentos

O SDF possui uma abordagem para detecção de *outliers*, que é comparável com Carminati et al. [11], que usa técnicas não supervisionados para fornecer diferentes tipos de análise estatística, e classificar corretamente transações como suspeitas. Para realizar os experimentos, utilizou-se um conjunto de dados contendo 14.021.371 registros de 1.388 clientes que sofreram evento fraudulento, entre janeiro e julho de 2020. Em linha com o descrito no Capítulo 1, os dados dos clientes possuem informações sensíveis e não podem ser fornecidos em conformidade com as regulações, leis e políticas de segurança vigentes.

Conforme descrito na Seção 3 o conjunto de dados tem 2 classes de transações: legítima e fraudulenta. No intuito de comparar o desempenho dos algoritmos supervisionados e não-supervisionados, fez-se necessário tratar a distribuição esparsa e desequilibrada do conjunto de dados, onde considerou-se a independência entre os atributos. Para tanto, aplicou-se as técnicas descritas por West et al. [3], dividindo o grande conjunto de dados em *datasets* menores, conforme o tipo da transação (Boletos de Pagamento, Cheques, DOC, TEC e TED) e número da conta.

Em seguida, realizou-se uma análise exploratória aprofundada e multidimensional, correlacionando as modalidades de atuação utilizadas pelos fraudadores, com as classificações

dos tipos de fraudes, indicadas pelos especialistas da área de prevenção e combate à fraudes. Isso, permitiu entender as formas de atuação dos fraudadores, conforme ilustrado na Figura 3.1 da Seção 3.2.

O entendimento de que os golpes correspondem a 62,48% dos casos e que o uso de engenharia social é o método mais utilizado (28,78%), a frente, inclusive, das fraudes cibernéticas, permitiu a criação das regras gerais destinadas à geração dos perfis de análise de comportamento, e a divisão mais adequada dos dados com base na conta de usuário e tipo de transação, aumentando a previsão e mitigando o problema a distribuição esparsa e desequilibrada dos dados.

Sobre o conjunto de dados gerados, foram avaliados 8 algoritmos (supervisionados e não supervisionados) para realizar a comparação de precisão, com base no referencial teórico utilizado neste trabalho, a saber: Distância de Cook [17]; *Extra Trees* [3] [4]; HBOS [12]; *Isolation Forest* (iForest) [19]; *Random Forest* [22]; *R-Student* [3] [4]; *One-class Support Vector Machine* (SVM) [20]; e XBOS [12].

Os algoritmos foram configurados da seguinte forma: Para a Distância de Cook e *R-Student*, utilizou-se o valor da transação como variável de resposta; Uma transação é considerada *outlier* quando a Distância de Cook é superior a 3 vezes a média (μ) de todas as outras transações; definiu-se parâmetros do *Random Forest* para **contaminação**='auto' e **behavior**='*new*'; Árvores extras e floresta aleatória tem **n_estimators**=tamanho do conjunto de dados do cliente, **max_depth**=7 e **random_state**=0; O SVM de uma classe foi definido com **gamma**='auto' e **nu**=0.01; e uma vez que o XBOS usa *clustering* k-means e temos duas classes, definiu-se o número de *clusters* para 2.

Como cada dado do cliente foi dividido por tipo de transação e encontrou-se conjuntos de dados com menos de 10 registros. Neste cenário, alguns algoritmos perderam a precisão, ou não puderam ser executados, ao se quebrar esses conjuntos de dados em treinamento e testes. Assim, usou-se a validação cruzada, que é mais confiável e oferece maior precisão do que a técnica de dividir os dados em treinamento/teste, onde cada dobra é usada em treinamento repetidamente. Assim, depois de executar o processo em dobras k-1, resumiu-se o desempenho usando a média e o desvio padrão.

Após o treinamento de todos os modelos, utilizou-se a curva Característica de Operação do Receptor (ROC) para estabelecer um limite adequado para classificar as transações como fraude ou não fraude, devido à imensa desproporção entre as classes. Fazendo isso, pôde-se gerar a matriz de confusão para cada resultado obtido.

O gráfico ROC é baseado em Taxa de Verdadeiro Positivo e Taxa de Falso Positivo. Para construir o gráfico ROC, o FPR (*False Positive Rate*) é plotado no eixo x e no TPR (*True Positive Rate*, ou *Sensitivity*, ou *recall*) no eixo y.

Sem perda de generalidade, chamaremos as classes de Fraude e Não Fraude. Assim,

o TPR mede a taxa de fraudes corretamente identificada. Por outro lado, a FPR mede a taxa de não fraudes incorretamente identificadas como fraudes.

Uma forma de apresentar estatísticas para avaliar um modelo de classificação é cruzar a classe prevista pelo modelo, e a classe real da amostra, conforme citado por Prati et al. [23]. A Figura 4.2 representa a organização da matriz de confusão citada pelo autor.

		Classes Atuais	
		Fraude	**Não Fraude**
Classes Preditas	**Fraude**	Verdadeiro Positivo	Falso Positivo
	Não Fraude	Falso Negativo	Verdadeiro Negativo

Figura 4.2: Matriz de Confusão.

A matriz de confusão é uma tabela com duas linhas e duas colunas que relata o número de falsos positivos, falsos negativos, verdadeiros positivos e verdadeiros negativos. Quando uma transação fraudulenta é classificada como positiva, ela é chamada de TP (*True Positive*). Quando uma transação normal é classificada como positiva, ela é chamada de FP (*False Positive*). O mesmo ocorre na situação oposta em que a FN é uma Falsa Negativa, e a TN corresponde a um Verdadeiro Negativo.

Para avaliar os modelos, derivou-se medições da matriz de confusão, reduzindo suas quatro células principais a um único índice numérico de qualidade. Para isso, foram comparadas as matrizes com os registros de fraude. Nesta fase, especialistas da área de prevenção e combate a fraudes foram envolvidos. Eles contribuíram para confirmar TPs - em caso de correspondências de classificação de fraude para registros de fraude - e TNs - quando modelos e especialistas concordam que o evento não é uma fraude. Essa informação tornou-se uma nova dimensão nos conjuntos de dados.

Utilizou-se o *F1-Measure*, que é uma média harmônica de precisão e revocação. A precisão é a fração das instâncias relevantes entre as instâncias recuperadas, e a revocação mede a quantidade total de instâncias relevantes que foram recuperadas.

Essa métrica é melhor do que o cálculo de precisão porque a quantidade de acessos para uma transação legítima é muito maior do que a indicação correta de que a transação é uma fraude, dado o desequilíbrio entre essas duas classes.

Em muitos casos, esse desequilíbrio é tão acentuado que encontramos muitos conjuntos de dados com apenas uma classe (Não Fraude). Para analisar esses casos, utilizou-se o *One-Class SVM*.

O exemplo da Tabela 4.1 ilustra a aplicação do algoritmo de Cook [17] para calcular Di (Equação 2.4) em duas transações distintas, identificando pontos que influenciam negativamente o modelo de regressão de mínimos quadrados, através da remoção de

Tabela 4.1: Exemplificação dos valores para cálculo da distância Di de Cook.

# Obs.	**HAT**	***S***	***ei***	***ei*^2**	***S*^2**	***h*ii/(1-*h*ii)^2**	**Di**
204	0,041	88966,30	326.600,32	1,067e+11	3,245e+10	0,044	0,15
209	0,024	89077,26	60.933,09	3,713e+09	3,253e+10	0,025	0,003

Yi (atributos de uma determinada transação, baseado no *timestamp*) do conjunto de observações. Trata-se de uma técnica de análise de resíduos que permite identificar *outliers* influentes em um conjunto de variáveis preditoras.

Um alto valor de Di indica que a transação gera elevados valores de resíduo ***e***i, que distorcem o resultado e a precisão da regressão linear, sendo que a observação que gera maior distorção, é considerada um possível *outlier*.

Abaixo seguem algumas interpretações utilizadas no modelo do protótipo:

- As observações que obtiverem um Di superior a 3 vezes a média dos Di's calculados para o conjunto de transações, são consideradas como um possível *outlier*;
- Quanto maior o valor de ***e***i mais longe o estimador $\widehat{Y}$i estará do valor ajustado; e
- O *leverage* de ***h***ii é a métrica que indica o quão longe está o valor da transação, de uma variável dependente, em relação às demais.

As fórmulas de cálculo do Di de Cook estão descritos no capítulo 2.

Para a validação dos experimentos foram filtradas 213 transações, no período entre janeiro e julho de 2020, que possuem valores acima de R$ 50.000,00. A Tabela 4.1 ilustra os valores das variáveis para cálculo de Di para duas transações: 204 e 209. A Figura 4.3 ilustra os resultados obtidos para a transação (204), classificada como *outlier* por possuir um *score* Di elevado, e a transação (209), que possui um *score* Di baixo.

Os pontos em vermelho da Figura 4.3 representam as transações consideradas *outliers*, e sinalizadas como possíveis fraudes, utilizando o cálculo da distância de Cook. O modelo propõe que os *outliers* equivalem a (3x) o valor da média de Di. Neste experimento, a média calculada de Di foi de $\mu = 0{,}004590$. Neste sentido, as transações são sinalizadas como suspeitas de fraudes quando o valor de Di $> 0{,}01377$.

O *Isolation Forest* é um algoritmo de detecção de anomalias baseado em árvores, que são estruturas criadas para isolar anomalias. O resultado é que exemplos isolados têm uma profundidade relativamente curta nas árvores, enquanto os dados normais são menos isolados e têm maior profundidade. O *One-Class SVM* teve um dos melhores desempenhos sendo mais eficaz para conjuntos de dados de classificação desequilibrados onde não há nenhum ou muito poucos exemplos da classe minoritária, ou em conjuntos de dados onde não há estrutura coerente para separar as classes que poderiam ser aprendidas por um algoritmo supervisionado.

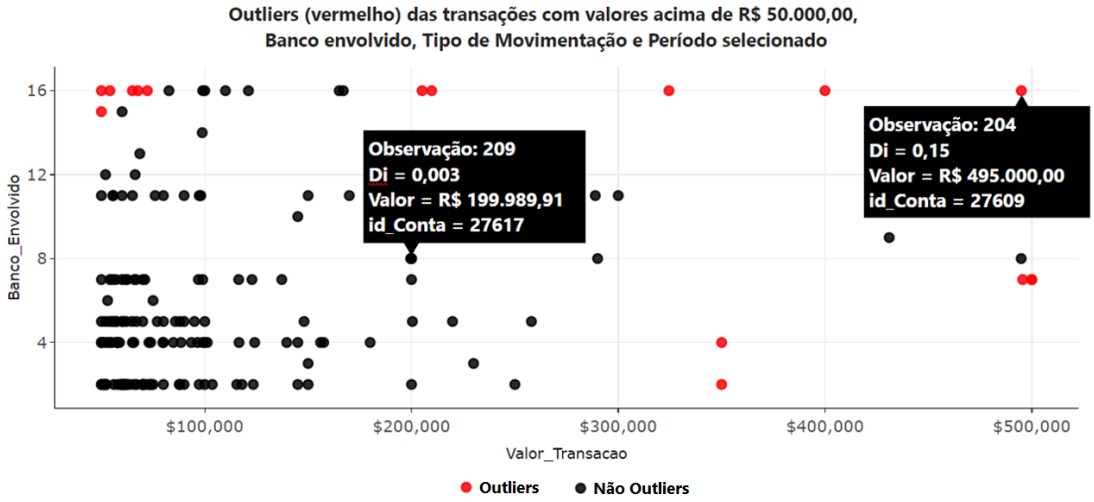

Figura 4.3: Identificação de *Outliers*.

XBOS é um algoritmo de detecção de anomalias baseado em *cluster*. Ele usa *k-means* como um algoritmo de *clustering*. No XBOS, um pequeno aglomerado perto de um *cluster* maior é tratado como se fosse um *cluster* médio, de modo que os pontos de dados são classificados como "não tão anômalos".

A Figura 4.3 apresenta um comparativo do desempenho dos algoritmos XBOS, HBOS e *Isolation Forest*. O XBOS é baseado em cluster; o HBOS é baseado em histograma; e o *Isolation Forest* é baseado no isolamento das anomalias. O *Isolation Forest* apresentou 3 falsos positivos e 4 falsos negativos. O HBOS teve um desempenho um pouco melhor apresentando 1 falso positivo e 1 falso negativo. O XBOS teve o melhor desempenho identificando corretamente todas as transações legítimas e fraudulentas.

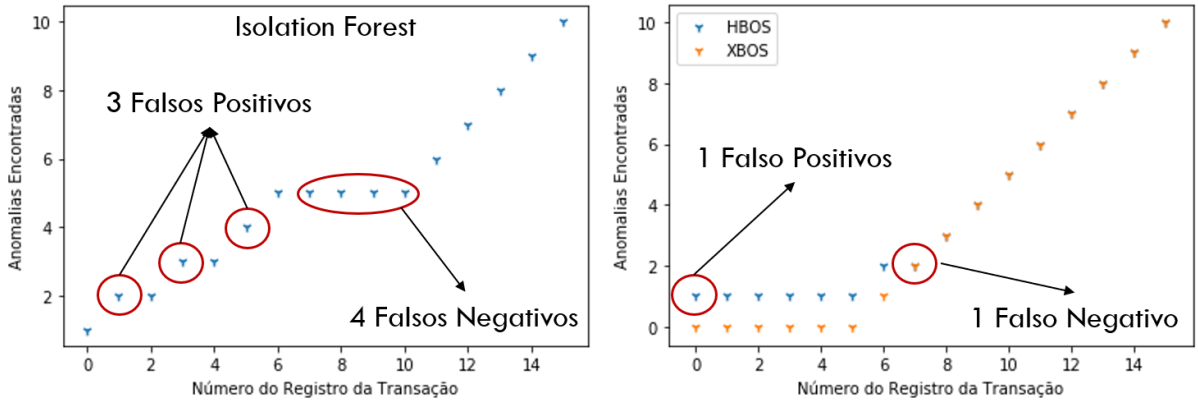

Figura 4.4: Comparativo entre *Isolation Forest*, HBOS e XBOS na sinalização de fraudes.

Os valores de *F1-Measure* foram usados para eleger os algoritmos que têm melhor desempenho (ver Tabela 5.1), e para apresentar suas previsões de forma visual na interface

SDF. Algoritmo de detecção de *outliers* como a distância de Cook, iForest, SVM e XBOS, tiveram um desempenho melhor do que outros neste cenário de domínio de fraude.

A ferramenta utilizada para geração dos *dashboards* é o SAS VA, atendendo a diretriz de utilizar soluções já adquiridas e com contratos de suporte ativo, possuir uma maior integração para se consumir os *datasets* e modelos construídos em outros módulos do SAS, além de manter a curva de aprendizado o mais baixa possível no Banco X.

4.3 Análise Humana

As soluções atuais de mercado negligenciam o potencial das técnicas de VA para integrar a análise humana ao processo e não permitem a inserção de regras para calibrar o modelo. A maioria dos casos fraudulentos não são facilmente previsíveis por regras comuns e requerem alguma investigação humana.

Por meio do FDS os especialistas da área de prevensão e combate à fraudes exploram várias transações simultaneamente. Durante essa exploração, os investigadores usam sua experiência pessoal para decidir se uma transação deve ser considerada fraudulenta ou não. O FDS permite a realização de tarefas de exploração e navegação nos dados, para obter insights que os especialistas nem estavam procurando, ou dar-lhes dicas para examinar partes específicas no conjunto de transações.

Há evidências experimentais de que a análise humana reduz os falsos positivos. O item **(a)** da Figura 4.5, sem a análise humana, dentro do universo de 10 fraudes reportadas, o modelo identificou apenas 3 *outliers*, ou ocorrências de fraudes. Com o ajuste e calibração sugeridos pelos especialistas da área de combate e prevenção à fraudes, o modelo passou a considerar o intervalo de tempo entre transações, conta envolvida e valor, após implementação de regra de negócio sugerida pelos especialistas.

Esta nova regra de comportamento permitiu calibrar o modelo de forma que fosse possível identificar as 10 ocorrências de fraude, conforme ilustrado no item **(b)** da Figura 4.5.

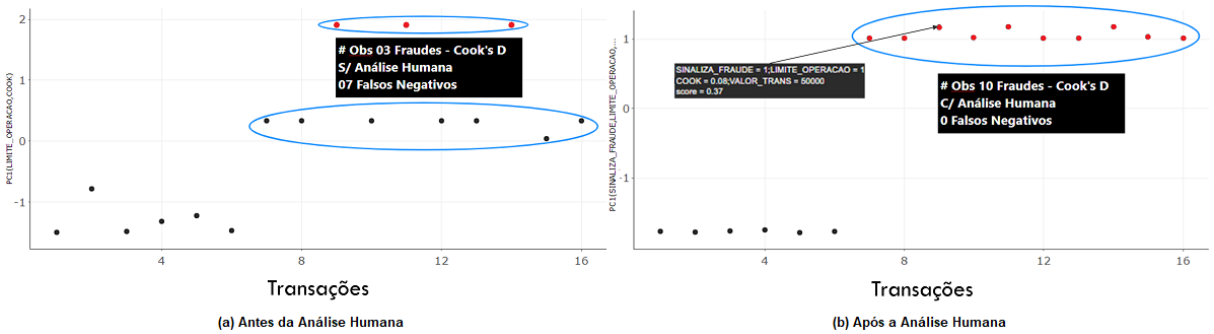

Figura 4.5: Análise humana para melhoria do desempenho

Capítulo 5

Resultados Obtidos

Diferente das publicações existentes, nossa abordagem constrói um modelo para cada conta do cliente com base no histórico de movimentações financeiras. Por meio de algoritmos de identificação de *outliers* o SDF sinaliza transações suspeitas, que são plotadas em consoles de VA para permitir a exploração e análise visual das possíveis movimentações fraudulentas.

5.1 *Dashboard* de *Visual Anaytics*

O SDF traz múltiplas perspectivas de dados conectados, uma vez que os dados de transação são compostos por múltiplas dimensões heterogêneas que precisam ser analisadas entre si. Ele fornece múltiplas perspectivas sobre os dados em várias visualizações conectadas.

O modelo permite interações e análises humanas, visando ajustes e calibrações que reduzem os falsos positivos. O sistema *online* acelera o processo de investigação e tomadas de ações tempestivas, evitando prejuízos por meio de bloqueios das transações suspeitas.

O SDF apresenta na área D da Figura 5.1, quatro dos algoritmos com maior precisão listados na área B da Figura 5.1, considerando os *Box Plots* gerados a partir da métricas F1.

Assim, é possível que modelos apresentem diferentes índices de desempenho conforme o comportamento específico do cliente. Neste sentido, o uso de uma única medida pode trazer a falsa impressão de que um determinado modelo, que apresente uma maior precisão em algum caso, deva ser utilizado em todas as situações. Por essa razão, o SDF é composto por diferentes pontos de vista exibindo diferentes aspectos e previsões dos melhores algoritmos (ver área D da Figura 5.1).

As técnicas de visualização foram escolhidas com relação à adequação de seus atributos visuais, utilizando diferentes codificações para alcançar o melhor equilíbrio possível entre capacidade de distinção, separabilidade e apresentação de informações importantes.

A: Visão temporal - Informações de janela de 6 meses. Com base em uma data de referência, o eixo x contém transação de uma janela de 6 meses, a partir de um tipo específico (por exemplo, Pagamentos, Cheques e Transferências de Contas), enquanto o eixo y representa a quantidade total de dinheiro transacionada por dia e segmentada por valor.

B: *Box Plots F1-Measure* por Algoritmo. Uso da pontuação F1 como uma maneira probabilística de medir a precisão dos algoritmos. Este gráfico contém o *Box Plots* para pontuações de F1 calculados usando a matriz de confusão de cada usuário, que foram geradas por diferentes algoritmos, com base nos dados dos clientes avaliados nos dados dos últimos 6 meses. Os algoritmos utilizados foram aqueles definidos na Seção 4.2. O i-Forest, *One-Class SVM*, Cook's Di, HBOS e XBOS apresentaram o melhor desempenho conforme ilustrado no gráfico e cujos resultados estão descritos na Tabela 5.1 do Capítulo 6.

C: Informações de transação. Essas informações são responsivas ao período selecionado do gráfico A. Apresenta data, hora, conta e valor para uma faixa para o período selecionado.

D: *Outliers* - Transações suspeitas. Este gráfico é dinâmico e é responsivo, e indica transações sinalizadas como *outliers* e anomalias, com base no período selecionado do gráfico A e transações selecionadas da tabela C. Especialistas da Área de Prevenção e Combate de Fraudes podem visualizar informações adicionais como pontuação e algoritmo usado para previsão.

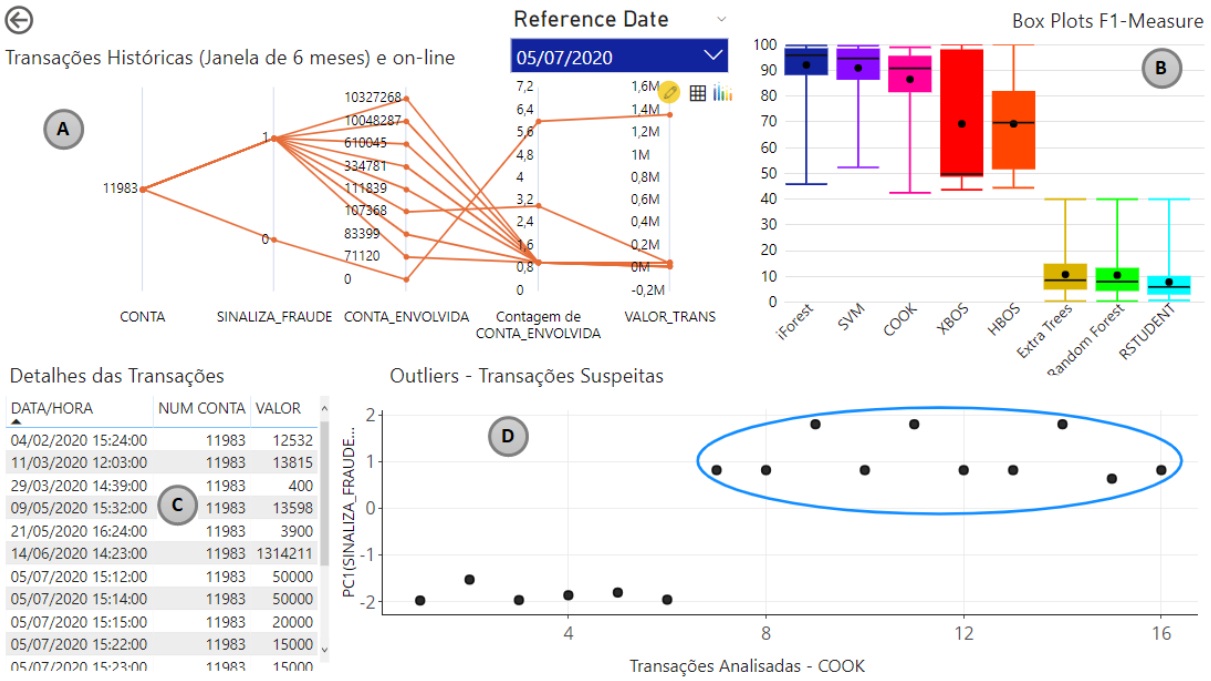

Figura 5.1: VA *Dashboard*.

5.2 Mapeamento do Destino dos Recursos

Por meio do sistema é possível mapear padrões de comportamentos dos fraudadores e identificar as contas envolvidas nos casos de fraude. Em diversos casos observa-se o mesmo padrão de golpe e de uso de técnicas de invasão eletrônica. A Figura 3.1 do Capítulo 3.2, ilustra o percentual de ocorrências em que os fraudadores aplicaram as mesmas modalidades de engenharia social (28,78%), vendas falsas de produtos ou serviços (19,33%), invasão de máquina (16,28%) e clonagem de WhatsApp (12,12%), em diversos clientes.

Outra contribuição do SDF é a capacidade de indicar as contas para onde o recurso financeiro foi destinado, assim como, o tipo de transação utilizada para a transferência do monetária. A Figura 5.2 evidencia esta funcionalidade, onde observa-se as contas que estiveram mais envolvidas com as fraudes registradas, e que há uma maior utilização de transferências via boleto de pagamento, devido à menor rastreabilidade e demora na identificação da fraude por parte do cliente. Como exemplo a Figura 5.2 apresenta o caso a conta de ID 52411719000170, que esteve envolvida em 85 casos de fraudes e recebeu o valor de R$ 887.260,75.

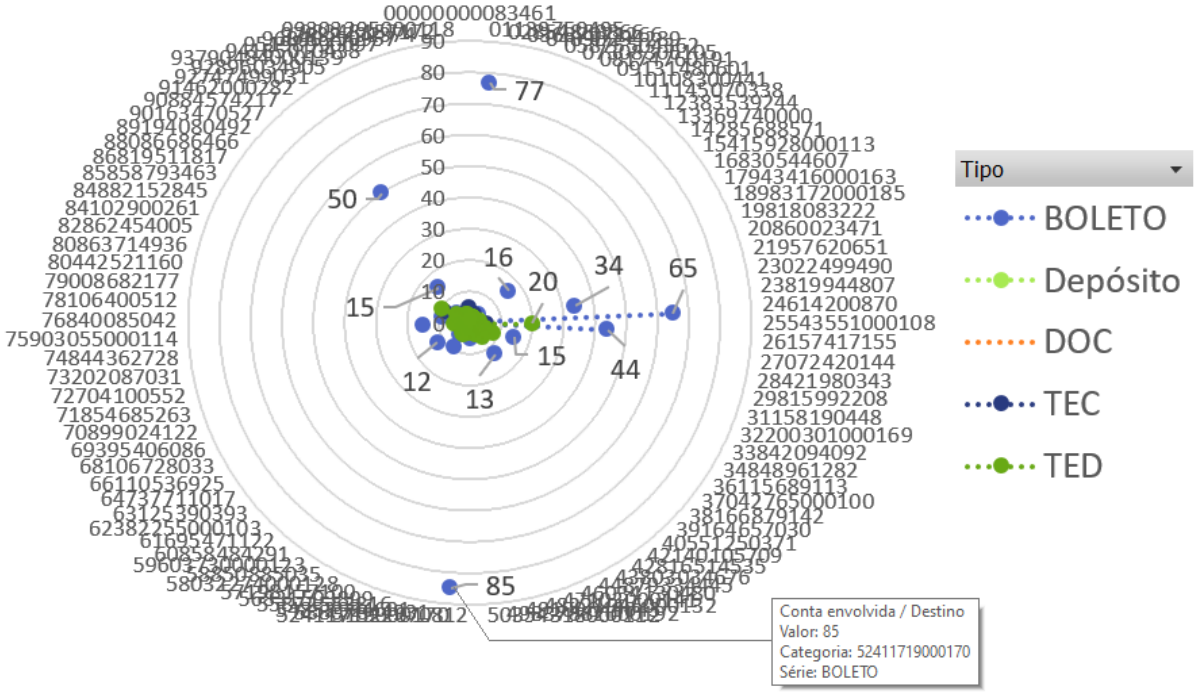

Figura 5.2: Trasações fraudulentas destinadas às contas de um mesmo favorecido.

5.3 Identificação dos Comportamentos das Fraudes

Atualmente, o SDF possui algumas regras de sinalização para alguns padrões de movimentação identificados ao longo do projeto, são elas: Identificação de operações sucessivas, em horários muito próximos (< 30min.) que diferem do comportamento padrão dos últimos 6 meses de movimentação das contas; Identificação de transações com valores similares, elevados e que fogem o padrão da conta; e Análise de movimentações similares, com o mesmo valor, para contas de mesma titularidade, ou não, em um curto intervalo de tempo. A figura Figura 5.3 apresenta um conjunto de transações classificadas como fraudes que se enquadram nas regras citadas anteriormente.

DATA_TRANS	ID_TRANS	COD_TRANS	TIPO_TRANS	VALOR_TRANS	CONTA	BANCO_ENVOLVIDO	AGENCIA_ENVOLVIDA	CONTA_ENVOLVIDA
5/7/20 15:27	32873	5472	D	15000	11983	104	401	71120
5/7/20 15:30	32870	5472	D	15000	11983	33	573	10327268
5/7/20 15:25	32869	5472	D	15000	11983	341	5206	107368
5/7/20 15:36	32867	5472	D	15000	11983	341	465	111839
5/7/20 15:23	32865	5472	D	15000	11983	341	5206	107368
5/7/20 15:22	32864	5472	D	15000	11983	341	5206	107368
5/7/20 15:15	32872	5472	D	20000	11983	341	8605	334781
5/7/20 15:14	32871	5472	D	50000	11983	33	1326	10048287
5/7/20 15:12	32868	5472	D	50000	11983	237	6669	83399
5/7/20 15:34	32866	5472	D	50000	11983	341	3114	610045

Figura 5.3: Transações sinalizadas como fraudes com base em regras de comportamento.

5.4 Comparativo de Desempenho dos Algoritmos

Para a realização dos comparativos de desempenho dos algoritmos avaliados durante este trabalho, foram considerados a soma de todos os resultados das matrizes de confusão dos 1388 clientes, que sofreram algum tipo de fraude entre os meses de janeiro e julho de 2020. Observou-se que algoritmos supervisionados têm impacto negativo na precisão, pois esperam que alguma variável seja explicada e consideram que muitos atributos são irrelevantes durante os processos de análise e correlações. Veja os resultados na Tabela 5.1 abaixo:

Com base na F1-measure que é a média harmônica entre precisão e revocação, o *Isolation Forest* apresentou melhor desempenho com uma precisão de 95%. Por sua vez, o *One-Class SVM* apresentou um desempenho muito próximo, com uma precisão de 94,4%. O algoritmo com o terceiro melhor desempenho foi o Cook com 90,8%. Estes resultados e os dos demais algoritmos podem ser visualizados na coluna F1-Measure da Tabela 5.1.

Os resultados obtidos reduzem a lacuna existente, em que 97% das fraudes identificadas são reportadas pela própria vítima, o cliente. Os algoritmos que obtiveram o melhor desempenho (vide Tabela 5.1) possuem uma precisão entre 90% e 95% na identificação

Tabela 5.1: Desempenho dos algoritmos na detecção de fraudes

Algoritmos	**F1-Measure**	**FPR**	**TPR**
i-Forest	0.951	0.031	0.994
One-Class SVM	0.944	0.027	0.987
Cook's Di	0.908	0.024	0.940
HBOS	0.681	0.021	1.000
XBOS	0.580	0.218	0.589
Extra Trees	0.070	0.025	0.061
Random Forest	0.065	0.025	0.057
R-Student	0.049	0.023	0.028

de transações fraudulentas. Caso o sistema estivesse em operação no período de janeiro à julho de 2020, teria evitado cerca de R$ 29,04 milhões de reais em prejuízos.

Com este enfoque, o projeto adquire um aspecto diferenciado e complementar aos estudos anteriores, tendo como principais contribuições os pontos listados a seguir:

1. Abordagem única de VA interativo, capaz de sinalizar, em tempo real, transações suspeitas de fraudes;
2. Modelo que permite ajustes e calibrações que, através de interações e análises humanas e uma janela deslizante de 6 meses de dados, reduzem os falsos positivos;
3. *Framework* capaz de endereçar problemas reais, reduzindo prejuízos financeiros;
4. Detecção de anomalias em grandes conjuntos de dados; e
5. FDS escala a produtividade permitindo que mais transações sejam analisadas pelos especialistas da área de prevensão e combate à fraudes.

Capítulo 6

Conclusões e Trabalhos Futuros

A metodologia e algoritmos utilizados neste projeto podem sinalizar, em tempo real, se as transações bancárias são potencialmente fraudulentas, para subsidiar ações oportunas, e interromper e/ou reverter as consideradas como indevidas.

Os resultados experimentais obtidos mostram evidências sugeridas por [3] e [4], em que técnicas supervisionadas, como *Extra Trees*, HBOS, *R-Student* e *Random Forest*, têm um melhor desempenho para detectar transações legítimas. Em contraste, elas têm menor precisão quando usadas para classificar eventos fraudulentos. Os resultados expostos na Tabela 5.1 mostram que modelos como i-Forest, *One-Class SVM*, Cook's Di, HBOS e XBOS que são usados na identificação de *outliers*, têm um melhor desempenho e apresentam evidências a favor da hipótese de que no domínio de fraudes os atributos são independentes e não correlacionados.

Em contra partida, algoritmos de detecção não supervisionados e a integração humanas no processo de investigação de transações financeiras suspeitas, trouxeram melhores índices de desempenho do F1-Measure, em comparação aos demais tipos de algoritmos. Além de um maior valor de F1, os algoritmos i-Forest, *One-Class SVM*, Di de Cook, HBOS e XBOS apresentaram uma taxa de descobertas falsas (FDR) mais reduzidas, quando comparado com os algoritmos supervisionados, como mostrado na Tabela 5.1.

Dentre os algoritmos avaliados o *Isolation Forest* apresentou melhor desempenho com uma precisão de 95%, por meio do particionamento recursivo, isolando os pontos cujo caminho possui o maior comprimento com base na distribuição Gausseana.

O SDF propõe uma série de *dashboards* para apoiar os investigadores durante seu processo de tomada de decisão. Há evidências experimentais que não rejeitam a hipótese de que a análise humana reduz os falsos positivos, permitindo interações pelos especialistas da área de prevenção e combate à fraude avalie.

Este conjunto de visualizações apresenta uma variedade de níveis de abstração do mesmo subconjunto dos dados. As técnicas de visualização selecionadas, bem como as

técnicas de interação, levaram em consideração os requisitos de design, derivados da colaboração com nossos especialistas em fraudes, que tinham pouca experiência com ferramentas de exploração visual.

As previsões foram realizadas utilizando dados do mundo real, onde pôde-se demonstrar que é possível escalar o método para casos extremos, realizando as tarefas necessárias de forma adequada e com o desempenho necessário.

O Anexo I apresenta cópia do artigo publicado na IEEE - *Internacional Conference on Machine Learning and Applications* (ICMLA) 2020.

As tarefas envolvidas no SDF são semelhantes às de diferentes domínios de detecção de eventos, a abordagem proposta pode ser transferível para cenários como análise de risco de malware, monitoramento de transações e outros tipos de detecções de anomalias, fazendo uso dos algoritmos apresentados neste trabalho. Entretanto, apesar de ter apresentado uma melhor precisão do que os algoritmos supervisionados o XBOS, que é baseado em análise de clusters, apresentou uma taxa de falsos positivos 10 vezes superior aos demais. Em termos de trabalhos futuros, sugere-se uma investigação aprofundada dos algoritmos de cluster neste cenário.

Outra possibilidade de estudo é o uso de algoritmos de redes neurais, *encoder* e *decoder* para avaliar diferentes tendências e padrões de movimentações, utilizando os 6 meses de dados. Entretanto, devido à quantidade de informações, essas técnicas inviabilizariam as análise em tempo real, já que demandam uma maior quantidade de recursos computacionais do que as disponíveis atualmente no ambiente.

Uma vez descobertos que algoritmos de detecção *outlier* que possuem um melhor desempenho neste domínio específico, outros trabalhos poderiam comparar o desempenho e precisão de algoritmos, tais como: *Local Outlier Factor* (LOF); *k-NN Global Anomaly Score*; *Connectivity-based Outlier Factor* (COF); *Local Correlation Integral* (LOCI); *Local Outlier Probability* (LoOP); e *Cluster-based Local Outlier Factor* (CBLOF).

Por fim, em trabalhos futuros pode-se aplicar métodos de *ensemble*, que usam vários algoritmos de aprendizagem para obter melhor desempenho preditivo do que poderia ser obtido por meio de qualquer um dos algoritmos de aprendizagem utilizados individualmente.

Referências

[1] Aigner, Wolfgang, Silvia Miksch, Wolfgang Mueller, Heidrun Schumann e Christian Tominski: *Visualizing time-oriented data - A systematic view.* Computers & Graphics-UK, vol. 31:pp. 401–409, 2007, ISSN 0097-8493. 1, 8, 16

[2] Kielman, Joe, Jim Thomas e Richard May: *Foundations and Frontiers in Visual Analytics Introduction.* Information Visualization, vol. 8:pp. 239–246, 2009, ISSN 1473-8716. 2, 3

[3] West, Jarrod e Maumita Bhattacharya: *Intelligent financial fraud detection: A comprehensive review.* Computers & Security, vol. 57:pp. 47–66, 2016, ISSN 0167-4048. 3, 8, 15, 23, 24, 34

[4] Hoogs, Bethany, Thomas Kiehl, Christina Lacomb e Deniz Senturk: *A genetic algorithm approach to detecting temporal patterns indicative of financial statement fraud.* Intelligent Systems in Accounting, Finance and Management, vol. 15:pp. 41–56, 2007. doi:10.1002/isaf.284. 3, 8, 24, 34

[5] Mariano, A.M e M.S Rocha: *Revisão da literatura: Apresentação de uma abordagem integradora.* AEDM International Conference – Economy, Business and Uncertainty: Ideas for a European and Mediterranean industrial policy, vol. 33:pp. 1–26, 2017. 6

[6] Lu, Yafeng, Rolando Garcia, Brett Hansen, Michael Gleicher e Ross Maciejewski: *The State-of-the-Art in Predictive Visual Analytics.* Computer Graphics Forum, vol. 36:pp. 539–562, 2017, ISSN 0167-7055. 7

[7] Chandola, Varun, Arindam Banerjee e Vipin Kumar: *Anomaly detection: A survey.* ACM Comput. Surv., vol. 41:pp. 15–58, 2009, ISSN 0360-0300. 7, 10, 11

[8] Fan, M, J Stallaert e AB Whinston: *The Internet and the future of financial markets.* Communications of the ACM, vol. 43:pp. 82–88, 2000, ISSN 0001-0782. 7

[9] Huang, Mao Lin, Jie Liang e Quang Vinh Nguyen: *A Visualization Approach for Frauds Detection in Financial Market.* Information Visualization, páginas 197+, 2009. 8

[10] Bolton, RJ e DJ Hand: *Statistical fraud detection: A review.* Statistical Science, vol. 17:pp. 235–249, 2002, ISSN 0883-4237. 8

[11] Carminati, Michele, Roberto Caron, Federico Maggi, Ilenia Epifani e Stefano Zanero: *Banksealer: An Online Banking Fraud Analysis and Decision Support*

System. ICT Systems Security and Privacy Protection, vol. 428:pp. 380–394, 2014, ISSN 1868-4238. 8, 14, 15, 19, 20, 23

[12] Goldstein, Markus e Andreas Dengel: *Histogram-based Outlier Score (HBOS): A fast Unsupervised Anomaly Detection Algorithm*. Citeseer, vol. 34:pp. 59–63, 2012. 8, 9, 12, 24

[13] Leite, Roger A., Theresia Gschwandtner, Silvia Miksch, Simone Kriglstein, Margit Pohl, Erich Gstrein e Johannes Kuntner: *EVA: Visual Analytics to Identify Fraudulent Events*. IEEE Transactions on Visualization and Computer Graphics, vol. 24:pp. 330–339, 2018, ISSN 1077-2626. 8, 19

[14] Zimek, Arthur, Erich Schubert e Hans Peter Kriegel: *A survey on unsupervised outlier detection in high-dimensional numerical data*. Statistical Analysis and Data Mining: The ASA Data Science Journal, vol. 5:pp. 363–387, 2012. doi:10.1002/sam.11161. 10, 11

[15] Schubert, Erich, Arthur Zimek e Hans Peter Kriegel: *Local outlier detection reconsidered: a generalized view on locality with applications to spatial, video, and network outlier detection*. Data Mining and Knowledge Discovery, vol. 28:pp. 190–237, 2014, ISSN 1573-756X. 10, 11

[16] Akoglu, Leman, Hanghang Tong e Danai Koutra: *Graph based anomaly detection and description: a survey*. Data Mining and Knowledge Discovery, vol. 29:pp. 626–688, 2015, ISSN 1384-5810. 10

[17] Cook, R. Dennis: *Detection of influential observation in linear regression*. Technometrics, vol. 19:pp. 15–18, 1977, ISSN 00401706. 11, 24, 25

[18] Verbeke, Wouter, Karel Dejaeger, David Martens, Joon Hur e Bart Baesens: *New insights into churn prediction in the telecommunication sector: A profit driven data mining approach*. European Journal of Operational Research, vol. 218:pp. 211–229, 2012. 12

[19] Fei, Tony Liu, Ming Ting Kai e Zhou Zhi-Hua: *Isolation-based anomaly detection*. ACM Trans. Knowl. Discov. Data, página 39, 2012, ISSN 1556-4681. 12, 24

[20] Max, Kuhn e Johnson Kjell: *Applied predictive modeling*. Springer, 2013. 13, 24

[21] Chapman, P., J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer e R. Wirth: *Crisp-dm 1.0—step-by-step data mining guide*. https://www.the-modeling-agency.com/crisp-dm.pdf, acesso em 2019-09-03. 13, 14

[22] Li, Yaqi, Chun Yan, Wei Liu e Maozhen Li: *A principle component analysis-based random forest with the potential nearest neighbor method for automobile insurance fraud identification*. Applied Soft Computing, vol. 70:pp. 1000–1009, 2018, ISSN 1568-4946. 24

[23] RC, Prati, Batista G e Monard MC: *Curvas roc para avaliacao de classificadores*. IEEE journal of Latin America, vol. 6:pp. 215–222, 2008. 25

Anexo I

Publicação IEEE ICMLA 2020

O IEEE é a maior organização profissional técnica do mundo, dedicada ao avanço da tecnologia para o benefício da humanidade.

Com a missão de se tornar reconhecida pelas contribuições de tecnologia e promover a inovação tecnológica e a excelência para o benefício da humanidade, todos os anos o IEEE promove conferências internacionais nos mais diversos ramos da ciência e tecnologia.

Para promover e difundir o conhecimento o IEEE oferece uma ampla gama de publicações, que tornam possível a troca de conhecimento técnico e informações entre os profissionais de tecnologia. Este conteúdo é fornecido por meio da Biblioteca Digital IEEE Xplore ®.

Este anexo traz a cópia do artigo publicado na conferência internacional do IEEE - *Internacional Conference on Machine Learning and Applications* (ICMLA) 2020, classificada como A3 no Qualis de eventos do CAPES.

Todos os artigos selecionados para este evento serão publicados no periódico IEEE COMPUTER SOCIETY, ISSN: 1086-3702.

A proposal for online analysis and identification of fraudulent financial transactions

Rodrigo Araujo Lima Torres
Business Intelligence, Risks and Analytics Information
Sicoob - Brazilian Credit Cooperative System
Federal District, Brazil
https://orcid.org/0000-0001-7485-4904

Marcelo Ladeira
Department of Computer Science
UnB - University of Brazilia
Federal District, Brazil
https://orcid.org/0000-0003-1542-6293

***Abstract*—Financial institutions handle with hundreds of thousands of wire transactions per day and need to ensure security and quality for their customers. Searching on predefined patterns is insufficient to identify frauds due to continuous evolution of fraudulent methods used by criminals. Systems used for this purpose are based on the application of some methods of Artificial Intelligence, neglect human process analysis and make little use of Visual Analytics (VA) techniques. Frauds detection domain involves time-oriented and multivariate aspects to identify anomalous transactions making fraud detection a difficult task. We propose the creation of a model for each customer based on his/her behavior, using techniques of identification of outliers and conducting analysis through VA to reduce the false positive rate in the identification of fraudulent financial transactions process. We apply this approach to a real Brazilian financial institution with a daily volume of more than 30 million of financial transactions. Our framework includes a hybrid approach: (1) use of unsupervised outlier detection algorithms; and (2) use of VA to support the real time human analysis with the aim of reducing the incidence of false positives. Potential fraudulent information are presented using VA techniques allowing specialists to evaluate suspicious transactions with no increase of the normal processing times. The initial results obtained sign that there are experimental evidence that our approach can overcome the performance of the fraud detection method today used at the Brazilian institution.**

***Keywords*-Fraud Detection, Visual Analytics, Financial Frauds**

I. Introduction

Brazilian banks as well as other financial institutions around the world suffer impacts of monetary transaction frauds. The National Confederation of Shopkeepers (CNDL) and the Brazil's Credit Protection Service (SPC Brazil) conducted a research showing that 46% of Brazilian people suffered some sort of scam on the Internet between 2018 and 2019. The estimated losses are equivalent to R$ 1.8 billion (about US$ 400 million), impacting 12.1 million people.

The Brazilian bank that supported this work, from now on called Bank X, has 4.4 million customers throughout the country and is present in all Brazilian states and Federal District. It registers more than 30 million transactions per day, like the biggest banks in Brazil. Although most of these transactions are legitimate, 5 to 10 customers report suffering financial fraud each month. Currently, 73% of transactions are made in digital channels, where 53% of it are by mobile applications. Bank X reported 1,388 customers were victims of some sort of fraud from January 2020 to July 2020. During this period, Bank X had recorded losses of R$ 32.27 million (about US$ 7 million) which could not be recovered. Table I describes the financial impact by type of transaction and the number of customers who have suffered some kind of fraud.

Bank X, and other institutions have difficulty to process, in real time, all the volume of transactions data and identify those that are fraud in a timely manner. Bank transaction monitoring tasks involve data of a complex nature [1], based on temporal attributes, which require sophisticated means for analysis and exploration. Only 0.0001% of total transactions are fraudulent, being hard to differentiate them from legitimate ones.

Table I
Financial Impacts by Type

Transaction Type	# Customers	Finacial Losses - R$
Payment Slips	632	5,756,276.90
Checks	1	15,000.00
DOC	17	44,044.80
TEC (Intercredis)	73	613,938.96
TED	678	25,841,114.92
Total	**1,388**	**32,270,375.58**

Table of financial impacts and number of frauds by type of movement.

Commercial fraud detection solutions in the market often involve data mining techniques, neglect human process analysis, and make little use of VA (Visual Analysis) techniques. Kielman et al. [5] describe fraud detection as an open VA problem that requires visual exploration, discovery, and analysis. Looking at financial fraud scenarios, West et al. [2] and Hoogs et al. [3] have concluded in their work that supervised techniques, such as logistic regression, support vector machine and random forest, have demonstrated a low accuracy when used to classify fraudulent events or they produces many false positives.

Our approach is different. We propose the FDS (Fraud Detection System) that builds a model for each customer account based on their financial behavior. FDS identify transaction outliers and VA techniques is used to reduce the false positive rate. The FDS has the following phases: Profile Generation Based on Customer Behavior; Analysis of Online Transactions; Flag Suspicious Online Transactions, using VA; and Human Evaluation and Intervention.

We used real data from Bank X that was built in collaboration with domain experts from the same institution. With FDS running, the experts can act without impact or delay the legitimate online transactions. The same team can analyze a larger set of reports and VA graph analysis helps to identify a greater number of fraud occurrences.

The main contribution of our work is the proposal of an approach to deal with a huge daily volume of financial transactions, allowing to discover and flag, in real time, possible fraud without generating impacts or delays in processing times of legitimate transactions.

Datasets used in this work are real and very sensitive. It is mandatory to comply with the privacy and security regulations. So, we considered existing legislation on the protection of customers'

personal data, and financial and institutional policies. Federal laws regulating manipulation, storage, and use of customers' personal and financial data were considered. It is not allowed to provide the dataset and codes used, to avoid misuse of this information by fraudsters.

The remainder of this paper is structured as follows. Section II details related works on financial fraud domain. Section III presents the Fraud Detection System Framework. Section IV describes the experiments carried on and comments the results. On Section V We present the VA dashboard of selected algorithms. Section VI contains conclusions and future works.

II. Related Work

Despite the number of systems and surveys that focus on fraud detection, they produce too many false negatives, that is, undetected frauds are classified as legitimate. Huang [4] work demonstrate that false negatives can cause more costly damages then false positives where a true transaction is classified as fraudulent. Due to the false negatives problem Kielman [5] states that the process of detecting fraud in banking transactions is a VA problem.

To identify the most impactful studies in frauds domain, and also using VA as an approach to signal suspicions transactions, we have used a word cloud to evaluate terms from titles, key-words and abstract of 518 articles and find the most used terms. Bibliografic-coopling were used on it to identify tendencies and to select the most impactful articles based on the number of citations.

Since 2011, the number of publications related to VA approach has grown to identify bank fraud in real time, and as an alternative capable of processing the volume of transactions involved in this scenario. In 2015, Leman et al. [12] address the use of dynamic graphical models for capturing the interdependencies of data objects, essential in the context of real-time data analysis, and in 2017 Lu et al. [13] present the state of art in VA.

Roger Leite [11] inspired our FDS framework with semi-supervised models, capable of providing an online method of evaluation, scoring each customer transaction based on how much it diverge from the customer historical patterns of transfers and payments. They also present graphic interfaces for human analyses and decisions support, to reduce false positives.

Chandola et al. [6] defines as anomaly detection, the problem of finding patterns that do not conform to expected behavior. The main contribution of their work was the demonstration of effective scatter chart models for identifying outliers.

Carminati et al. [10] identify frauds as a complex problem because there is no specific variable to be explained, in this domain. Their work contribution was describing the use of unsupervised algorithms to identify outliers, seeking for patterns without the need for prior training or data labeling.

Markus Goldstein and Andreas Dengel [9] contribute with the comparison of HBOS, XBOS and other unsupervised models, relative to runtime and accuracy for each of them to detect anomalies. The histogram-based outlier detection algorithm (HBOS) has a shorter linear execution time, by assuming the independence of features. XBOS is a cluster-based anomaly detection algorithm. It has is relatively slower than HBOS but has a higher accuracy and a superior performance then other models.

The Cook's [8] distance calculation algorithm, cited by [6], indicates the distortion degree that transactions generate in the model. Chandola et al. [6] presents how this model can contribute to fraud detection identifying points that negatively influence the least squares regression model. It is a residual analysis technique that allows signaling influential outliers in a set of predictor variables.

In [7], the author presents the area under the curve (AUC) metric as a performance measure. Reference [9] shows the use of AUC for performance comparison among different technics.

III. The Proposed Fraud Detection System

We use an anonymized data set from Bank X which registers more than 30 million transactions each day. We use a 6-months data window of this dataset to build a dataset containing 14,021,371 records from 1,388 customers who suffered some kind of fraud between January and July of 2020. This dataset has 2 transactions classes: legitimate and fraudulent. Fraudulent ones equal 1,626 records within the total amount in this dataset, representing 0.0001% of the sample.

The sparce and imbalanced distribution of the dataset proved to be a huge challenge to be solved. So, we have decided to work under the hypothesis of independent and uncorrelated attributes following Carminati et al. [10] studies. This approximation allows a much easier visualization and interpretation of models and results, on the top of a reduced temporal and spatial complexity. The 6-months data window is updated once a week. Each customer's account is treated separately to avoid transactions influences from one to another. We have used predictive analysis, data mining and machine learning (ML) to identify outlier transactions for each customer's dataset. These processes mitigate the under-training due to the lack of historical data, building well-trained profiles, and calibrating models for evolution of users' financial behavior in the future.

For model calibration, we slide the 6-month data window over time, analyzing new transactions. Behavior analyses considers transactions with the information contained in business engine that follows norms and rules of Central Bank of Brazil.

Experts suggested to implement business rules filters that reduce false positives. Also, we have excluded transaction below R$ 100 (about US$ 20) since the cost involved in assessing these cases are greater than the fraud value, and there is no records of stealthy frauds involving amounts below this amount. To avoid distortions on the customer's model, we created an attribute where experts may set transactions confirmed as fraudulent, so our model ignores it during analysis. Finally, we have implemented a rule to flag sequence of transactions of the same customer, destination, type, and category during a time stamp less than 30 minutes. Transactions with same characteristics occurring very close in time are not usual and fraudster tends to follow this pattern intending to maximize their profits, especially if they get access to customer's account.

FDS framework is a self-adaptative profile-based model, designed for anomaly detection, which computes individual customer financial transactions created on basis of his/her historical transactions. It is structured in four phases as follows:

Profile Generation Based on Customer Behavior. For each customer account FDS generates profiles based on customer's account transaction history (A, Fig. 1). Profile generation is a batch process not linked with the other phases. It has 6-months customer's data sliding windows, which is updated with a delay of one week.

Analysis of Online Transactions. Represents the real-time analysis for fraud detection, taking into account historical and online transactions (B, Fig. 1). Regression residuals are calculated seeking to identify how negatively the new transaction influences the model. It is verified whether the new transaction represents an outlier, distancing from the clusters. Finally, we validate business rules related to behavior considering known type of frauds (e.g., destination, type, category, and time stamp inferior to 30 minutes).

Flag Suspicious Online Transactions. Transactions that negatively affect regression or violate business and behavior rules, are flagged as possible fraud. This information is plotted on dashboards used by experts (C, Fig. 1).

Human Evaluation and Intervention. Experts from Fraud Prevention and Combat Area explore suspicious flagged transaction data, applying filters in VA dashboards. Experts can reject or approve a suspicious transaction through (ALC/CTR) control and authorization layers (D, Fig. 1). A delay (e.g., 30 minutes) in suspicious transactions are applied to enable specialists to run deeper analysis.

Fig. 1 presents the FDS framework integrated with current Bank X's transactional processing system. Each presented layer represent existing systems that are already in daily production, where: **(1) CTE:** Execution Control is responsible for making calls to other layers; **(2) AUD:** Records all transactions for auditing purposes; **(3) CTA:** Authenticate Users/Customers; **(4) CTR:** Transaction Control (team performance today); **(5) CLO:** Control of Operational Limits (Business Rules); **(6) ALC:** Elevation Control/Authorizations; **(7) FDS:** Fraud Detection System. The last column represents the FDS. It is being integrated on the daily production system.

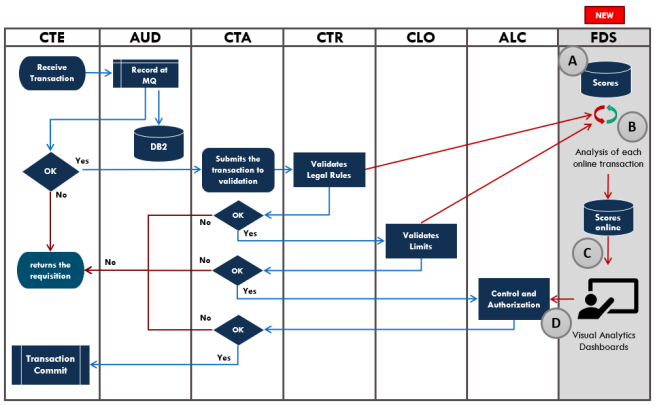

Figure 1. Fraud Detection System Transaction Flow.

To implement a fully integrated, real-time fraud detection framework, FDS should consider Bank X's existing architecture definitions. Solutions responsible for ETL (Extract, Transform and Load), Modeling and ML should be those already acquired and with active support contracts in the institution. At Letter A from Fig. 1, for ETL process, we used IBM Data Stage cluster which has 4 servers in cluster running on top of Suze ZLinux Enterprise V. 12, with 16 cores, and 32 GB of RAM, each. SAS Enterprise Guide 7.1 (64-bit) was used to apply business rules generate profiles for customers. It was processed on 2 cluster servers with 32 cores and 96 GB of RAM. Letter B from Fig. 1 represents real-time outlier analysis over SAS output data using PySpark event Streaming and Kafka. These two components run over 4 Red Hat Enterprise Linux 64 bits, with 16 cores and 64 GB of RAM, each. Letter C from Fig. 1 is implemented using SAS VA on a single servers with 32 cores and 96 GB of RAM. Team intervention occurs through internal system of ALC and CTR, and is represented by letter D from Fig. 1. A total of 224 cores and 672 GB of RAM were allocated to this environment.

IV. Experiments and Interpretation of Results

Our model has an outlier detection approach, which is comparable with Carminati et al. [10], who uses non-supervised technics to provide different kinds of statistical analysis, and correctly ranks complex transactions as suspicious. To carry out the experiments, we used a dataset containing 14,021,371 records from 1,388 customers who suffered fraudulent event, between January and July of 2020, which cannot be provided por security compliance, but is detailed here.

As described on Section III dataset has 2 transactions classes: legitimate and fraudulent. Those classes are highly imbalanced where fraudulent class has just 1,626 records from the total, representing 0.0001% of the sample. Dataset contains 6 months of data and was used to create models based on each customer behavior patterns.

To treat sparce and imbalanced distribution of the dataset we have considered the independence and uncorrelation between attributes. West et al. [2] state that data splitting allows to identify techniques which are more suitable for fraud detection on a specific type of transactions. In our case the large dataset was split into smaller ones based on type o transaction (e.g. Payment Slips, Checks, DOC, TEC, and TED) and account number, enabling to identify which technique has a higher efficiency to flag anomalies and suspicious transactions. Then, we performed an in-depth exploratory and a skew analysis of the dataset, allowing us to understand its main features. With that we have created a model for each user account and transaction type, enhancing prediction and mitigating imbalance issue.

Transactions dataset is composed by the features: unique identification, date and timestamp, code to identify transaction types, values, agency, account, account involved in the transfer or payment, customer risk, person identification (e.g. name, ID, cell phone, city, customer income, birth date), and if it represents a debit or credit into account. Those data can be categorical, numerical, continuous, or temporal. The multivariate aspects are an indicative of complex nature [1] of data, which require a sophisticated combination of different techniques in order to be tackled. We have prepared data to treat time stamps, string conversions, data normalization (Gaussian distribution), and selection of features that have little correlation using PCA (Principal Component Analysis) and Recursive Feature Elimination (RFE). These methods allow extracting important variables from a highly dimensional dataset. With fewer variables the visualization becomes more significant.

We have evaluated 8 algorithms (supervised and non-supervised) to conduct precision comparison, based on this work references, such as: Cook's Distance [8]; Extra Trees [2] [3]; HBOS [9]; Isolation Forest (iForest) [15]; Random Forest [17]; R-Student [2] [3]; One-Class Support Vector Machine (SVM) [16]; and XBOS [9].

The algorithms were set as follows: For Cook's Distance and R-Student we used transaction value as the response variable; An transaction is considered as outlier when Cook's Distance is superior to 3 times the mean (μ) from all other transactions; we defined Isolation Forest parameters to **contamination**='auto' and **behaviour**='new'; Extra Trees and Random Forest got **n_estimators**=customer dataset size, **max_depth**=7 and **random_state**=0; One-Class SVM was defined with **gamma**='auto' and **nu**=0.01; and Once XBOS uses k-means clustering and we have two classes, we have defined the number of clusters to 2.

Since each customer data were split by type of transaction, we have found smaller dataset with less than 10 records. In this scenario some algorithms lost accuracy or are not able to run if we break these datasets into training and testing. So, we used Cross Validation which is more reliable and offers greater accuracy than the technique of dividing the data into training/testing, where each fold is used in training repeatedly and one-fold at a time. So, after running the process in k-1 folds, we can summarize the performance in each fold using the mean and standard deviation.

After training all models, we used the Receiver Operating Characteristic (ROC) to set an appropriate threshold to classify transactions as fraud or non-fraud, due to immense disproportion between classes. Doing this, we managed the confusion matrix for each obtained result.

The ROC chart is based on True Positive Rate, and False Positive Rate. To construct the ROC chart, the FPR (False Positive Rate) is plotted on the x-axis and TPR (True Positive Rate, or Sensitivity, or recall) on the y-axis.

Without loss of generality, we will call the classes as Fraud and Not Fraud. So, TPR measures the rate of frauds correctly identified. In other hand, FPR measures the rate of not frauds correctly identified.

A way to present statistics for evaluating a classification model is by cross-tabbing the class predicted by the model, and the actual class of the sample, as quoted by [14]. This tab is known as a contingency table or confusion matrix.

Confusion matrix is a table with two rows and two columns that reports the number of false positives, false negatives, true positives, and true negatives. When a fraudulent transaction is classified as positive, it is called TP (True Positive). When a normal transaction is classified as positive, it is called FP (False Positive). The same occurs on the opposite situation where FN is a False Negative, and TN corresponds to a True Negative. Figure 2 represents confusion matrix organization, where the following measures, discussed by [14], were applied.

		Actual Classes	
		Fraud	**Not Fraud**
Predicted Classes	**Fraud**	True Positives	False Positives
	Not Fraud	False Negatives	True Negatives

Figure 2. Confusion Matrix.

To evaluate models, we derivate measurements from the contingency matrix, reducing its four main cells to a single numerical quality index. For this, we have compared the matrices with fraud records. In this phase we have involved experts from Fraud Prevention and Combat Area. They have contributed confirming TPs - in case of fraud classification matches to fraud records - and TNs - when models and experts agree that the event is not a fraud. This information became a new dimension in the datasets.

FDR (False Discovery Rate) was obtained to calculate F1-Measure, which is a harmonic mean of precision and recall. Precision is the fraction of relevant instances among the retrieved instances, and the recall measures the total amount of relevant instances that were retrieved.

This metric is better than accuracy calculation because the amount of hits for a legitimate transaction is much higher than the correct indication that the transaction is a fraud, given the imbalance between these two classes.

In many cases this imbalance is so sharp that we have found many datasets with only one class (Not Fraud). To analyze such cases, we have used SVM (Support Vector Machine) One-Class Classification.

To exemplify this visualization processes, we have selected a customer who suffered financial frauds. We used Cook's Distance that can be used to estimate the influence of a transaction when performing a least-squares regression analysis, by removing (Xi, Yi) from the set of observations.

Y-axis from Figure 3 represents Cook's distance (Di), indicating observations transactions which have large residuals, and distort the result and accuracy of linear regression, being considered as Outlier. X-axis represents the transaction number from customer dataset. In this model, a transaction is considered an outlier when Di is superior to 3 times the mean (μ) from all other transactions, like obs. #204.

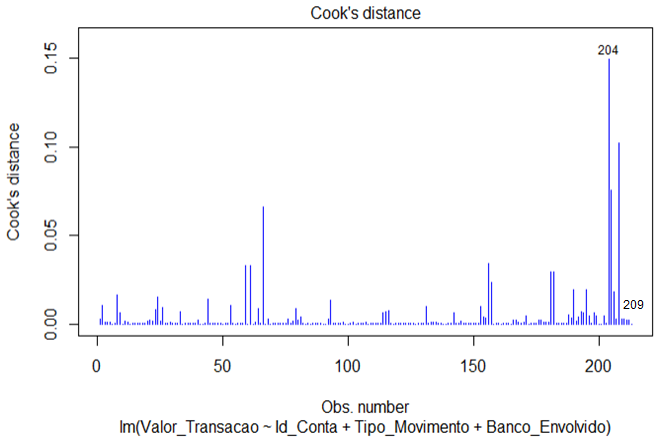

Figure 3. Cook's Di calculated for anomaly detection at case of study.

FDS implements four major predictions using algorithms listed on graph B from Fig. 4, considering F1s obtained.

F1-Measure were used to elect the algorithms which perform better (see Table II), and to present their predictions in a visual manner at FDS interface. Outlier detection algorithm, such as Cook's distance; iForest, SVM and XBOS, performed better than others in this fraud domain scenario.

iForest is a tree-based anomaly detection algorithm. Tree structures are created to isolate anomalies. The result is that isolated examples have a relatively short depth in the trees, whereas normal

data is less isolated and has a greater depth in the trees. SVM One-Class Classification had one of the best performances being more effective for imbalanced classification datasets where there are none or very few examples of the minority class, or in datasets where there is no coherent structure to separate the classes that could be learned by a supervised algorithm.

XBOS is a cluster-based anomaly detection algorithm. It uses k-means as a clustering algorithm. In XBOS, a small cluster near a larger cluster is treated as if it were an average cluster, so the data points are rated "not so anomalous."

V. VA Dashboard of Selected Algorithms

So, it is possible that one model is better than others for a specific customer behavior, but worse than others on a particular case. In such cases, using a single measure may give the false impression that precision can be evaluated using only that measure. For that reason FDS is composed by different views displaying different aspects and predictions from the best algorithms (see Figure 4).

Visualization techniques were chosen with respect to the suitability of their visual attributes, using different visual encodings to achieve the best possible balance between distinguish ability, separability, and pop out of important information.

A: Temporal View - 6-month Window Information Based on a reference date, x-axis contains transaction of a 6-month window, from a specific type (e.g., Payments, Checks and Account Transfers), while the y-axis represents the total amount of money transacted per day, and segmented by value.

B: Box Plots F1-Measure per Algorithm Use of the F1 score as a probabilistic way to measure algorithms precision. This graph contains the Box Plot for F1 scores calculated using confusion matrix of each user, which were generated by different algorithms, based on customers data evaluated on last 6-month data.

C: Transaction Information This information is responsive to selected period from graph A. It presents date, time, account, and value for a range for the selected period.

D: Outliers - Suspicious Transactions This graph is dynamic and is responsive, and indicates transactions flagged as outliers and anomalies, based on selected period from graph A and selected transactions from table C. Experts from Fraud Prevention and Combat Area can visualize additional information like score and algorithm used for prediction.

FDS brings multiple connected data perspectives, since transaction data are composed of multiple heterogeneous dimensions that need to be analyzed in relation to each other. It provides multiple perspectives on the data in multiple connected views.

VI. Conclusion

Our methodology and algorithms can signal, in real time, whether the banking transactions is potentially fraudulent, to subsidize taking timely actions, and interrupt and/or reverse those considered as undue ones.

We use unsupervised outlier detection algorithms and integrate human analysis into the investigation process of suspicious financial transactions, reducing the incidence of false positives on fraud detection and predictions. Fraudulent information is presented using VA technics allowing Fraud Prevention and Combat Area to evaluate suspicious transactions, reducing false positives.

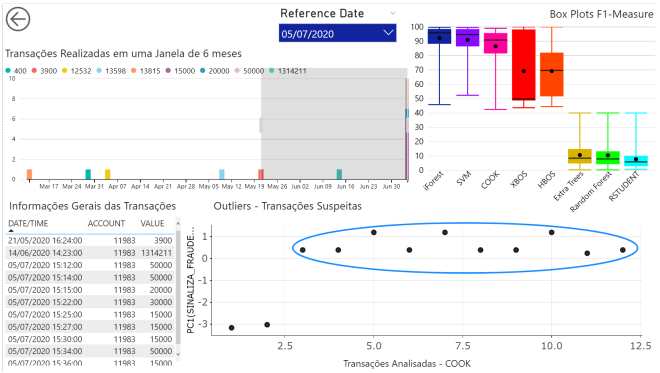

Figure 4. VA Dashboard.

Instead of running queries in spreadsheets and judging alarms by a single overall score value, our FDS (Fraud Detection System) propose a series of dashboards to support investigators during their decision-making process.

This set of views presents a variety of abstraction levels of the same subset of the data. We selected our visualization techniques as well as the interaction techniques with special consideration of our design requirements, derived from discussions with our collaborating domain experts, who had limited experience with visual exploration tools.

We evaluated prediction using real world data and could demonstrate that it is possible to scale well even for extreme cases and to perform the required tasks in a suitable and appropriate way.

Experiments from [2] and [3] have been confirmed, proving supervised technics, such as Extra Trees, HBOS, R-Student, and Random Forest, have a better performance for detecting legitimate transactions. In contrast, they have lower precision when used to classify fraudulent events. So, Models such as i-Forest, SVM One-Class Classification, Cook's Di, HBOS and XBOS which are used for outlier identification, have a better performance, and confirmed the hypothesis of independent and uncorrelated attributes in fraud domain. We have observed that supervised algorithms get negative impact on precision as they expect some variable to be explained and consider many attributes as irrelevant during analysis and correlations. See results on Table II.

Table II
Algorithms Performance in Fraud Detection

Algorithms	F1-Measure	FPR	TPR	FDR
i-Forest	0.951	0.031	0.994	0.088
SVM	0.944	0.027	0.987	0.095
Cook's Di	0.908	0.024	0.940	0.122
HBOS	0.681	0.021	1.000	0.484
XBOS	0.580	0.218	0.589	0.429
Extra Trees	0.070	0.025	0.061	0.918
Random Forest	0.065	0.025	0.057	0.925
R-Student	0.049	0.023	0.028	0.781

Additionally to a higher F1-Measure, i-Forest, SVM One-Class Classification, Cook's Di, HBOS and XBOS have a lower FDR when compared to the others, as shown on Table II.

Based on our study, we also propose possible future research

directions in the field. Since the tasks involved in FDS are similar in different event detection domains, our approach may be transferable to other domains too, such as malware risk analysis, transaction monitoring and other kind of anomaly detections.

It could be better studied different tendencies on 6-months data, using neural networks, encoder and autoencoders. Due to the amount of information used these technics would demand more processing and the focus of this work is on real-time analysis.

Once discovered that outlier detection algorithms has a better performance in this specific domain, further works could also compare performance and accuracy of algorithms, such as: Local Outlier Factor (LOF); k-NN Global Anomaly Score; Connectivity-based Outlier Factor (COF); Local Correlation Integral (LOCI); Local Outlier Probability (LoOP); and Cluster-based Local Outlier Factor (CBLOF).

ACKNOWLEDGMENT

Sicoob (Brazilian Credit Cooperative System) supported this work and agreed to provide real data, respecting all the compliances, customer privacy and security regulations. PPCA program from University of Brazilia, funded by the Brazilian Ministry of Education, University and Research. Thanks for God and family who gave support and encouragement.

REFERENCES

[1] Aigner, Wolfgang, Silvia Miksch, Wolfgang Mueller, Heidrun Schumann and Christian Tominski, "Visualizing time-oriented data - A systematic view," Computers & Graphics-UK, vol. 3, pp. 401–409, 2007, ISSN 0097-8493.

[2] West, Jarrod and Maumita Bhattacharya, "Intelligent financial fraud detection: A comprehensive review," Computers & Security, vol. 57, pp. 47–66, 2016, ISSN 0167-4048.

[3] Hoogs, Bethany, Thomas Kiehl, Christina Lacomb and Deniz Senturk, "A genetic algorithm approach to detecting temporal patterns indicative of financial statement fraud," Intelligent Systems in Accounting, Finance and Management, vol. 15(1-2), pp. 41–56, 2007, doi:10.1002/isaf.284.

[4] Huang, Mao Lin and Liang, Jie and Nguyen, Quang Vinh, "A Visualization Approach for Frauds Detection in Financial Market," Information Visualization, pp. 197+, 2009, DOI 10.1109/IV.2009.23.

[5] Kielman, Joe, Jim Thomas and Richard May, "Foundations and Frontiers in Visual Analytics Introduction," Information Visualization, vol. 8, pp. 239–246, 2009, ISSN 1473-8716.

[6] Chandola, V., Banerjee, A., and Kumar, V., "Anomaly Detection: A Survey," ACM Computing Surveys, vol. 41(3), pp. 1–58, 2009, doi:10.1145/1541880.1541882.

[7] Verbeke, Wouter, and Dejaeger, Karel, and Martens, David, and Hur, Joon, and Baesens, Bart, "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach," European Journal of Operational Research, vol. 218(1), pp. 211–229, 2012.

[8] R. Dennis Cook, "Detection of Influential Observation in Linear Regression," Technometrics, vol. 19, pp. 15–18, 1977, ISSN 0040-1706.

[9] Goldstein, Markus, and Dengel, Andreas, "Histogram-based Outlier Score (HBOS): A fast Unsupervised Anomaly Detection Algorithm," Citeseer, pp. 59–63, 2012.

[10] Carminati, Michele, and Caron, Roberto, and Maggi, Federico, and Epifani, Ilenia, and Zanero, Stefano, "Banksealer: An Online Banking Fraud Analysis and Decision Support System," ICT Systems Security and Privacy Protection, vol. 428, pp. 380–394, 2014, ISSN 1868-4238.

[11] Leite, Roger A., and Gschwandtner, Theresia, and Miksch, Silvia, and Kriglstein, Simone, and Pohl, Margit, and Gstrein, Erich, and Kuntner, Johannes, "EVA: Visual Analytics to Identify Fraudulent Events," IEEE Transactions on Visualization and Computer Graphics, vol. 24, pp. 330–339, 2018, ISSN 1077-2626.

[12] Akoglu, Leman, and Hanghang, Tong, and Danai Koutra, "Graph based anomaly detection and description: a survey," Data Mining and Knowledge Discovery, vol. 29, pp. 626–688, 2015, ISSN 1384-5810.

[13] Yafeng, Lu, and Garcia, Rolando, and Hansen, Brett, and Gleicher, Michael, and Maciejewski, Ross, "The State-of-the-Art in Predictive Visual Analytics," Computer Graphics Forum, vol. 36, pp. 539–562, 2017, ISSN 0167-7055.

[14] RC Prati, and G Batista, and MC Monard, "Curvas ROC para avaliação de classificadores", IEEE journal of Latin America, vol. 6, pp. 215–222, 2008, NBR 6023.

[15] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou, "Isolation-Based Anomaly Detection", ACM Trans. Knowl. Discov. Data, 39 pages, 2012, ISSN 1556-4681.

[16] Max Kuhn, and Kjell Johnson, "Applied Predictive Modeling", Springer, 2013, ISBN-10: 1461468485, ISBN-13: 978-1461468486.

[17] Li, Yaqi and Yan, Chun and Liu, Wei and Li, Maozhen, "A principle component analysis-based random forest with the potential nearest neighbor method for automobile insurance fraud identification", Applied Soft Computing, vol. 70, pp. 1000–1009, 2018, ISSN 1568-4946.