UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Detecção de fraudes em cartão de crédito: um caso de uso de modelos supervisionados no e-commerce brasileiro

Rafael Belmiro Cristovão

Dissertação de Mestrado do Programa de Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria (MECAI)

ICMC USP SÃO CARLOS

Rafael Belmiro Cristovão

Detecção de fraudes em cartão de crédito: um caso de uso de modelos supervisionados no e-commerce brasileiro

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria. *EXEMPLAR DE DEFESA*

Área de Concentração: Matemática, Estatística e Computação

Orientador: Prof. Dr. Gustavo C. Buscaglia

USP – São Carlos
Fevereiro de 2023

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

Belmiro Cristovão, Rafael
B451d Detecção de fraudes em cartão de crédito: um caso de uso de modelos supervisionados no e-commerce brasileiro / Rafael Belmiro Cristovão; orientador Gustavo Carlos Buscaglia. -- São Carlos, 2023.
81 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2023.

1. FRAUDE NO COMÉRCIO E NA INDÚSTRIA. 2. ESTATÍSTICA APLICADA. 3. APRENDIZADO DE MÁQUINA. I. Carlos Buscaglia, Gustavo, orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

Rafael Belmiro Cristovão

Credit card fraud detection: a case study of supervised models in brazilian e-commerce

Master dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master – Professional Masters in Mathematics, Statistics and Computing Applied to Industry. *EXAMINATION BOARD PRESENTATION COPY*

Concentration Area: Mathematics, Statistics and Computing

Advisor: Prof. Dr. Gustavo C. Buscaglia

USP – São Carlos
February 2023

Este trabalho é dedicado à minha família e amigos,
sem os quais esse momento seria impossível.

AGRADECIMENTOS

Os agradecimentos são direcionados às seguintes pessoas: A minha mãe, Marlene, e minha tia, Marli, que lutaram diariamente para me oferecer uma vida melhor. Vocês estão sempre comigo.

À minha noiva, Beatriz, que tenho a honra de dividir meus dias, alegrias e tristezas. Você me torna uma pessoa melhor.

Às minhas irmãs, Daniela e Jessica, por me ensinarem tanto.

Ao meu orientador, Gustavo, por não me deixar desistir e me ajudar até nos piores momentos.

Ao meu amigo e ex-chefe, Roan, por ser um grande líder e viabilizar que eu realizasse o curso.

Aos meus colegas de trabalho, que participaram da construção do pouco que sei e topam participar de longas discussões.

Por fim, à todos meus amigos, por tornarem meus dias melhores.

“Mesmo quando tudo parece desabar,
cabe a mim decidir rir ou chorar,
ir ou ficar,
desistir ou lutar;
porque descobri, no caminho incerto da vida,
que o mais importante é o decidir.”
(Cora Coralina)

RESUMO

CRISTOVAO, R. B. **Detecção de fraudes em cartão de crédito: um caso de uso de modelos supervisionados no e-commerce brasileiro**. 2023. 81 p. Dissertação (Mestrado – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

As tentativas de fraude têm crescido com a chegada de novas tecnologias de comunicação e a digitalização de processos, resultando em grandes perdas financeiras para as instituições. Consequentemente, os métodos de detecção e prevenção de fraudes se tornaram um importante tema a ser explorado.

A fraude de cartão de crédito é uma das formas mais populares de fraude devido à disseminação das compras online, facilidade de utilização de cartões de crédito de terceiros e falta de camadas de validação, como senhas e chips, que acontecem na maioria das compras presenciais.

Muitas técnicas diferentes de extração de variáveis e aprendizado de máquina são utilizadas na criação de modelos de prevenção e detecção à fraude. A necessidade de rápida adaptação às mudanças de comportamento, distribuições desbalanceadas e a demora na obtenção da informação de transações fraudulentas são alguns dos desafios que os modelos de prevenção de fraudes devem lidar.

Neste trabalho comparamos diferentes modelos de aprendizado de máquina utilizando-se de uma base de transações reais de uma loja do comércio eletrônico brasileiro, aplicando diversos algoritmos de previsão para comparação de desempenho. Além disso, estudamos o impacto de uma abordagem de aprendizado *online* como alternativa à queda de performance na presença de *concept drift*.

Os experimentos desenvolvidos mostraram que os algoritmos baseados em árvores de decisão possuem os melhores desempenhos na base estudada, sendo o *Gradient Boosting Decision Tree* o algoritmo com melhor resultado. A partir da comparação dos cenários de aprendizado, foi possível identificar que a atualização com lotes semanais melhora o desempenho do algoritmo ao longo do tempo, sendo capaz de reduzir em até 30% os gastos com *chargeback* na presença de concept drift.

Palavras-chave: Fraude, Cartão de Crédito, Ecommerce, Detecção de fraude, Compra Online.

ABSTRACT

CRISTOVAO, R. B. **Credit card fraud detection: a case study of supervised models in brazilian e-commerce**. 2023. 81 p. Dissertação (Mestrado – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

Fraud has grown significantly with the development of new communication technologies and the processes digitalization, resulting in huge financial losses for institutions. Consequently, fraud detection and prevention methods are important topics to explore.

Credit card fraud is one of the most frequent type of fraud due to the popularization of online shopping, ease of using third party credit cards and the lack of validation layers, such as password and chip verification, which are commonly used in face-to-face purchases.

Many different techniques for extracting features and machine learning algorithms are used to create fraud prevention and detection models. The need to quickly adapt to new types of fraud, unbalanced distributions and the delay in obtaining information on fraudulent transactions are some of the challenges that fraud prevention models must deal with.

In this work, we use a real Brazilian e-commerce databaset to compare different machine learning algorithms and study the online learning approach as an alternative to deal with concept drift.

The experiments showed that the decision tree based algorithms performed better and the Gradient Boosting Decision Tree was the best. Moreover, the comparison of different learning strategies revealed that the online learning approach improved the algorithm's performance in the presence of concept drift, reducing by up to 30% the losses with chargebacks.

Keywords: Fraud, Credit Card, Ecommerce, Fraud Detection, Online Purchase.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

SUMÁRIO

CAPÍTULO

1

INTRODUÇÃO

Conhecido popularmente como e-commerce, o comércio eletrônico tem mudado a maneira como os consumidores adquirem produtos e serviços. Kalakota e Whinston, citados por Albertin (1998), definem comércio eletrônico como a compra ou venda de qualquer informação, serviço ou produto através de redes de computadores. Inicialmente, a mudança aconteceu majoritariamente na compra de produtos não perecíveis, como livros, celulares e computadores, onde o ambiente de compras foi transferido da loja física para o ambiente online em grande parte das vezes. Depois, passou a se popularizar para outras categorias de produtos, como perecíveis e digitais. Toda a disseminação da modalidade tem refletido nos números, estima-se que as vendas por e-commerce mundial atingirão 3.914 trilhões de dólares em 2023 (eMarketer, 2022b), um crescimento de 9.7% em relação ao ano anterior. No Brasil, o crescimento esperado é ainda maior, de 17.2% (eMarketer, 2022a).

O cartão de crédito é um dos meios de pagamento mais utilizado nas compras online, sendo escolhido em 87.5% das compras (MoIP Pagamentos, 2010). Entretanto, a facilidade de utilização que faz o cartão de crédito ser o principal meio de pagamento também o torna vulnerável à fraudes. Estima-se que o e-commerce brasileiro sofre mais de R$ 3.6 mil em tentativas de fraudes por minuto (CLEARSALE, 2020), causando grande impacto na receita da loja online. Por este motivo, o tema chama atenção de pesquisadores e trabalhos científicos e se tornou um tópico de pesquisa frequente com o crescimento do interesse em big data e aprendizado de máquina (NIU; WANG; YANG, 2019).

Existem muitas definições de fraude, que são bastante abrangentes. A *Association of Certified Fraud Examiners* (ACFE) define fraude como o uso de sua ocupação para enriquecimento pessoal através do uso indevido deliberado ou má aplicação dos recursos ou ativos da organização empregadora (ABDALLAH; MAAROF; ZAINAL, 2016), já a definição do dicionário Houaiss da Língua Portuguesa para fraude é "qualquer ato ardiloso, enganoso, de má-fé, com o intuito de lesar ou ludibriar outrem, ou de não cumprir determinado dever"(OLIVEIRA, 2016). No

mercado de cartões de crédito, a fraude acontece quando uma pessoa (fraudador) utiliza o cartão de crédito de outra pessoa (proprietário do cartão) por motivos pessoais, sem o consentimento do proprietário e o do emissor do cartão de crédito. Quando isso acontece, o proprietário do cartão entra em contato com o emissor para o não reconhecimento da transação e o início da disputa de *chargeback*.

Para combater ações de fraudadores, as instituições financeiras brasileiras vêm adotando sistemas e métodos estatísticos cada vez mais sofisticados, a fim de separar o cliente legítimo do fraudador (Fidel Beraldi, 2014). Um sistema antifraude, normalmente, possui múltiplas camadas de controle, que podem ser automatizadas ou supervisionadas por humanos (CARCILLO *et al.*, 2019). Dentro dos objetivos do sistema, os mais comuns são a prevenção e a detecção de fraudes.

A camada de prevenção foca em evitar a fraude antes que aconteça, por exemplo, criando processos de validação no fluxo de compra, como o uso de senhas ou a verificação do código validador do cartão, conhecido como CVV. Já a detecção de fraudes objetiva identificar uma fraude o mais rápido possível após sua ocorrência (Yufeng Kou *et al.*, 2004). Normalmente, ela é formada por: motor de escoragem, motor de controle de risco e análise manual. No motor de escoragem, a cada transação é atribuído um escore, normalmente ligado à probabilidade da transação ser fraudulenta. Posteriormente, o motor de controle de risco avalia qual a decisão será tomada a partir do apetite a risco, que por sua vez é controlado pelo ponto de corte do escore e das regras de negócio. Por fim, ainda é possível a utilização da análise manual com o objetivo de avaliar a veracidade da transação e evitar que as fraudes tenham sucesso (POZZOLO *et al.*, 2018). Além disso, também pode-se utilizar de métodos de análise para avaliação das transações sinalizadas pelo processo de *chargeback* e a detecção de novos casos de fraude a partir das fraudes encontradas. Nesse caso, identificar fraudes pode não evitar a perda financeira, pois a transação pode já ter sido concluída, porém a importância se dá pela oportunidade de utilizar as informações das fraudes detectadas na retroalimentação e melhoria do sistema antifraude.

Apesar das evoluções dos últimos anos, ainda existem diversas oportunidades de melhorias na detecção e prevenção a fraudes, pois as técnicas implementadas por várias instituições permanecem insuficientes e limitadas (SADGALI; SAEL; BENABBOU, 2020) dado a grande variedade de desafios que são encontrados no tema. Os desafios encontrados nos sistemas antifraudes são diversos e variam desde de problemas tecnológicos até a utilização de modelos de aprendizado de máquina. Um dos desafios mais emblemáticos é a dificuldade em mapear o comportamento do fraudador, dado que tentam se passar por bons compradores e utilizam-se de várias estratégias, mudando-as quando não oferecem os resultados esperados. Além disso, muitos fatores também influenciam o comportamento dos bons compradores, como períodos promocionais ou o lançamento de um novo produto. A mudança de comportamento pode causar o que é definido como *concept drift*, que ocorre quando um modelo treinado aprende determinados tipos de padrões dos consumidores e dos fraudadores, porém o comportamento muda e o modelo não se adapta rápido o suficiente para separar os bons e maus compradores eficientemente

(JANSSON; AXELSSON, 2020). Outras dificuldades frequentemente citadas são:

- Distribuição desbalanceada: dado que os consumidores mal intencionados representam uma pequena parcela do total de consumidores, a quantidade de transações não fraudulentas é muito maior do que as tentativas de fraudes;
- Ataques de fraude em brechas dos sistema: como dito, o fraudador se adapta às respostas que obtém do sistema antifraude. Sendo assim, quando uma fragilidade é encontrada, pessoas mal intencionadas passam a abusar da falha, aumentando a quantidade de fraudes desta modalidade, caracterizando um processo conhecido como ataque de fraude;
- Falta de informação das transações rejeitadas: uma transação rejeitada pela sistema antifraude tem o processo de compra interrompido, sendo assim, não é possível saber se a transação é de um comprador legítimo sem a utilização de fluxos alternativos de avaliação dessas transações;
- Demora na maturação dos dados: quando uma transação é aprovada, a informação sobre a variável resposta, se é fraudulenta ou não, demora à ser obtida uma vez que a informação depende do titular do cartão de crédito identificar o uso indevido na fatura e fazer o processo de sinalização com o banco;
- Alta dimensionalidade: a fraude é um evento complexo e existem diversos fatores que impactam na previsibilidade do evento e na mudança de comportamento dos compradores e fraudadores. Por esse motivo, a quantidade de variáveis utilizada pode ser muito grande;
- Performance do sistema: a competitividade do e-commerce fez com que os sites aumentassem a preocupação com o tempo de resposta das avaliações das transações. Sendo assim, os sistemas antifraudes precisam fazer a avaliação o mais rápido possível.

Por conta da grande variedade de desafios presentes na prevenção e detecção de fraudes, a literatura sobre o tema é bastante variada. Pode-se encontrar trabalhos que variam de acordo com tipo de fraude, parte do problema e técnica utilizada. A abordagem supervisionada possui melhores resultados (NIU; WANG; YANG, 2019) mas possui suas limitações, como a necessidade de processos de geração de variável resposta e a baixa performance padrões de fraude ainda não mapeados. Os modelos não supervisionados são uma alternativa para lidar com essas limitações (ABDALLAH; MAAROF; ZAINAL, 2016) a partir da detecção de padrões de comportamento *outliers* sem a necessidade de supervisão. A abordagem semi-supervisionada também é utilizada, principalmente para disseminação da informação de variável resposta e na combinação de modelos, utilizando técnicas não supervisionadas na construção de *features* e segmentação de transações para utilização nos modelos supervisionados. Além disso, com a surgimento e evolução de técnicas de aprendizado por reforço e aprendizado *online*, cresceu o

interesse da aplicação dessas metologias na detecção de fraudes como opção para lidar com o *concept drift* presente no tema.

Neste trabalho, buscamos comparar diferentes algoritmos preditivos supervisionados utilizando-se uma base real de transações de uma loja do comércio eletrônico brasileiro. Além disso, o trabalho estudar a aplicação de uma abordagem dinâmica de atualização de modelos baseado em *ensembles* feito com lotes temporais apresentada por Pozzolo *et al.* (2018). O atual sistema antifraude da loja é composto de um modelo de risco de fraude, regras de negócio e análise manual. A adaptação do sistema para novos tipos de fraude é, majoritariamente, feita a partir da criação de novas regras, por isso, a utilização de um modelo dinâmico pode diminuir a necessidade de intervenções manuais no sistema, aumentar a performance e melhorar os indicadores de fraude.

Os algoritmos baseados em árvores de decisão se mostraram as melhores alternativas, sendo o *Gradient Boosting Decision Tree* do framework LightGBM o algoritmo que apresentou o melhor resultado. O estudo do aprendizado *online* mostrou que a abordagem de criação de *ensembles* com base em modelos semanais pode reduzir em até 30% os gastos com *chargeback* em períodos de mudanças nos padrões de fraude, entretanto, o aumento do número de semanais para além de duas se mostrou pouco impactante no ganho de desempenho.

O restante do trabalho segue a seguinte organização: no primeiro capítulo, Fraudes no comércio eletrônico, entraremos no contexto do comércio eletrônico descrevendo o seu crescimento e funcionamento atual. Além disso, falaremos sobre o mercado de cartões e o impacto da fraude, passando por seus principais conceitos. Posteriormente, no capítulo Sistemas Antifraudes, detalharemos como a literatura lida com os sistemas antifraudes, descrevendo suas componentes e as principais formas de modelagem. No capítulo Metodologia e Experimentos, faremos o detalhamento do caso de uso estudado. Por fim, na Conclusão, trataremos sobre os principais resultados obtidos no trabalho e possíveis próximos passos.

CAPÍTULO

2

FRAUDES NO COMÉRCIO ELETRÔNICO

2.1 Comércio eletrônico

O desenvolvimento tecnológico tem mudado a maneira como fazemos negócio, o surgimento da internet e dos microcomputadores possibilitaram a digitalização do comércio, dando origem ao comércio eletrônico, popularmente conhecido como e-commerce. Kalakota e Whinston, citados por Albertin (1998), definem comércio eletrônico como a compra ou venda de qualquer informação, serviços e produtos através de redes de computadores. Já Albertin (2000) oferece uma definição mais ampla, como a realização de toda a cadeia de valores dos processos de negócio em um ambiente eletrônico, por meio da aplicação intensa das tecnologias de comunicação e de informação, atendendo aos objetivos de negócio.

De forma geral, uma das principais características do e-commerce é o relacionamento de fornecedores e consumidores dentro de ambientes digitais. O modelo de negócio de uma loja online pode ser classificado de acordo com o agente presente em cada um das pontas da negociação. Os modelos de negócio mais populares estão descritos na tabela 1.

A primeira transação via internet aconteceu em 1982, porém a popularização do e-commerce iniciou na década de 1990, quando começaram a surgir as primeiras lojas online abertas para o público. No Brasil, as primeiras lojas iniciaram na segunda metade da década de 1990, com um aumento significativo de novas lojas no início da década de 2000 (CARDOSO; KAWAMOTO; MASSUDA, 2019). A Figura 1 passa por alguns dos principais acontecimentos no mercado brasileiro.

Desde o início das lojas onlines, o comércio eletrônico têm se tornado cada vez mais frequente no cotidiano da população, como podemos ver na evolução de seus números, que crescem ano após ano. A Figura 2 mostra a evolução do faturamento do e-commerce no primeiro semestre de cada ano.

A facilidade e rapidez na compra de produtos foram fatores que tiveram impacto na

		Consumidores	
		Pessoa Jurídica	Pessoa Física
Fornecedores	Pessoa Jurídica	*Business to Business* (B2B) - Negociações são feitas de empresa para empresa, entre duas pessoas jurídicas.	*Business to Consumers* (B2C) - É o modelo usual do varejo offline. Nele, o negócio ocorre entre empresa e pessoa física.
	Pessoa Física	*Consumers to Business* (C2B) - É a transação realizada entre o consumidor e uma empresa, onde o fornecedor é a pessoa física e o consumidor, a pessoa jurídica.	*Consumers to Consumers* (C2C) - Negociação feita por pessoas físicas com moderação de uma pessoa jurídica. O Modelo de negócio está em crescimento com o surgimento dos *Marketplaces*, plataformas onde pessoas físicas e jurídicas podem fazer anúncios e achar consumidores interessados.

Tabela 1 – Modelos de negócio presentes no comércio eletrônico. Adaptado de Cardoso, Kawamoto e Massuda (2019)

Figura 1 – Principais marcos do ecommerce. Fonte: Ebit | Nielsen.

popularização das lojas online. O cartão de crédito é um meio de pagamento que corrobora com tais fatores, sendo a forma de pagamento mais popular na modalidade por conta da sua simplicidade e conveniência (Fidel Beraldi, 2014). Para o usuário, o seu funcionamento é bastante simples: inserindo as informações do cartão de crédito, o logista as valida e busca aprovação junto à empresa emissora do cartão. Em caso de aprovação, o pagamento é realizado. A simplificação para o usuário é possível pela atuação de diversos agentes que viabilizam o pagamento via cartão de crédito. A próxima seção é dedicada à compreensão mais detalhada do mercado de cartões e sua funcionalidade.

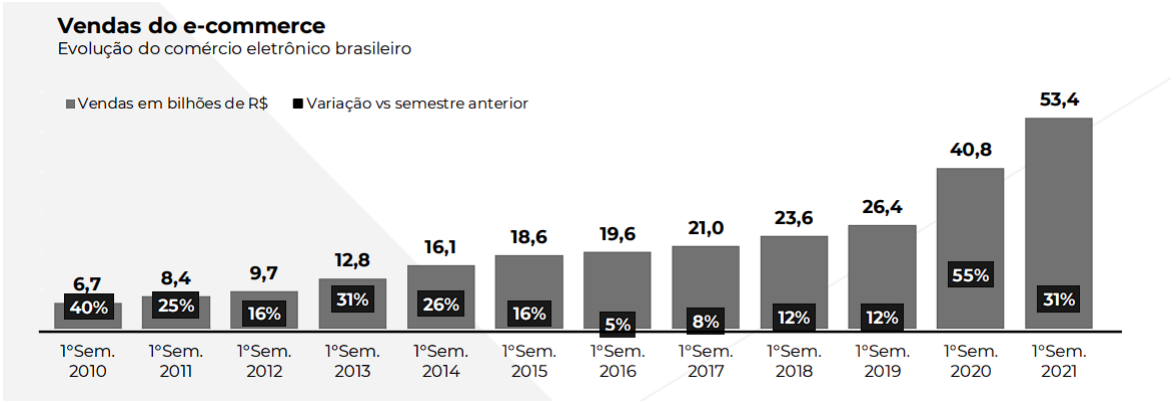

Figura 2 – Fonte: Ebit | Nielsen - Webshoppers 44 l Brasil – Vendas em Bilhões de Reais por semestre, Var% semestre contra semestre anterior.

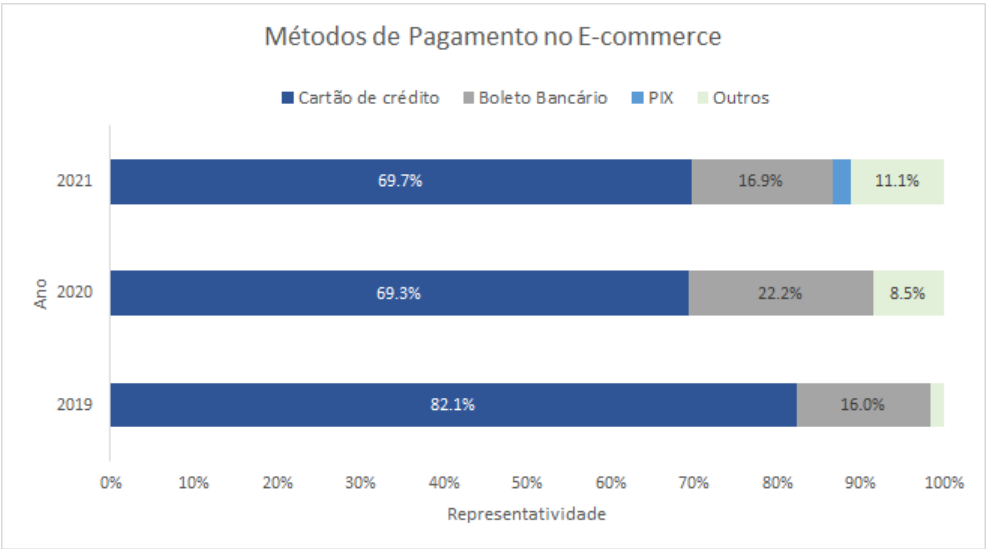

Figura 3 – Utilização dos métodos de pagamento no ecommerce. Fonte: Neotrust - Adaptado pelo autor do Relatório E-commerce 2021 e projeções 2022.

2.2 Mercado de cartões

O cartão de crédito é um meio de pagamento que funciona como uma forma de crédito imediata, registrando a intenção de pagamento em uma data futura. Eles fazem parte do Mercado de Cartões que é formalmente definido como um mercado de dois lados (M2L). Os M2L tem como característica a existência de uma plataforma que organiza e permite o encontro de dois grupos distintos de consumidores (Fidel Beraldi, 2014). No caso do Mercado de Cartões, os grupos são as pessoas portadoras do cartão de crédito, que pretendem realizar o pagamento, e as redes de estabelecimentos comerciais, que oferecem produtos e serviços. O seu funcionamento depende de diversos agentes, entre eles estão:

- Bandeiras: As bandeiras são instituições que criam as plataformas de pagamento. Elas são responsáveis por administrar as políticas das operações, manter a rede de comunicação global e tornar a plataforma atraente, aumentando o número de pagamentos com cartões.

Sua fonte de receita é composta pelas taxas cobradas dos estabelecimentos e a aplicação de multas devido ao não cumprimento das políticas. Alguns exemplos bem conhecidos são a Visa e a MasterCard.

- Emissores: São os responsáveis por conceder crédito, emitir o cartão e manter relacionamento com o titular do cartão de crédito. Suas principais fontes de receita com o mercado de cartões são juros relacionados ao financiamento, anuidades e outros serviços agregados ao funcionamento do cartão. Normalmente, são instituições financeiras que atuam como emissores, onde os principais exemplos são os bancos.

- Adquirentes: São empresas responsáveis por intermediar as transações financeiras junto aos estabelecimentos. Suas principais fontes de receita são as taxas cobradas das transações, os aluguéis das máquinas de vendas e as taxas cobradas dos estabelecimentos pela antecipação dos valores a serem recebidos. Alguns exemplos são a Cielo e a Rede.

- Estabelecimentos: Pessoas físicas ou jurídicas que querem aceitar o pagamento via cartão de crédito.

- Titulares: Pessoas físicas ou jurídicas que estão dispostos a possuir um cartão de crédito.

Os cartões tiveram suas primeiras utilizações no ambiente físico, por isso, seu fluxo básico de funcionamento depende das máquinas distribuídas pelas adquirentes que são utilizadas para capturar as transações. Neste caso, o fluxo de compra simplificado é:

1. Portador ou titular sinaliza o estabelecimento que o pagamento será feito utilizando-se o cartão de crédito;

2. Os dados do cartão são capturados via leitura do chip ou tarja magnética no terminal de venda, normalmente, seguido de solicitação de senha;

3. Dados são enviados para a Bandeira;

4. As Bandeiras repassam às informações aos emissores que, utilizam suas análises de crédito e fraude para decidir a aprovação da transação;

5. Na data de pagamento acordada, os emissores repassam os valores correspondentes às adquirentes e os lançam na fatura dos titulares;

6. Titulares pagam a fatura e as adquirentes enviam o dinheiro para o estabelecimento.

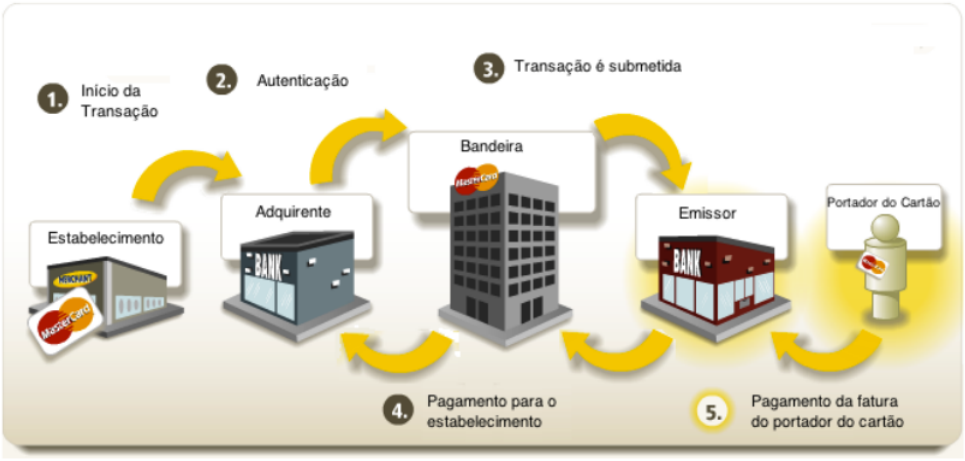

Figura 4 – Fluxo da transação. Fonte: Mastercard. Consultado em Fidel Beraldi (2014).

Com o surgimento do pagamento online, apareceram novas necessidades que motivaram a criação de novos agentes, dentre eles:

- *Gateways*: Funcionam como as maquininhas do mundo físico mas no ambiente online. São empresas que conectam os estabelecimentos às adquirentes, integrando a infraestrutura computacional de ambas as partes. O gateway é responsável pela captura dos dados e sua transmissão para as adquirentes.
- Subadquirentes: Funcionam como facilitadores, os subadquirentes mantém contato com diversas adquirentes e oferecem acesso à elas para seus clientes.

As transações que possuem captura das informações do cartão fisicamente, via chip ou tarja magnética, são conhecidas como transações de cartão presente (CP) . No ambiente online, o envio das informações para a realização do pagamento com cartão é feita pelo usuário, caracterizando as transações de cartão não presente (CNP). O ambiente CNP trouxe novos desafios nas questões relacionadas à validação das informações e no combate à fraude, pois não fazem uso dos processos físicos de autenticação do portador do cartão, como a digitação de senha eletrônica ou a leitura de chip. Por conta disso, ambientes CNP possuem maiores taxas de tentativas de fraudes (Fidel Beraldi, 2014), o que faz o tema constante preocupação para as lojas online. Na próxima seção, discutiremos mais profundamente o tema da fraude.

2.3 Mercado da fraude

As fraudes estão presentes em toda história da humanidade e pode acontecer em diferentes lugares e de muitas formas (Dal Pozzolo *et al.*, 2018). Por conta disso, há muitas definições de fraude, sendo algumas bastante abrangentes. A The Association of Certified Fraud Examiners (ACFE) possui uma definição direcionada ao vínculo empregatício, da fraude como o uso de sua

ocupação para enriquecimento pessoal através do uso indevido deliberado ou má aplicação dos recursos ou ativos da organização empregadora (ABDALLAH; MAAROF; ZAINAL, 2016), já a definição do dicionário Houaiss da Língua Portuguesa é mais geral, dada por "qualquer ato ardiloso, enganoso, de má-fé, com o intuito de lesar ou ludibriar outrem, ou de não cumprir determinado dever"(OLIVEIRA, 2016).

Além da definição de fraude, existem diversas teorias que tentam explicar a motivação das pessoas cometerem esse tipo de crime. Uma teoria amplamente utilizada no mercado é o Triângulo da Fraude. A teoria é baseada no trabalho de (CRESSEY, 1953) e simplifica as condições para o surgimento da fraude em três pilares:

- **Oportunidade**: Condição que faz com que o indivíduo tenha a circunstância propícia para realizar o crime, como por exemplo o conhecimento de brechas sistêmicas/processuais/legais, ausência de fiscalização e rede de contatos.
- **Incentivo**: Fatores que fazem com que o indivíduo cogite a possibilidade de fraude. Por ser algo pessoal, possui alto grau de subjetividade. Alguns incentivos bem conhecidos são a ganância e a necessidade de obter-se dinheiro.
- **Racionalização**: Comportamento pelo qual o indivíduo entende como justificável e legítimo o ato de fraudar, por exemplo, a falta de valorização no trabalho e a subestimação do impacto quando a fraude é feita contra grandes empresas.

A teoria do Triângulo da Fraude é amplamente utilizada, sendo citada por mais de oito mil trabalhos acadêmicos e sendo base para o nascimento de novas teorias (TICKNER; BUTTON, 2020). Entretanto, o triângulo da fraude também é bastante criticado pela sua simplicidade, seu baseamento puramente psicológico e por não retratar a complexidade das fraudes e sua dinâmica de grupo (HUBER, 2017). Além desses fatores, também é importante considerar o esforço necessário para a realização da fraude, o valor capturado, o risco de descoberta do crime e suas consequências.

No mercado de cartões de crédito, a fraude acontece quando uma pessoa (fraudador) utiliza o cartão de crédito de outra pessoa (proprietário do cartão) por motivos pessoais sem o consentimento do proprietário e o do emissor do cartão de crédito. Quando isso acontece, o fraudador consegue acesso ao produto ou serviço comprado enquanto que o proprietário do cartão entra em contato com o emissor para o não reconhecimento da transação, dando início a disputa de *chargeback*.

O *chargeback* não ocorre somente nos casos de fraudes, mas para todos os casos em que é necessário reverter transações realizadas com cartões. Os estabelecimentos recebem um código, conhecido como *Reason Code* com a justificativa do estorno juntamente com a sinalização do valor financeiro. Existem diferentes tipos de *Reason Codes*, que podem ser resumidos nas seguintes categorias:

- Desacordo comercial: Ocorre quando existe alguma divergência entre o titular do cartão e o estabelecimento em relação ao produto ou serviço negociado. Algumas motivações comuns são a não entrega ou demora no envio do produto e produtos com defeitos ou diferentes do originalmente solicitado.
- Erro sistêmico: Transações que foram efetivadas com algum tipo de erro em seu processamento.
- Fraude: Como descrito anteriormente, acontece quando o titular do cartão não reconhece uma transação em sua fatura e entra em contato com o seu emissor para disputa de quem arcará com o valor perdido. É o tipo de *chargeback* que é de interesse do presente trabalho.

As fraudes causam muitos prejuízos no mercado de cartões. No comércio eletrônico, o impacto é ainda maior devido às transações serem mais arriscadas. Estima-se que as perdas por fraudes em ambientes CNP corresponde a 66% das perdas por fraude em cartão de crédito nos países da Europa, e 78% nos EUA (BRYAN *et al.*, 2017). A situação no Brasil é ainda mais crítica, o país possui uma das maiores taxas de fraudes *online* do mundo (CLEARSALE, 2018) e sofre mais de R$ 3.6 mil em tentativas por minuto (CLEARSALE, 2020).

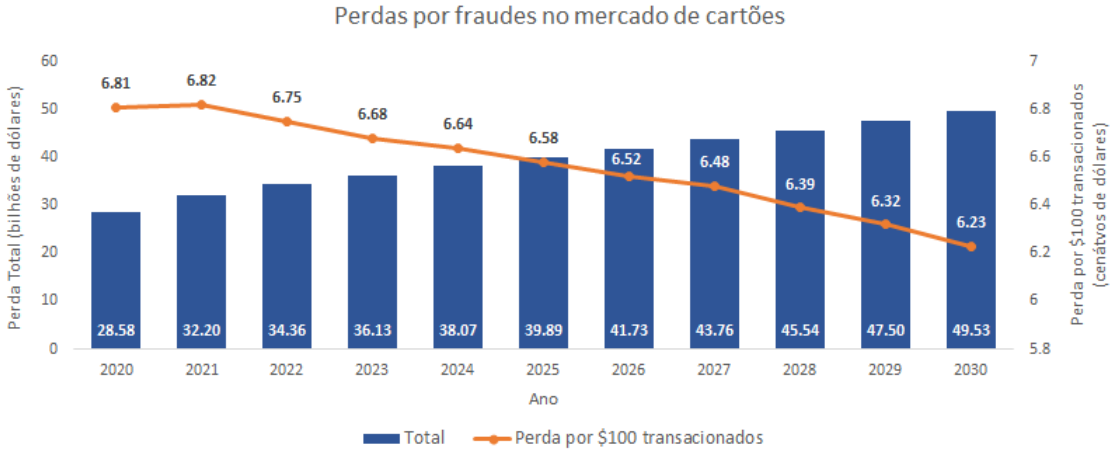

Figura 5 – Perdas por fraude global no mercado de cartões de crédito considerando todos os agendes do mercado: emissores, comerciantes, adquirentes em todos os tipos de cartões. Valor da perda anual e a perda para cada $100 transacionados. Dados realizados de 2020 até 2022 e estimados de 2022 a 2030. Fonte: Adaptado de The Nilson Report de dezembro de 2021.

As perdas provenientes dos produtos e serviços adquiridos através de transações fraudulentas são apenas parte do Custo Total com Fraudes (em tradução livre de *True Cost Of Fraud*), e por isso é necessário levar em consideração outros fatores na avaliação do impacto total, esse é um indicador frequentemente utilizado para avaliação da contratação de produtos e serviços de combate à fraude. Alguns custos são facilmente mensuráveis, enquanto que outros podem requirir estimação ou avaliação qualitativa, por serem intangíveis monetariamente. Outros exemplos de custos são:

- Potenciais multas e taxas aplicadas pelas bandeiras
- Custo de oportunidade em relação à venda perdida
- Custos operacionais de atendimento oferecido à vítima
- Gastos com sistemas de avaliação e proteção contra fraudes
- Perda de vendas legítimas devido ao falso positivo dos sistemas de proteção
- Custos operacionais com equipes de combate à fraude
- Impactos na reputação das marcas do logista, bandeira e emissor
- Perda de consumidores devido à experiência negativa

É importante notar que o valor gasto com fraudes é nocivo para todo ambiente, pois os gastos dos emissores e lojistas podem ser compensados com maiores taxas de juros no cartão de crédito e produtos e até serviços mais caros para o consumidor final.

As fraudes são vistas como um negócio para os fraudadores. Eles atuam em grupos bem organizados, trocam informações e estão constantemente pensando em novas formas de coletar informações de possíveis vítimas e de obter sucesso nas tentativas de fraude. Algumas formas comuns de coleta de informações são a criação de páginas falsas, distribuição de programas maliciosos e o vazamento de dados de empresas e instituições. Posteriormente, os dados coletados são utilizados em tentativas de fraudes. Como comentado anteriormente, existem diversos tipos de fraudes e novos tipos surgem frequentemente, de modo que, a listagem de todos os tipos é uma tarefa bastante difícil. No ambiente de compras de cartões de crédito *online*, alguns dos tipos mais comuns são:

- Fraude deliberada: Acontece quando o fraudador consegue acesso aos dados do cartão de outra pessoa e utiliza-os para a realização de compras. Nota-se que as outras informações utilizadas não precisam necessariamente ser do titular do cartão de crédito. O criminoso pode, por exemplo, criar contas com dados falsos ou usar informações de laranjas no momento da compra. Por se tratar de uma definição bastante ampla, é comum a classificação desse tipo de fraude em subgrupo que compartilham características na sua operacionalização. Alguns exemplos são:
  - Invasão de contas: Acontece quando o fraudador consegue acesso à conta da vitima cadastrada na loja e realiza as compras a partir dela. A principal característica é a compra ser realizada em nome de clientes antigos do estabelecimento.
  - Teste de cartão: Comum quando os criminosos possuem uma base grande de cartões de crédito ou utilizam-se de geradores. Para testar os cartões disponíveis, realizam

compras inicialmente de pequeno valor, que são seguidas de compras de valores mais altos, em caso de sucesso das primeiras.

  - Interceptação de mercadorias: Nessa modalidade, o produto é interceptado em alguma etapa da logística de entrega. Normalmente, esse tipo de fraude está associado ao aliciamento de algum funcionário que tem acesso à mercadoria durante sua entrega.
  - Triangulação: Os fraudadores utilizam-se de plataformas de *marketplace* ou mediadores para o anúncio de produtos. Quando uma pessoa com real interesse na mercadoria faz a compra, seus dados são usados para a compra do produto em outro site utilizando-se de um cartão de crédito de terceiro. Então, o fraudador recebe o dinheiro da venda e a pessoa recebe a mercadoria sem desconfiar que seus dados foram utilizados em uma compra fraudulenta.
  - Aquecimento de contas: Na tentativa de obter uma maior taxa de sucesso, o fraudador cria uma nova conta e realiza algumas transações legítimas para, posteriormente, utilizá-la em compras fraudulentas.

- Autofraude: Nessa modalidade, o criminoso está disposto a utilizar-se de seus próprios dados nas tentativas de fraude. Ele utiliza-se de seus dados cadastrais a fim de dificultar o processo de autenticação das plataformas de combate à fraude. Além disso, o cartão utilizado na compra pode pertencer a outra pessoa ou ao próprio fraudador. No segundo caso, após receber o produto, age de má fé na alegação de fraude junto ao emissor do cartão.

- Fraude Amiga: Fraude amiga ou amigável acontece quando uma pessoa próxima ao titular utiliza-se de seu cartão para realização da compra, que não é reconhecida posteriormente por esquecimento ou falta de aval.

As fraudes são um grande risco à saúde financeira dos agentes dos mercados de cartões. Por conta disso, cada vez mais, investe-se em pessoas, processos e tecnologias para combate à fraude. No próximo capítulo, discutiremos as diferenças, características e funcionamentos dos sistemas antifraudes no ambiente de compras CNP.

CAPÍTULO

3

SISTEMAS ANTIFRAUDES

3.1 Definição

No comércio eletrônico, o cartão de crédito utilizado na compra não está presente no ponto de venda, criando uma situação conhecida como Cartão Não Presente (CNP). Nesse cenário, o banco emissor deixa de se responsabilizar pelas perdas financeiras causadas pelas fraudes, com isso os lojistas se tornam os principais interessados na autenticação das transações. Para conter a ação dos fraudadores, as lojas podem desenvolver sistemas internos, fazer a contratação de empresas especializadas ou utilizarem de ambas alternativas. Os sistemas antifraudes, de maneira geral, buscam analisar as características da transação, do comprador, dispositivo, dados históricos, entre outras variáveis, para decidir se segue ou não com a aprovação do pedido de compra online (Fidel Beraldi, 2014). Normalmente, um sistema antifraude possui múltiplas camadas de controle, que podem ser automatizadas ou supervisionadas por humanos (CARCILLO *et al.*, 2019). Algumas camadas comuns dos sistemas antifraudes são:

- Validação de Dados do Emissor: A instituição emissora do cartão de crédito faz a conferência das informações do cartão para processamento da transação. Entre as informações conferidas estão o código de verificação do cartão ou *Card Verification Value* (CVV), o estado do cartão de crédito (se está ativo ou bloqueado) e o limite de crédito fornecido para o titular. Normalmente, transações que não passam na etapa de verificação são reprovadas por falha no processo de pagamento.

- Sistema de Regras: São um conjunto de regras lógicas desenvolvidas por analistas de fraudes da forma: *Se* <condição> *Então* <ação>. Elas podem ser voltadas à política da loja, por exemplo, no caso de não liberar a compra de bebidas alcoólicas para pessoas com idade menor que 18 anos, ou para gerenciamento de risco, inclusive, utilizando-se de Modelos Estatísticos. Pela facilidade de implementação e entendimento, os sistemas de regras são amplamente utilizados pela indústria, entretanto, a complexidade da manutenção e a

limitação na detecção de padrões deram espaço para a utilização de métodos de previsão mais completos.

Devido à popularidade, alguns trabalhos acadêmicos são dedicados parcialmente ou exclusivamente aos sistemas de regras antifraude. Os principais desafios abordados são: selecionar o conjunto de regras, como em Milo, Novgorodov e Tan (2016) e Gianini *et al.* (2020), e a criação de regras para avaliação dos analistas, como em Oliveira (2016).

- Regras de Pontuação: É similar ao Sistema de Regras, entretanto, cada regra atribui um valor adicional de pontuação e a decisão final é tomada com base na pontuação final da transação, conhecida como escore (ou *score*). O resultado final, o escore, é parecido com a saída de um modelo estatístico, porém o sistema é baseado no conhecimento de domínio dos analistas, por isso, é dependente da capacidade deles encontrarem os padrões de fraude e seu valor associado.

- Modelos Estatísticos: Os modelos estatísticos ou modelos de aprendizado de máquina são as formas mais efetivas de combate à fraude. Eles são puramente baseados em dados mas usam o conhecimento de domínio dos analistas na criação de características que vão ser utilizadas pelas técnicas, popularmente conhecidas como *features*. A partir das *features*, o modelo estatístico pode detectar os padrões nos dados e estimar a probabilidade da transação ser fraude, que também é conhecida como escore. O escore calculado pode ser utilizado para decisão no Sistema de Regras em conjunto com outras informações ou sozinho com a definição de cortes limiares, onde atribui-se uma ação para cada faixa de valor. Além de gerar o escore da transação, os modelos estatísticos podem ser utilizados com diversas finalidades dentro de um sistema antifraude. Alguns exemplos de utilização são: criação de *features* (KUMARI; KANNAN; MUTHUKUMARAVEL, 2014), definição da estratégia de aprovação ou reprovação dos pedidos com base no escore de fraude (LI *et al.*, 2018), previsão das taxas de *chargeback* do sistema (LI *et al.*, 2018), etc.

- Análise Manual: Algumas transações podem ser selecionadas para uma análise mais detalhada ou para seguir protocolos de autenticação mais rígidos, como no caso de um especialista de domínio analisar a transação com objetivo de definir sua legitimidade. A decisão do analista pode ser feita durante o fluxo de decisão da transação, neste caso a análise manual funciona como o decisor final para as transações que as etapas anteriores do sistema não conseguiram classificar, ou posteriormente ao fluxo com o objetivo de coletar informações sobre as transações fraudulentas e melhorar a performance geral do sistema. Ambos cenários precisam ser usados com moderação devido à limitação da mão de obra humana, que reflete no alto custo para análises de transações.

Atualmente, é crescente a utilização de outras camadas de controle que buscam validar a identidade do usuário, conhecidas como segundo fator de autenticação. A utilização é parecida com a análise manual, objetivando autenticar as transações que não tiveram uma classificação

pelas camadas anteriores, porém sem intervenção humana. Diferentes tecnologias são usadas, como biometria de impressão digital, biometria de face e a confirmação da transação via telefone ou e-mail cadastrado.

Com o avanço tecnológico, os sistemas antifraudes têm se tornado cada vez mais complexos e cheios de possibilidades, entretanto não existe receita para a utilização de seus componentes. Diferentes estratégias de negócio utilizam-se de diferentes componentes para a prevenção e detecção de fraudes e o resultado final depende da interação de todas as partes do sistema.

3.2 Desafios dos modelos estatísticos na previsão de fraudes

Os métodos estatísticos são componentes importantes dos sistemas antifraudes por conseguirem obter padrões de fraudes automaticamente por meio dos dados (ABDALLAH; MAAROF; ZAINAL, 2016), conseguir analisar uma grande quantidade de informações ao mesmo tempo e identificar padrões complexos no dados (POZZOLO *et al.*, 2018), entretanto a utilização de algoritmos matemáticos no combate à fraude também enfrenta desafios. Alguns desafios comuns são:

- Desbalanceamento: As transações fraudulentas representam apenas uma pequena parte da quantidade total das transações, sendo esse cenário conhecido como um problema de classes desbalanceadas, dado que a distribuição das transações tem grande acúmulo para a classe de transações legítimas (não fraudulentas). Determinados algoritmos de aprendizado de máquinas requerem estratégias para lidar com esse tipo de problema (POZZOLO *et al.*, 2018). Ademais, algumas métricas de performance usuais para problemas de classificação tornam-se pouco informativas.

- *Concept Drift*: O fraudador está sempre tentando se passar por um consumidor comum a fim de obter sucesso na realização da fraude (SADGALI; SAEL; BENABBOU, 2020). Além disso, o *feedback* recolhido com a resposta do sistema gera adaptações por parte dos criminosos, de tal forma que o sucesso na realização da fraude pode reforçar o comportamento utilizado. Isso pode gerar um grande volume de tentativas de fraudes com as mesmas características com o intuito de abusar de falhas e as reprovações de transações fraudulentas podem resultar na criação de novos tipos de fraude na tentativa do criminoso em obter sucesso. O comportamento dos compradores legítimos também pode ser impactado por períodos promocionais, mudanças na economia, lançamento de novos produtos, etc. Portanto, a distribuição da fraude e a sua relação com as *features* podem mudar ao longo do tempo, efeito conhecido como *concept drift*, que trás a necessidade de revisões e atualizações frequentes para evitar a perda do desempenho dos sistemas de antifraudes ao longo do tempo (POZZOLO *et al.*, 2018).

- Alta dimensionalidade: A fraude é um problema complexo, em que a sua avaliação depende de muitos fatores, por exemplo, o histórico de compras do usuários, os dados utilizados na transação, dados financeiros do cartão de crédito, informações do dispositivo utilizado na compra, etc. A fim de detectar os padrões relacionados ao comportamento dos fraudadores, é importante considerar o máximo de informações disponíveis para a criação do modelo estatístico, o que pode causar problemas computacionais e atrapalhar sua performance e acurácia (THUDUMU *et al.*, 2020).

- Tempo de chegada da variável resposta: A resposta sobre uma transação, isto é, se uma transação é fraudulenta ou legítima, pode ser conhecida pela análise manual da transação ou pela sinalização de estorno pelo processo de *chargeback*. Na primeira alternativa, é possível uma resposta em tempo curto, normalmente entre 24 e 48 horas, porém a análise é possível para uma pequena amostra de transações devido ao alto custo (POZZOLO *et al.*, 2018). Então, a resposta é obtida pelo fluxo de *chargeback* para a grande parte das transações. Nesse caso, o tempo de resposta é longo por depender da identificação da fraude pelo dono do cartão e todo o fluxo de estorno do banco emissor. As transações que não foram sinalizadas pelo banco podem ser consideradas como legítimas após a passagem do tempo de notificação, que pode ser definido assumindo certo tempo de reação dos clientes para identificação e reporte das fraudes. A demora da informação pode causar uma desatualização no sistema e ignorar informações recentes dos *feedbacks* da análise manual, pode resultar em uma performance pior quando comparado com sistemas que usam a informação eficientemente (POZZOLO *et al.*, 2018).

- Perda de informação das transações rejeitadas: Uma vez que a transação é rejeitada perdemos a informação sobre sua resposta, isto é, não conseguimos a confirmação se a transação é fraudulenta. Mesmo nos casos em que as transações são revisadas manualmente, caso não sejam aprovadas pelas componentes do sistema, as respostas coletadas teriam os vieses dos analistas que fizeram a avaliação da transação. No caso de reprovações sem análise manual, a perda de informação é ainda maior, já que não teremos nenhuma informação adicional de análise. Nesse caso, na maioria das vezes, não saberemos se a transação é legítima e não existe possibilidade da identificação de fraude por parte do titular do cartão.

- Necessidade de resposta em curto espaço de tempo: Para uma boa experiência dos consumidores nas compras onlines, as lojas de e-commerce têm grande preocupação com o tempo de resposta dos sistemas de pagamento, incluindo o processo de avaliação de fraude. Por conta disso, aplicações de detecção de fraude em compras online precisam lidar com recursos limitados de tempo e memória computacional ao mesmo tempo que asseguram boa capacidade de predição (ABDALLAH; MAAROF; ZAINAL, 2016). Sendo assim, a performance computacional deve ser levada em consideração para avaliação de qualquer sistema antifraude utilizado no e-commerce.

Para lidar com os desafios, diferentes técnicas e algoritmos são propostos na literatura. Na próxima seção, detalharemos algumas abordagens utilizadas por diferentes autores.

3.3 Técnicas de aprendizado de máquina nos sistemas antifraudes

Por conta da grande variedade de desafios presentes na prevenção e detecção de fraudes, a literatura sobre o tema é bastante variada. Pode-se encontrar trabalhos que variam de acordo com tipo de fraude, parte do problema e técnica utilizada. A abordagem supervisionada aparece mais frequentemente em modelos de detecção de fraude, existem trabalhos que utilizam métodos estatísticos clássicos e outros que aplicam técnicas mais recentes de aprendizado profundo. Entre as técnicas clássicas, temos Regressão Logística (Niu, Wang e Yang (2019); Fidel Beraldi (2014); Carcillo *et al.* (2019); Oliveira (2016); Correa Bahnsen *et al.* (2016)), kNN (Niu, Wang e Yang (2019)), SVM: Niu, Wang e Yang (2019); Sadgali, Sael e Benabbou (2020); Whitrow *et al.* (2009)), Redes neurais (Aleskerov e Rao (1997)), Árvore de decisão (Niu, Wang e Yang (2019); Correa Bahnsen *et al.* (2016)), Florestas Aleatórias (Niu, Wang e Yang (2019); Correa Bahnsen *et al.* (2016)), Extreme gradient boosting (XGB) (Niu, Wang e Yang (2019)), Árvores de regressão (Soemers *et al.* (2018)) e Naíve Bayes (Whitrow *et al.* (2009)). Muitos trabalhos recentes focam em métodos de aprendizado profundo como uma evolução para sistemas antifraudes, como Redes Neurais profundas, Long Short-term Memory (LSTM) e Gated Recurrent Units (GRUs) citados em Sadgali, Sael e Benabbou (2020) e Generative Adversarial Networks (GAN) (NIU; WANG; YANG, 2019). Os modelos de sequência, como LSTM e GRUs procuram encontrar padrões a partir de um conjunto das últimas transações do usuário e são alternativas mais utilizadas em cenários em que não existem muitas informações das transações realizadas, por exemplo, no caso do modelo ser baseado somente nas informações do cartão de crédito utilizado, localidade e do valor da transação. Nesses cenários, o conjunto de variáveis não é suficiente para a boa performance das outras técnicas.

Ademais, devido a necessidade constante de aprendizado de novos padrões de fraudes, alguns autores focam em métodos de aprendizagem por reforço, como Contextual Multi-Armed Bandit (Soemers *et al.* (2018)) e Dinamic Model Averaging (Fidel Beraldi (2014)) e na utilização de técnicas de aprendizado *online* ((Dal Pozzolo, 2015); Soemers *et al.* (2018)). Além da utilização para atualização dos modelos, outra possível utilização das técnicas de aprendizagem por reforço são focadas na automatização e controle dos cenários de indicadores. No geral, essa etapa fica no motor de controle de risco e utiliza o escore de fraude da transação, o valor da compra e o ponto de corte do escore para decidir quais transações serão aprovadas, reprovadas ou passarão por análise manual dos especialistas. Li *et al.* (2018) faz uma proposta de estratégia de controle de risco baseada em dados e cita a pequena quantidade de artigos focados em discutir como fazer o controle de maneira ótima.

Os modelos não supervisionados também possuem ampla utilização nas pesquisas, principalmente para a geração de variáveis. As principais estratégias para a geração são baseadas em modelos de detecção de outliers, como One-Class SVM (OCSVM) (Niu, Wang e Yang (2019)). Restricted Boltzmann Machine (RBM) (Niu, Wang e Yang (2019)), Z-score (Carcillo *et al.* (2019)), variações de PCA (Carcillo *et al.* (2019); Abdallah, Maarof e Zainal (2016)), Isolation Forest (Carcillo *et al.* (2019); Abdallah, Maarof e Zainal (2016)), Gaussian Mixture (Carcillo *et al.* (2019); Abdallah, Maarof e Zainal (2016)), Regras de associação Fuzzy Association Rules (FAR) (Sadgali, Sael e Benabbou (2020); Oliveira (2016)) e Modelos Escondidos de Markov (Kumari, Kannan e Muthukumaravel (2014); Lucas *et al.* (2020)).

A seguir, apresentaremos as técnicas supervisionada e não-supervisionadas mais populares nos sistemas antifraudes.

3.4 Métodos supervisionados

Em Aprendizado de Máquina, os métodos supervisionados são aqueles que se utilizam de um conjunto de informações passadas para previsão do evento de interesse. No caso de modelos de detecção de fraude, a técnica escolhida utiliza-se de um conjunto de transações passadas no qual as variáveis de interesse, se cada um das transações é legítima ou não, estão disponíveis para atribuir a probabilidade da transação ser fraudulenta. Modelos supervisionados funcionam bem para a detecção de padrões de fraudes passados (CARCILLO *et al.*, 2019), por isso, são amplamente estudados. A escolha da técnica pode variar de acordo com o cenário do problema, o conjunto de variáveis, necessidade de interpretabilidade e capacidade computacional do ambiente, por isso, muitos trabalhos focam na comparação da performance de diferentes técnicas, como em Whitrow *et al.* (2009), Niu, Wang e Yang (2019) e Bhattacharyya *et al.* (2011). A seguir, vamos revisar alguns métodos frequentemente utilizados.

3.4.1 Regressão logística

A Regressão Logística é uma técnica muito popular para problemas supervisionados de detecção de fraude devido sua simplicidade e boa performance (B. Baesens; V. Van Vlasselaer; W. Verbeke, 2015). A técnica foi desenvolvida por Davic Cox em 1958 e é um modelo de regressão em que o evento previsto Y, conhecido como variável resposta, é categórico (NIU; WANG; YANG, 2019). Os métodos de regressão buscam explicar a variável resposta, também chamada de variável dependente, por meio de um conjunto de características $X = (x_0, ..., x_n)$, também chamadas de variáveis independentes, explicativas ou *features*. No modelo de regressão clássico, estimamos a equação:

$$Y = b_0 + b_1 x_1 + ... b_n x_n \tag{3.1}$$

onde $b = (b_0, ..., b_n)$ são os coeficientes que mensuram os impactos das variáveis explicativas na variável resposta.

No caso em que a variável resposta é categórica binária, ela apresenta apenas dois valores, então consideramos a função logística:

$$f(z) = \frac{1}{1+e^{-z}} \tag{3.2}$$

Para formulação do modelo de regressão:

$$P(Y=1|X) = \frac{1}{1+e^{-(b_0+b_1x_1+...b_nx_n)}} \tag{3.3}$$

Reformulando em termos de log da proporção de chance, temos:

$$ln\left(\frac{P(Y=1|X)}{P(Y=0|X)}\right) = b_0+b_1x_1+...b_nx_n \tag{3.4}$$

O método de máxima verossimilhança pode ser usado para a estimação dos coeficientes $b = (b_0,...,b_n)$.

3.4.2 SVM

Máquinas de Vetores de Suporte (SVM) são modelos de aprendizado de máquina que tem a capacidade de fazer regressões e classificações. Se considerarmos o problema de classificação binário, podemos construir um classificador linear da forma:

$$y(x) = w^Tx+b \tag{3.5}$$

onde w^T e b são parâmetros a serem estimados a partir dos dados de treinamento com N exemplos de vetores $x_1,...,x_n \in X$, onde X é o espaço das variáveis, e suas respectivas respostas $y_1,...,y_n \in \{-1,1\}$. No caso em que o conjunto de treinamento é linearmente separável, é possível utilizar um hiperplano para separar as classes -1 e $+1$ perfeitamente (BISHOP, 2007).

$$f(x) = w^Tx+b = 0 \tag{3.6}$$

A equação do hiperplano 3.7 separa o espaço X em duas regiões, então pode-se utilizar uma função sinal g para definição das classes.

$$g(x) = \begin{cases} +1, & \text{se } w^Tx+b > 0 \\ -1, & \text{se } w^Tx+b < 0 \end{cases} \tag{3.7}$$

Utilizando-se a função $f(x)$ é possível encontrar diversos hiperplanos que separam as duas classes, o SVM escolhe o hiperplano que maximiza a margem, definida como a menor distância entre as instâncias das classes e o hiperplano. Sendo assim, podemos representar o treinamento do classificador como um problema de otimização.

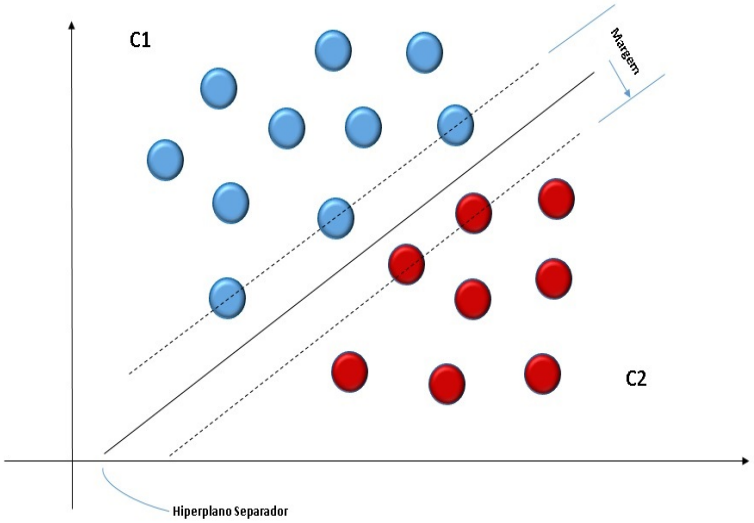

Figura 6 – SVM e o Hiperplano Separador.

Para o caso linearmente não separável, insere-se um termo de erro e_k para tornar a resolução possível. Além disso, pode-se utilizar uma função ϕ conhecida como núcleo ou *kernel* para mapeamento dos pontos de treinamento para um espaço diferente de X, normalmente não linear em relação às variáveis.

Na prática, dado N exemplos de treinamentos:

$$\{(x_i, y_i)\}_{i=1}^{N}, \quad x_i \in R^n, y_i \in \{-1, 1\} \tag{3.8}$$

O hiperplano separador é encontrado resolvendo o seguinte problema de otimização:

$$\min_{w,b,e} \quad \Phi(w,b,e) = \frac{1}{2}\sum_{i=1}^{M} w_i^2 + C\frac{1}{2}\sum_{i=1}^{N} e_i$$

$$\text{sujeito a}$$

$$y_k(w^T\phi(x_k) + b) \geq 1 - e_k, \quad k = 1, ..., N$$

$$e_k \geq 0$$

onde $w = (w_1, ..., w_M)$ são os coeficientes, b é o intercepto, a função $\phi : R^n \rightarrow R^M$ é o núcleo e C é o parâmetro de regularização, que pondera a escolha entre um erro, $e = (e_1, ..., e_N)$, baixo nos dados de treinamento e a minimização da norma do vetor de pesos (NIU; WANG; YANG, 2019).

3.4.3 Redes neurais

As redes neurais são modelos matemáticos que foram criados inspirados no funcionamento do cérebro humano. Podemos tratar a estrutura das redes neurais como uma generalização dos modelos estatísticos existentes (B. Baesens; V. Van Vlasselaer; W. Verbeke, 2015), em que, suas unidades de processamento básicas, conhecidas como neurônios, estão conectadas por canais de comunicação associados ao peso de influência entre eles. As primeiras unidades de

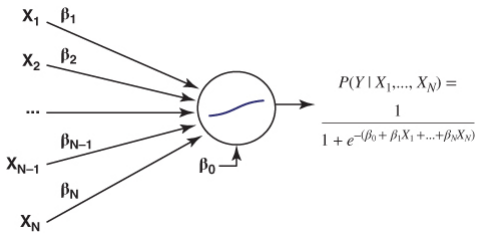

Figura 8 – Representação da Regressão Logística como uma Rede Neural. Fonte: B. Baesens, V. Van Vlasselaer e W. Verbeke (2015)

processamento, conhecida como camada de entrada, são as *features* utilizadas na previsão. Elas são conectadas às camadas de processamento ou escondidas, que trabalham como extrator de novas *features* e enviam a informação para a camada de saída fazer a previsão.

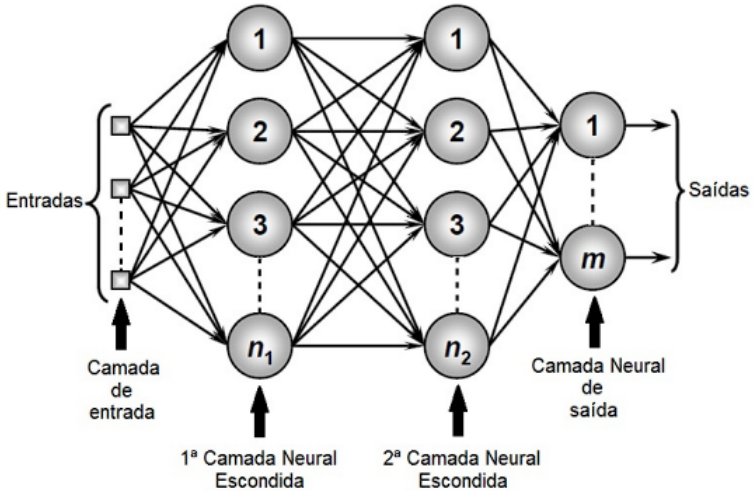

Figura 7 – Arquitetura Rede Neural Perceptron Múltiplas camadas. Fonte: Suterio (2017)

A Regressão Logística, por exemplo, pode ser caracterizada por uma Rede Neural com somente uma camada escondida com um neurônio. Nesse caso, a classificação final é feita com base na função logística (3.3) calculada no neurônio, conhecida como função de ativação.

As inúmeras possibilidades de arquiteturas das redes neurais é um de seus pontos fortes, pois permitem a representação de relações não lineares. Entretanto, a melhor representação para o problema pode ser difícil de ser encontrada devido ao número de hiperparâmetros a serem testados.

A estimativa dos pesos é mais complexa do que na Regressão Logística por conta da quantidade de conexões existentes e pode ser encontrada utilizando-se de algoritimos de otimização, que buscam minimizar uma função custo que geralmente representa o erro de previsão. Iniciam-se com um conjunto de pesos e os ajustam iterativamente até a convergência (B. Baesens; V. Van Vlasselaer; W. Verbeke, 2015).

3.4.4 Árvore de decisão

Árvore de decisão é um método de particionamento recursivo que utiliza uma estrutura de árvore para representar padrões dos dados (B. Baesens; V. Van Vlasselaer; W. Verbeke, 2015). A estrutura se assemelha à uma árvore pela utilização de nós e folhas, em que cada nó representa o teste de uma condição e as folhas o resultado final da classificação. Em outras palavras, as árvores de decisão utilizam-se de uma série de comparações sucessivas nos nós até alcançar uma de suas folhas, onde estará a previsão da árvore.

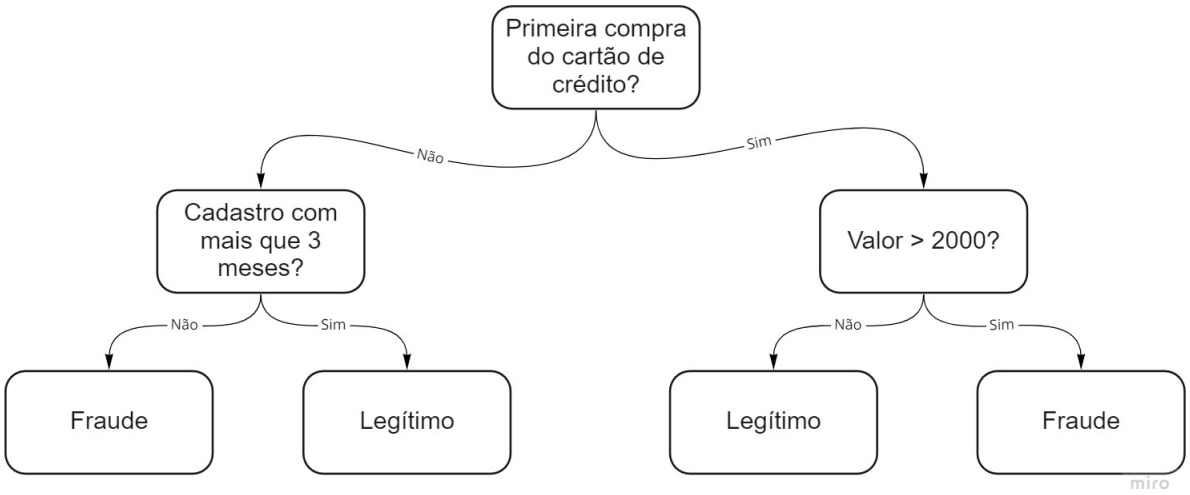

Figura 9 – Exemplo ilustrativo de Árvore de Decisão. Fonte: Elaborada pelo autor.

Existem vários algoritmos para construção de árvores de decisão na literatura. No geral, eles se diferenciam pela maneira em que decidem os nós de particionamento e o tamanho da árvore (B. Baesens; V. Van Vlasselaer; W. Verbeke, 2015). Um dos algoritmos mais populares é o C4.5, em que a cada etapa utiliza-se o ganho de informação para definição do atributo de separação. O ganho de informação é calculado utilizando a medida de entropia ou índice Gini para cada atributo e possibilidade de separação, então o cenário com maior ganho de informação é escolhido. A entropia de um conjunto representa o seu grau de pureza e pode ser calculada por:

$$Entropia(S) = -p_+ log_2 p_+ - p_- log_2 p_- \tag{3.9}$$

onde S é o conjunto de treinamento, p_+ e p_- são, respectivamente, a porção de exemplos positivos e negativos.

Já o ganho de informação representa o ganho esperado na entropia geral de S ordenada pelo atributo A.

$$\text{Ganho de informação}(S,A) = Entropia(S) - \sum_{v \in \text{valores}(A)} \frac{|S_v|}{|S|} Entropia(S_v) \tag{3.10}$$

3.4.5 Métodos baseados em árvores

Os modelos de árvores de decisão são modelos simples que conseguem atingir um bom nível de performance, além disso, sua importância é muito grande devido sua utilização em outra classe de modelos preditivos, conhecidos como *ensembles*. Os métodos *ensembles* são modelos preditivos que utilizam múltiplos modelos estatísticos para fazer a previsão final. A ideia principal é que a utilização de diversos modelos possa cobrir partes diferentes dos dados enquanto complementam suas fraquezas, para isso, as técnicas escolhidas precisam ser sensíveis à mudanças nos dados de treinamento (B. Baesens; V. Van Vlasselaer; W. Verbeke, 2015). As principais técnicas de composição são:

- *Bagging*: É abreviação de *Bootstrap aggregating*. A técnica consiste em gerar modelos treinados em diferentes amostras aleatórias com reposição dos dados iniciais e combinar as respostas de todos para fazer a previsão final, por exemplo, no caso de problemas de classificação, a combinação pode ser feita considerando a classe com maioria de classificações pelos modelos individuais.

- *Boosting*: É uma técnica iterativa em que um modelo inicial é estimado com igualdade de pesos das amostras e a cada iteração novos modelos são treinados utilizando pesos maiores para as observações previstas erroneamente. No geral, a previsão final é feita utilizando a média ponderada de cada um dos modelos.

Diversos algoritmos utilizando *ensembles* foram propostos, a Floresta Aleatória e o *Gradient Boosting* estão entre os mais populares.

3.5 Métodos não supervisionados

Quando uma base de dados histórica com marcações das transações fraudulentas não está disponível, os modelos não supervisionados são recomendados. Esse tipo de modelo não é baseado nas marcações históricas de fraudes, por isso, podem descobrir padrões ainda não detectados (Dal Pozzolo, 2015). No geral, a utilização de modelos não supervisionados em sistemas antifraudes visam a detecção de comportamentos *outliers*, a clusterização de transações e usuários ou a redução de dimensionalidade.

3.5.1 Detecção de outliers

As técnicas empregadas na detecção de fraudes são geralmente uma combinação de métodos de detecção de perfis e de *outliers*. Elas buscam modelar uma distribuição base que representa o comportamento normal e, em seguida, tentam detectar observações que mostram o maior desvio do comportamento esperado (Dal Pozzolo, 2015). Existem diversas técnicas que

podem ser usadas na tentativa de mapear o comportamento esperado da população e definir o nível de desvio, discutiremos brevemente a seguir algumas possibilidades:

- *Z-score*: É a forma de detecção de *outliers* mais comum na abordagem estatística, é a distância em número de desvios padrões que o valor está da média da distribuição. Para cada observação i, o *Z-score* pode ser calculado por:

$$Z\text{-}score_i = \frac{x_i - \mu}{\sigma} \tag{3.11}$$

  onde x_i é o valor observado e μ e σ são, respectivamente, a média e o desvio padrão da distribuição.

  No caso multivariado, isto é, quando existem variáveis *features* disponíveis, as covariâncias entre as distribuições precisam ser consideradas. Nesse caso, a distância de Mahalanobis pode ser usada como uma generalização do *Z-score*.

$$D_{mahalanobis} = \sqrt{(X-M)^T \cdot C^{-1} \cdot (X-M)} \tag{3.12}$$

  onde, X é o vetor com as observações das *features*, M é o vetor com cada uma das médias e C^{-1} é o inverso da matriz de covariância entre as *features*.

- Modelo de Mistura Gaussiana: Pode ser usado para a estimativa da distribuição conjunta das *features*, sendo a medida de *outliers* inversamente proporcional à densidade da distribuição no ponto da observação (CARCILLO *et al.*, 2018).

- Modelos Escondidos de Markov: Modelam o comportamento estocástico do padrão de compra do usuário e utilizam a probabilidade de aceitação da sequência de compras apresentada como medida de *outliers* (KUMARI; KANNAN; MUTHUKUMARAVEL, 2014).

- *Peer Group Analysis*: Um grupo de contas que se comportam de forma semelhante é definido. Quando o comportamento começa a desviar de seus pares, uma anomalia é sinalizada. Para comparação do comportamento de compra, pode-se utilizar testes estatísticos ou medidas de distância (B. Baesens; V. Van Vlasselaer; W. Verbeke, 2015).

- *One-Class SVM*: O *One-Class SVM* utiliza o mesmo algoritmo SVM descrito anteriormente para a detecção de *outliers*. Na sua utilização para a classificação o objetivo é encontrar o hiperplano separador entre as classes. Já no caso de detecção de *outliers*, a ideia é criar hiperplanos que contornam a maioria das observações, então qualquer ponto fora dos limites são considerados como *outliers* (CARCILLO *et al.*, 2019).

- *Floresta de Isolamento*: É uma algoritmo baseado em árvores de decisão para a detecção de *outliers*, onde as observações anômalas são isoladas usando um processo interativo de

criação de árvores binárias. Como observações diferentes são mais facilmente isoladas, a medida de *outlier* utilizada é o tamanho dos caminhos percorridos até as folhas finais das árvores criadas.

3.5.2 Clusterização

O objetivo da clusterização é encontrar subconjuntos do banco de dados que possuem alta homogeneidade entre seus elementos, enquanto mantém alta heterogeneidade entre os subconjuntos (B. Baesens; V. Van Vlasselaer; W. Verbeke, 2015).

No combate à fraude, a clusterização pode ser usada a fim de identificar grupos com comportamentos heterogêneos para utilização de diferentes modelos preditivos (CARCILLO *et al.*, 2018) e também para detecção de *outliers*. No primeiro caso, a diferença de informação e comportamento faz com que os padrões de fraudes sejam diferentes para os grupos, por exemplo, um cliente que está realizando a primeira compra não tem informações de consumo disponíveis então outras variáveis precisam ser consideradas a fim de detectar transações fraudulentas. Já a detecção de *outliers* é possível com a consideração de que itens que não se encaixam em algum dos grupos, ou seja, estão longe dos elementos mais representativos, possuem comportamento anômalo.

Existem diversos algoritmos de clusterização na literatura, que podem ser divididos em hierárquicos e não hierárquicos (B. Baesens; V. Van Vlasselaer; W. Verbeke, 2015). De forma geral, eles funcionam baseados em uma medida de distância entre os elementos e um critério de ligação entre os grupos formados. Várias medidas de distância são encontradas na literatura, alguns exemplos de distâncias são a distância Euclidiana (Equação 3.13) e Manhattan (Equação 3.14).

$$\text{Euclidiana} = ||a - b||_2 \tag{3.13}$$

$$\text{Manhattan} = ||a - b||_1 \tag{3.14}$$

3.6 Adaptabilidade dos modelos antifraudes

Os consumidores e os fraudadores estão constantemente mudando seus comportamentos de compras. Promoções, datas comemorativas e eventos específicos são alguns exemplos de fatores que podem resultar na mudança para do primeiro público. Já os fraudadores podem utilizar as respostas do sistema antifraude para tentar novas formas de obter sucesso e passam a fazer várias fraudes com as mesmas características quando encontram alguma vulnerabilidade, evento conhecido como ataque de fraude. Nesse cenário dinâmico, o sistema antifraude precisa se adaptar rapidamente, a fim de se manter efetivo na identificação de novos perfis de fraude.

Atualizações no sistema podem ser feitas a nível de processo, por exemplo, tornando obrigatório a apresentação do documento no caso da mercadoria ser retirada em uma loja física, regras de negócio, ao evitar a aprovação de pedidos com as características do ataque, ou no modelo estatístico de previsão, fazendo sua atualização.

Os modelos de detecção de fraude podem ser tratados em duas abordagens distintas: aprendizado estático e online. Na abordagem de aprendizado estático, um modelo de previsão é treinado e atualizado conforme a identificação de queda na performance ou periodicamente em um intervalo de tempo estabelecido. No caso da abordagem de aprendizado *online*, o modelo de previsão é atualizado assim que novas informações estão disponíveis (Dal Pozzolo, 2015).

Os algoritmos de aprendizado *online* podem funcionar com atualizações constantes, assim que uma nova predição é feita, ou utilizando lotes de exemplos, quando um número suficientemente grande de informação é acumulada, para mais informações sobre o funcionamento das técnicas de aprendizados *online* consultar Hoi *et al.* (2021). No caso de dados desbalanceados, a segunda opção é uma boa alternativa por conseguir capturar informações das duas classes de interesse. Em seu trabalho, Dal Pozzolo (2015) fez um detalhado estudo sobre o aprendizado *online* na detecção de fraudes. Nele, o autor cria diferentes cenários de comparação para um aprendizado baseado por *ensembles*. Como citado na Seção 3.4.5, os métodos *ensembles* utilizam-se de múltiplos algoritmos de aprendizagem para construção da predição final, com objetivo de conseguir desempenho superior do que ao obtido pelos algoritmos exclusivamente. No caso do trabalho de Dal Pozzolo (2015), a cada novo lote de tempo um algoritmo é treinado com os dados recentes e posteriormente utilizado para predição por um modelo final que utiliza como *features* os modelos criados para diversos lotes. Os cenários de treinamento de modelos testados variam de acordo com o algoritmo utilizado, o método de amostragem, a frequência de atualização do modelo, número de modelos utilizados no *ensemble* e estratégia de construção da amostra. Como conclusão, o autor cita a necessidade de acúmulo de exemplos das classes minoritárias na criação dos lotes de treinamento dos modelos, a melhor performance do algoritmo de Florestas Aleatórias e o resultado superior de cenários de aprendizados feitos a partir da criação de modelos em lotes diários.

É possível encontrar na literatura outros autores que estudam alternativas para a adaptabilidade dos modelo antifraudes devido ao *concept drift*. Fidel Beraldi (2014) aplica o método de Ponderação Dinâmica de Modelos (PDM), traduzido de *Dynamic Model Averaging*, baseado no modelo de regressão logística na detecção de fraudes. O método PDM se baseia em um conjunto finito de modelos candidatos, o autor utiliza-se de diversos modelos de regressão logística considerando subconjuntos das variáveis disponíveis, e assume que os dados seguem uma cadeia de Markov para fazer a atualização dinâmica dos pesos dos modelos e definir o modelo final de previsão. O modelo baseado no método PDM conseguiu melhorar em 34,5% o desempenho em relação à regressão logística padrão. Sadgali, Sael e Benabbou (2020) propõe um sistema de previsão baseado em três camadas, camada de autenticação, comportamental e

decisão. A atualização dos modelos é feita pelo constante treino dos modelos nos novos dados e a descoberta de novas regras de decisão utilizando a partir de Regras de Associação Fuzzy. Por fim, Soemers *et al.* (2018) propõe a utilização uma variação do algoritmo de árvore de decisão, o *Fast Incremental Model Trees - Drift Detection* (FIMT-DD), que é adaptado para lidar com atualizações a partir de novas instâncias de dados. O algoritmo FIMT-DD é utilizado para criar grupos de transações, nas quais serão utilizados pela técnica *contextual multi-armed bandit* para definição das transações que serão analisadas manualmente como *feedback* para a previsão. A previsão final é feita a partir da regressão logística feita para o grupo da transação e atualizada via Descida de Gradiente Estocástica.

O interesse na criação de modelos dinâmicos de previsão à fraude têm aumentado devido a alta presença de *concept drift* nos conjuntos de dados do tema e o crescimento de novas técnicas de aprendizado *online* e de aprendizado por reforço. Na Seção 4.3, aplicaremos a metodologia proposta por Dal Pozzolo (2015) a fim de entender o impacto de tal abordagem na base de dados estudada.

3.7 Métricas de avaliação

Os sistemas de detecção de fraudes de cartão de crédito têm como seu principal objetivo definir se cada uma das transações é fraude, portanto, trata-se de um problema de classificação. A avaliação de um modelo de classificação pode ser feita medindo o grau em que a classe sugerida pelo modelo é correspondente ao caso real (NOVAKOVIĆ *et al.*, 2017). Existem diversas medidas utilizadas para a avaliação da performance dos modelos de classificação e a seleção da medida mais apropriada depende do objetivo, do problema e de suas características.

Em problemas de classificação binária, isto é, quando existem apenas duas possíveis classes, por exemplo transações fraudulentas ou legítimas, a matriz de confusão é amplamente utilizada e ajuda a entender as diferentes métricas de avaliação. Nesse tipo de problema, é comum estabelecer uma classe base, que é chamada de negativo, e uma classe de interesse, também conhecida como positivo.

		Valor Real	
		Positivo	Negativo
Valor Predito	Positivo	Verdadeiro Positivo (VP)	Falso Positivo (FP)
	Negativo	Falso Negativo (FN)	Verdadeiro Negativo (VN)

Tabela 2 – Matriz de confusão

Da matriz de confusão podemos derivar as seguintes métricas:

Acurácia: É uma das métricas mais intuitivas em problemas de classificação, pois mede a porcentagem de classificações corretas do modelo.

$$\text{Acurácia} = \frac{\text{Quantidade Classificações Corretas}}{\text{Quantidade Total de Casos}} = \frac{VP+VN}{VP+FP+VN+FN} \tag{3.15}$$

As maiores desvantagens no uso da acurácia estão na consideração de que os tipos de erros tem mesma relevância e na dependência da distribuição das classes dos dados, por exemplo, no caso de distribuições desbalanceadas, como em problemas de fraudes, a classificação de todas as observações como a classe majoritária vai resultar em um bom nível de acurácia ainda que o modelo erre todas as classificações da classe minoritária.

Precisão: A precisão é definida por:

$$\text{Precisão} = \frac{VP}{VP+FP} \tag{3.16}$$

A precisão pode ser interpretada como o grau de acerto do classificador, em outras palavras, é a probabilidade do classificador acertar caso a previsão seja positiva.

Taxa de Falso Positivo: A taxa de falso positivo (TFP) é a porcentagem dos casos negativos que tiveram sua classificação feita como positivo pelo classificador.

$$\text{TFP} = \frac{FP}{FP+VN} \tag{3.17}$$

Sensibilidade: Também conhecida como taxa de verdadeiro positivo e *recall*, a sensibilidade é definida como a proporção de casos positivos cobertos pelo classificador. Por exemplo, no caso da fraude é a porcentagem do total de fraudes que o classificador consegue detectar como fraude.

$$\text{Sensibilidade} = \frac{VP}{VP+FN} \tag{3.18}$$

Especificidade: Se a sensibilidade passa o grau de cobertura dos casos positivos, a especificidade ou taxa de verdadeiro negativo passa a proporção de casos negativos que são corretamente previstos como negativo pelo classificador.

$$\text{Especificidade} = \frac{VN}{VN+FP} \tag{3.19}$$

No geral, as medidas derivadas da matriz de confusão falham em captar informação das distribuições das classes positiva e negativa ao mesmo tempo. Por conta disso, diversas medidas híbridas foram propostas com objetivo de avaliar o nível de discriminação dos modelos de classificação, principalmente visando problemas com desbalanceamento de classes.

F-Score: É um conjunto de métricas que combinam a precisão e o *recall*. A medida clássica F_1 é definida por:

$$F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{3.20}$$

As demais medidas F_n consideram um fator n que representa a quantidade de vezes em que o *recall* é considerado mais importante do que a precisão, de forma geral:

$$F_n = (1+n^2) \times \frac{\text{precision} \times \text{recall}}{n^2\text{precision} + \text{recall}} \tag{3.21}$$

Existem diversas críticas na literatura ao F-Score (POWERS; PROCESSING, 2015), principalmente voltadas ao viés da métrica para a classe majoritária.

AUC-ROC: A curva ROC é o resultado obtido pelo gráfico da relação entre a taxa de falso positivo, no eixo x, e o recall, no eixo y, para todos os possíveis pontos de corte para a classificação. A métrica final consiste no cálculo da área sobre a curva formada.

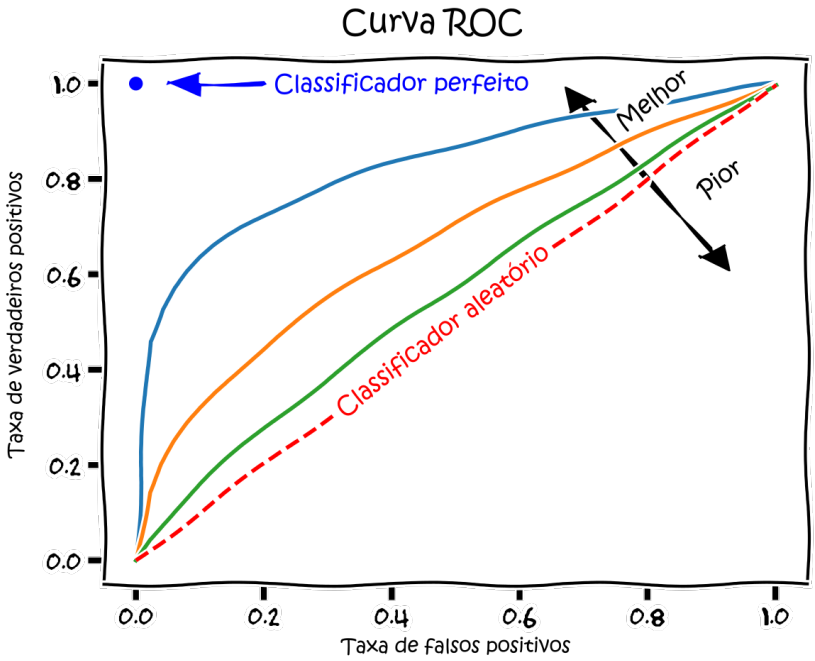

Figura 10 – Curva ROC. Fonte: MartinThoma, CC0, via Wikimedia Commons

A curva ROC é uma das métricas mais utilizadas para problemas de classificação devido sua independência das distribuições das classes, além de conter toda informação contida na matriz de erros e permitir uma comparação visual de diferentes classificadores (NOVAKOVIĆ *et al.*, 2017).

Estatística KS: Muito utilizada para avaliação de modelos de concessão de crédito (FANG; CHEN, 2019), a estatística KS é derivada do teste de hipótese não paramétrico de igualdade de distribuições contínuas, conhecido como teste Kolmogorov-Smirnov. Ela é obtida a partir da maior distância entre as distribuições de score acumuladas das transações fraudulentas e não fraudulentas.

$$F_{\text{Fraude}}(S) = \frac{\#\text{ transações fraudulentas com score} < S}{\#\text{ transações fraudulentas}} \tag{3.22}$$

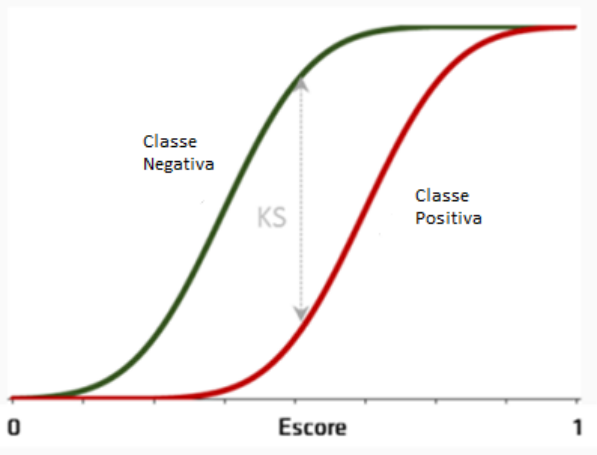

Figura 11 – Estatística KS. Fonte: Elaborada pelo autor.

$$F_{\text{Não-Fraude}}(S) = \frac{\text{\# transações legítimas com score < S}}{\text{\# transações legítimas}} \tag{3.23}$$

$$KS = Max(F_{\text{Não-Fraude}}(S) - F_{\text{Fraude}}(S)) \tag{3.24}$$

onde $S \in [0,1]$ são os pontos de cortes possíveis para a decisão do modelo.

Além de métricas clássicas de performance de modelos de classificação, as empresas de e-commerce costumam acompanhar indicadores de performance que podem ou não ter relações com as métricas vistas. Alguns indicadores importantes são:

Taxa de Aprovação: A taxa de aprovação pode ser calculada em valor monetário ou sobre a quantidade de transações e representa a porcentagem das transações (ou do valor monetário) que foi aprovada. O complementar, conhecido como taxa de reprovação ou negação, também é frequentemente acompanhado. A taxa de aprovação está diretamente relacionada com o ponto de corte escolhido para a classificação final das transações e com a decisão final do sistema. Existe grande interesse em maximizar esse indicador devido sua relação com a receita monetária das empresas, dado que uma venda que não é aprovada pelo antifraude não é concluída. Além disso, como citado em Oliveira (2016), esse indicador tem grande relação com a taxa de falsos-positivos, pois sistemas que possuem uma taxa de aprovação baixa (ou taxa de reprovação alta) têm maiores chances de reprovarem compras de bons consumidores.

Índice de Chargeback: Esse indicador pode ser tratado como a versão prática e monetária do falso-negativo (OLIVEIRA, 2016), portanto, se trata da taxa do valor total das transações que foram aprovadas e eram fraudes.

$$\text{Índice de Chargeback} = \frac{\sum \text{Valor transações de chargeback}}{\sum \text{Valor transações aprovadas}} \tag{3.25}$$

Na prática, esse é um dos principais indicadores de interesse das empresas, pois representa a porcentagem do faturamento perdido devido à aprovação de compras fraudulentas. Além disso, é a métrica acompanhada pelas bandeiras para definição de multas aos lojistas.

Taxa de Contato: É a porcentagem das transações que precisaram de algum tipo de contato com o comprador. Esse indicador é acompanhado, principalmente, por empresas que possuem soluções com análise manual ou algum tipo de segundo fator de autenticação.

3.7.1 Métricas na detecção de fraudes

No caso da detecção de fraudes, a tarefa de selecionar uma medida de performance não é uma tarefa trivial por conta das diferentes estratégias das empresas e das características dos ambientes antifraudes. Alguns desafios comuns citados por Dal Pozzolo *et al.* (2018) são:

- Classes desbalanceadas: A medida escolhida precisa ser robusta ao fato de existir uma quantidade muito maior de transações legítimas do que fraudulentas.
- Estrutura de custos: É difícil atribuir um custo a cada possibilidade dentro de um sistema antifraude, por exemplo, no caso de uma fraude sofrida existem custos bem definidos do produto perdido e de logística, mas também custos não óbvios de oportunidade, perda de reputação da marca, má experiência do usuário, etc.
- Tempo de detecção: No caso de um ataque de fraude, a detecção rápida das primeiras tentativas pode prevenir o surgimento de novas fraudes.
- Erros na variável resposta: As marcações finais das transações fraudulentas podem ser impactadas pelo viés de análise do analista e por erros nas notificações de fraudes, por exemplo, no caso de fraudes de baixo valor que não são reportadas ou problemas de compras legítimas que são reportadas como fraudes.

  Entre outros desafios, vale a pena citar:

- A comparação de dois sistemas antifraudes é complexa devido ao viés em favor do atual decisor das transações. A avaliação dos sistemas só é possível a partir das marcações de fraudes, como o estado final das transações dependem do modelo atual, existe viés de seleção. Por exemplo, uma transação legítima que foi negada pelo sistema atual não será considerada na avaliação.

  Algumas empresas adotam um grupo controle para uma amostra das transações, a fim de coletar informações da variável resposta com menos ruídos. O grupo controle pode ser feito, por exemplo, aprovando algumas transações independentemente da resposta do sistema antifraude para validar quais serão reportadas como fraudes. Alguns desafios em relação à utilização da ferramenta são o custo financeiro e a baixa incidência de fraudes.

Um bom modelo antifraude deve conseguir ordenar bem a probabilidade de fraude das transações, para que o escore seja usado por outras componentes do sistema antifraude e calibrada de acordo com a estratégia de risco do estabelecimento. Por conta disso, a AUC-ROC é a principal métrica utilizada nos trabalhos acadêmicos (Dal Pozzolo *et al.*, 2018). A estatística KS, mais utilizada em problemas de crédito, também é recomendada. Além disso, é crescente os trabalhos que utilizam métricas monetárias baseadas na estrutura de custos de problema, que tenta aproximar os indicadores de negócio com os do modelo. A utilização individual das métricas derivadas da matriz de confusão não são recomendadas devido à sua ruim interpretabilidade em problemas com classes desbalanceadas (POZZOLO *et al.*, 2018), porém podem ser utilizadas em conjunto com outras métricas e indicadores para a mensuração da performance do sistema.

3.8 *Features* de modelos antifraude

Um modelo de aprendizado de máquina é um algoritmo que recebe algum dado de entrada e retorna uma informação de saída, normalmente uma previsão, um agrupamento de dados ou uma decisão. Por exemplo, um modelo de previsão de preço de ações pode receber o histórico de preço das ações do mercado financeiro e gerar uma previsão do preço da ação para uma janela de tempo. Normalmente, os dados de entrada são um conjunto de características do evento de interesse, conhecido como *features*. Em qualquer projeto para desenvolvimento de modelos de aprendizado de máquina, a construção, processamento e a avaliação de *features* são partes indispensáveis. Em modelos antifraude a tarefa é umas das principais partes do desenvolvimento de soluções devido à complexidade do evento, sua rápida mudança, a grande quantidade de fatores que influenciam na detecção e prevenção de fraudes, como padrão histórico de compras do CPF utilizado na transação, informações de padrões de fraude e fluxo de venda da loja.

Muitos trabalhos realizados na academia utilizam somente um conjunto básico de informações das transações, que, em grande parte das vezes, passaram por algum tipo de transformação para que o dado original não fosse divulgado. O principal motivo é a sensibilidade das informações contidas nas bases de dados do tema, que pertencem a grandes instituições e possuem dados pessoais dos usuários como e-mail, endereço e informações de pagamento (cartão de crédito, conta bancária, etc). Entretanto, como mencionado em Whitrow *et al.* (2009), as informações básicas de uma transação, conhecidas como informações brutas, não são suficientes para uma detecção eficiente de transações fraudulentas, pois deixaria de considerar informações importantes como o padrão de compra do usuário.

Muitos autores passaram a estudar como derivar mais informações para utilizar nos modelos antifraude, principalmente utilizando dados históricos de compras, processo conhecido como *feature engineering* ou derivação de características. Bolton e Hand (2001) mostraram o grande impacto do uso de *features* relacionadas ao histórico das transações em métodos de

detecção não supervisionados. Whitrow *et al.* (2009) foram pioneiros no estudo de *features* agregadas que interpretavam características associadas ao histórico de transações. No estudo avaliaram o uso de *features* criadas a partir de agregações de características das transações feitas em três diferentes janelas de tempo, 1, 3 e 7 dias em duas bases de dados pertencentes a diferentes instituições bancárias. Por exemplo, uma das *features* avaliadas foi a quantidade de compras feitas pelo usuário nas últimas 24 horas. Eles mostraram que a janela de tempo escolhida para a consolidação tem grande impacto na performance e que o uso dessas *features* agregadas melhorou em 28% a detecção de fraude feita com o algoritmo Floresta Aleatória. Jha, Guillen e Christopher Westland (2012) propõem uma melhoria na abordagem de Whitrow *et al.* (2009) em uma base de dados de transações de cartões de crédito de uma empresa do mercado, acrescentando informações agregadas de transações fraudulentas e não fraudulentas na geração das *features* para um modelo de regressão logística. Correa Bahnsen *et al.* (2016) se aprofunda mais no estudo de *features* agregadas utilizando uma combinação de critérios para agregação, por exemplo, no lugar de utilizarem apenas o valor gasto pelo usuário nas últimas 24 horas, eles especificaram ainda mais a característica utilizando o país e o grupo de lojas em que as compras foram realizadas. Mensurando o resultado em um banco de dados disponibilizado por uma empresa europeia de processamento de cartões, conseguiram uma melhora de 25% quando comparado com modelos que utilizam *features* com um grau de agregação menor.

Outra técnica bastante comum na geração de *features* é a utilização de variáveis categóricas ou binárias para a representação de uma característica da compra, por exemplo, se o usuário já é cliente da loja ou não. No trabalho de Fidel Beraldi (2014) pode-se encontrar uma grande variedade de *features* brutas, como o e-mail utilizado e o horário da compra, além de variáveis categóricas que representam características da transação, por exemplo, se o CPF de cadastro da conta é o mesmo do titular do cartão de crédito, e variáveis agregadas, como o valor médio das compras feitas no estabelecimento da transação.

Pela natureza do evento, devido aos constantes surgimentos de novas formas de cometer fraudes, é crescente o número de pesquisas em métodos semi-supervisionados para sua detecção. Uma das formas mais estudadas é a criação de *features* baseadas em modelos não supervisionados e a utilização de modelos supervisionados para a predição final. Carcillo *et al.* (2019) aplicam múltiplos modelos não supervisionados de detecção de *outliers* em diferentes granularidades, no histórico de compras, em clusters de usuários e com base em comportamentos globais, para a utilização dos scores como *features* de uma Floresta Aleatória. Apesar dos resultados não convincentes das abordagens do histórico de compras do cartão de crédito e a global, a abordagem de clusters teve um resultado promissor no aumento da acurácia da detecção. Lucas *et al.* (2020) faz a modelagem do valor das transações e do delta de tempo entre elas considerando as transações feitas nos cartões de crédito e nos terminais de compra. Além disso, os modelos foram criados para transações fraudulentas e não fraudulentas utilizando a técnica de Modelos Escondidos de Markov (HMM), resultando em oito modelos diferentes. Utilizando-os como entrada para diversos algoritmos de aprendizado de máquina, conseguiram uma melhora na

curva AUC de 9.3% para a base de dados do e-commerce e 18.1% para a base de transações presenciais.

O aumento no interesse de novas formas de derivar informações para os modelos antifraude é justificável, dado a complexidade de se prever o evento e os pontos positivos e negativos de cada abordagem. As formas mais comuns de detecção são a busca por mudanças no comportamento do usuário e o mapeamento de comportamentos ligados à fraude, em que a complexidade está no número de fatores que podem causar essas mudanças. Esses fatores podem ser exclusivos de um consumidor, por exemplo, quando um usuário adquire um novo celular e é esperado que exista uma mudança no dispositivo usado para fazer as transações, ou externos, como eventos promocionais, performance econômica ou uma nova modalidade de fraude. Os modelos com base em *features* brutas falham em captar as mudanças devido a sua falta de informação histórica e de padrões de fraudes globais, enquanto que os métodos de agregação podem colocar informações resumidas do histórico do usuário e dos padrões de fraudes já vistos. Entretanto, desconsideram a tendência temporal das transações e podem perder muita informações históricas, caso o intervalo de tempo considerado no agrupamento for pequeno, ou inserir muito ruído caso o intervalo de tempo considerado for muito longo (LUCAS *et al.*, 2020). Como alternativa, métodos mais complexos de modelagem, como o Modelos Escondidos de Markov (HMM), procuram resolver essas falhas, porém podem acrescentar problemas de performance computacionais e aumentar o tempo de execução do modelo. Além disso, a análise da perspectiva de transação por transação pode deixar de identificar padrões globais de comportamentos fraudulentos, principalmente os não-vistos, sendo assim analises de comportamentos gerais podem melhorar a detecção (JHA; GUILLEN; Christopher Westland, 2012). Dado as características de grupo da fraude, o compartilhamento de informações entre os fraudadores e a dinâmica de rede possuem grande impacto na predição e devem ser consideradas. Por exemplo, com o vazamento de uma base de dados de cartões de crédito, é esperado que as tentativas de fraudes com as informações aconteçam em diferentes lojas, de modo que o compartilhamento das tentativas de fraudes pode dificultar a ação dos fraudadores.

Devido ao funcionamento único de cada fluxo antifraude e o acesso a diferentes fontes de informações e tecnologias, é difícil generalizar um conjunto de *features* e técnicas a serem utilizadas, sendo recomendável o teste de diferentes informações e técnicas de engenharia de *features* no problema selecionado. Além disso, o surgimento de novos processos de venda, como a feita por intermediação de aplicativos de comunicação e redes sociais, e a evolução dos existentes tornam necessárias análises recorrentes de novas informações e a reavaliação das *features* utilizadas nos modelos. Por conta disso, os sistemas de avaliação de risco de fraude podem passar por constantes atualizações para manutenção do seu desempenho. O próximo capítulo descreve um caso de construção de um modelo de aprendizado de máquina com atualização dinâmica para uma base de dados com transações reais de uma empresa do comércio eletrônico brasileiro.

CAPÍTULO

4

METODOLOGIA E EXPERIMENTOS

4.1 Introdução

Ao longo deste capítulo, discutiremos a parte experimental do trabalho, que tem como objetivos a criação de um modelo de detecção à fraude a partir da base de dados de uma loja de e-commerce brasileira, e o estudo de diferentes técnicas de aprendizado com o modelo selecionado.

A base de dados utilizada possui 11.211.709 transações resultantes do processo de compra de cartão não presente da loja que passaram por um processo de avaliação de fraude. A loja escolhida possui parte relevante da quota do mercado *online* brasileiro, onde atua como plataforma de venda para diversos logistas. Por conta disso, a base de dados possui grande quantidade de produtos e perfis de compradores. O processo de avaliação utilizado nas transações é composto por um modelo de previsão, regras de decisão e análise manual. Além disso, a base possui 419.895 marcações de fraude, 3.745% do total de transações, oriundas dos processos de *chargeback* e marcação interna na análise manual das transações. No primeiro caso, a identificação da fraude é feita pelo dono do cartão de crédito, após a aprovação errada de uma transação fraudulenta e passa por longo processo até a loja ser sinalizada. Por conta disso, as marcações de fraude desse processo podem demorar até meses para serem notificadas e são conhecidas como notificações atrasadas. Já as marcações provenientes da análise manual acontecem em poucos dias a partir da análise de transações suspeitas, evitando-as que aconteçam. A seguir, apresentaremos uma análise mais detalhada sobre a base de dados e depois estudaremos a aplicação das técnicas supervisionadas. Por fim, estudaremos a performance do modelo ao longo do tempo e alternativas para lidar com o *concept drift*.

As 11.211.709 transações contidas na base de dados ocorreram entre os dias 01/07/2021 e 01/10/2021 e possuem uma tendência sazonal decorrente do comportamento de compra dos consumidores, como pode ser observado na Figura 12.

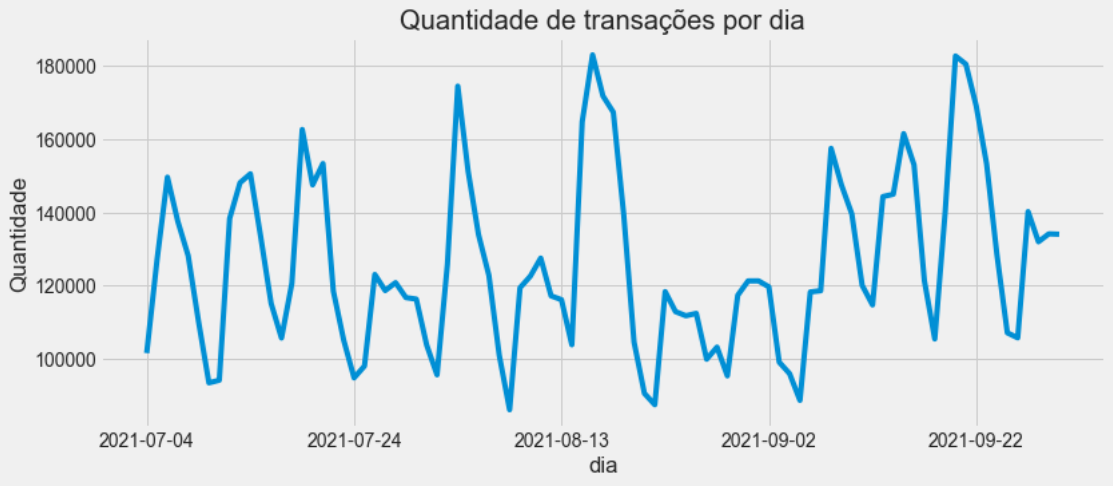

Figura 12 – Quantidade de transações por dia retirada da base de dados.

Como característico em bases de dados de fraude, existe desbalanceamento, sendo apenas 3.745% do total de transações fraudulentas.

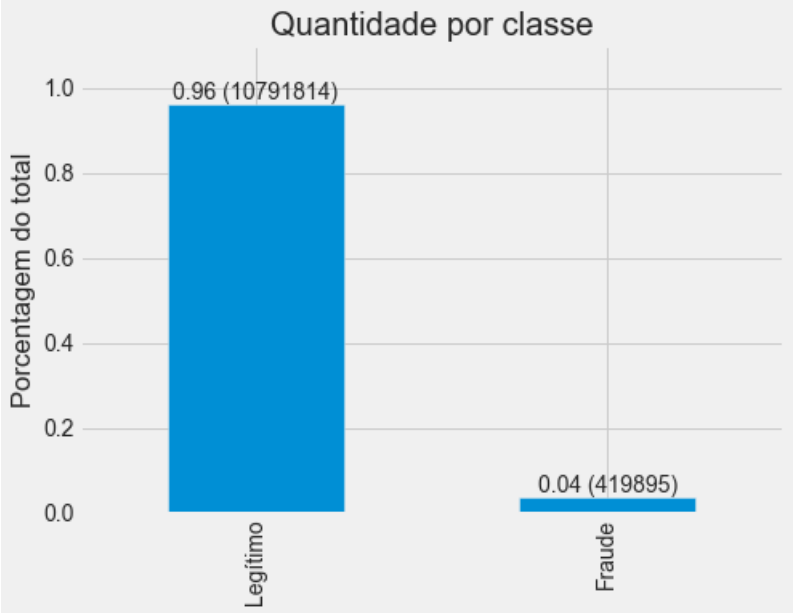

Figura 13 – Porcentagem (quantidade) de transações por categoria.

A porcentagem de fraude, também conhecida como exposição, por dia também parece ter certa sazonalidade. Entretanto, há um pico na segunda metade do mês 08/2021, com alguns dias atingindo uma taxa de fraude próxima de 7%.

Além da marcação de fraude, a base de dados possui 152 variáveis que podem ser utilizadas como *features*. Elas podem ser resumidas como:

- Categóricas: Informações que não assumem valor numérico. Por exemplo: Domínio de e-mail, cidade, CEP, etc.

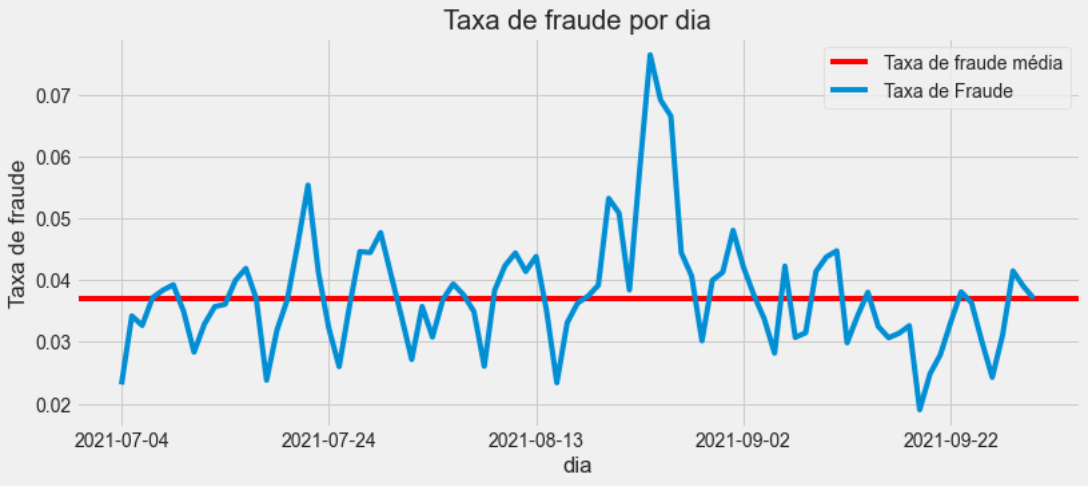

Figura 14 – Porcentagem de fraude por dia.

- Numéricas: Informações que assumem um valor mensurável. Por exemplo: Valor da compra, valor do frete, etc.

- Informações de rede: Variáveis derivadas a partir da caracterização das compras como um rede conectada por características das transações. Por exemplo, quantidade de cartões diferentes utilizados pelo e-mail da transação durante a última semana.

- Informações Agrupadas: Informações históricas de comportamento resumidas por alguma função de agregação. Por exemplo, quantidade de compras do CPF nos últimos 7 dias e 30 dias.

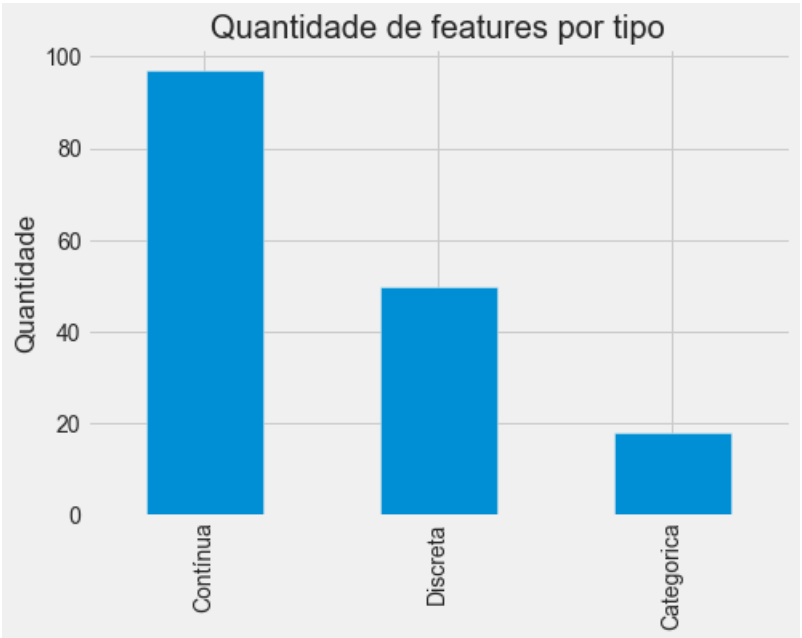

Figura 15 – Quantidade de *features* por tipo de dado.

Algumas variáveis possuem uma grande porcentagem de valores faltantes. As colunas que possuem mais de 40% dos dados faltantes não foram utilizadas na criação dos modelos. A tratativa dos valores faltantes das colunas que não foram retiradas depende do tipo de variável. Para variáveis categóricas, uma nova categoria de dados faltantes foi criada. Já para variáveis numéricas, as médias populacionais foram consideradas junto com a criação de uma variável binária indicando a falta de informação.

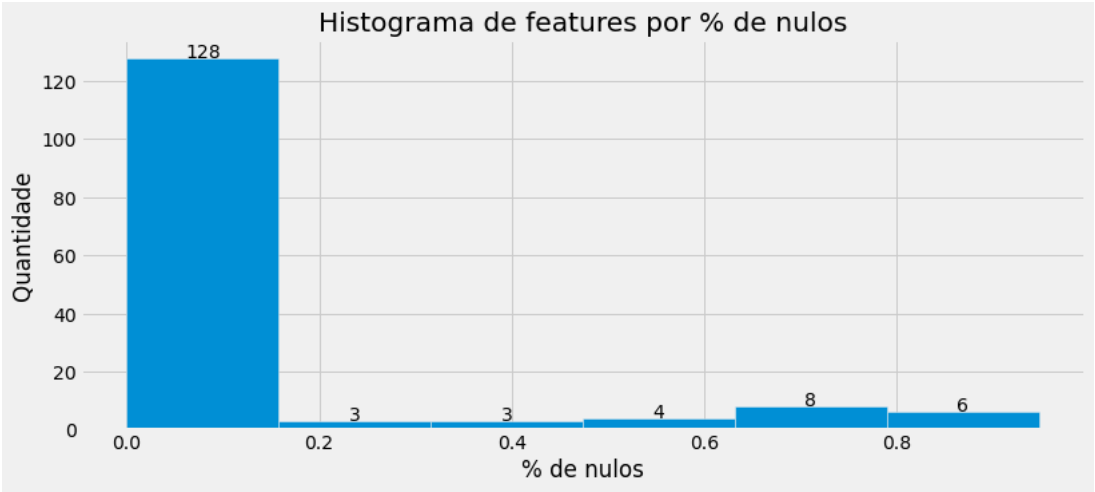

Figura 16 – Histograma de porcentagem de valores nulos por variável.

Como resumo do comportamento das variáveis contínuas utilizamos os gráficos *boxplots* separados pelos grupos de interesse (Figuras 17, 18 e 19). Os comportamentos são heterogêneos, com algumas *features* que apresentam variação entre os grupos, como a F39 e F58, e outras que não possuem grande variação, como F93 e F94. Além disso, algumas variáveis não possuem grande variação de valores, assumindo o valor médio para a grande maioria das transações, com isso, valores diferentes são representados como comportamento anômalos, como podemos observar nos gráficos das variáveis F111 e F117.

Para avaliação do comportamento das *features* categóricas, fizemos a visualização da porcentagem de fraude por porcentagem total de pedidos para cada uma das categorias (figura 20). Podemos notar que algumas variáveis possuem categorias que separam bem grupos com maior concentração de fraudes e possuem poucas categorias, como as de F0 até F5. Já outras variáveis possuem mais grupos e também parecem ser efetivas, como as F18 e F22. Existem também variáveis que parecem ser muito específicas e possuem muitos valores possíveis, criando grupos com poucas transações e diversos níveis de risco, como as F29 e F30. Variáveis categóricas que possuem mais do que 10 mil valores possíveis foram desconsideradas na criação dos modelos por serem muito específicas.

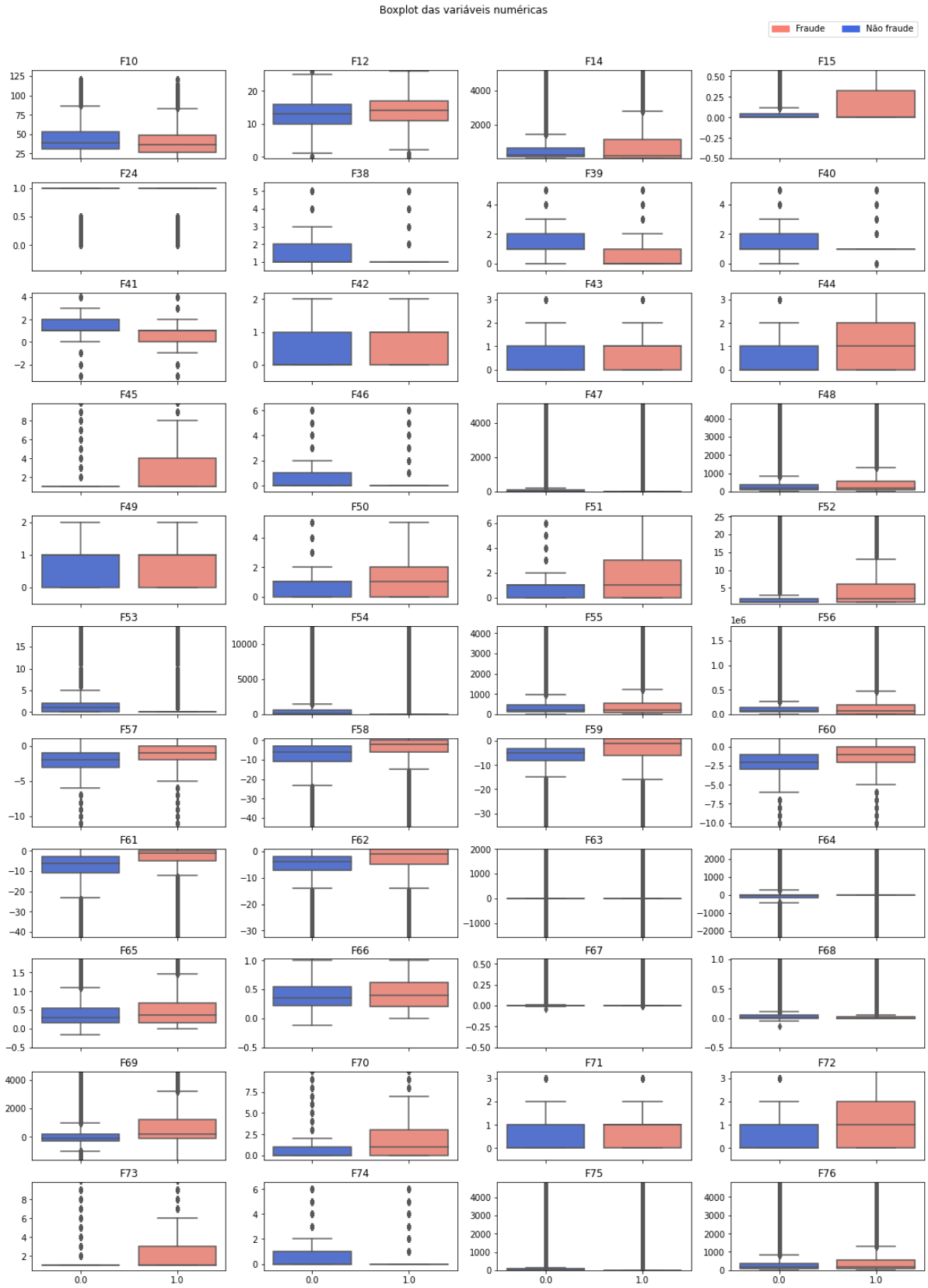

Figura 17 – *Boxplot* das variáveis numéricas encontrada na base de dados para cada uma das classes.

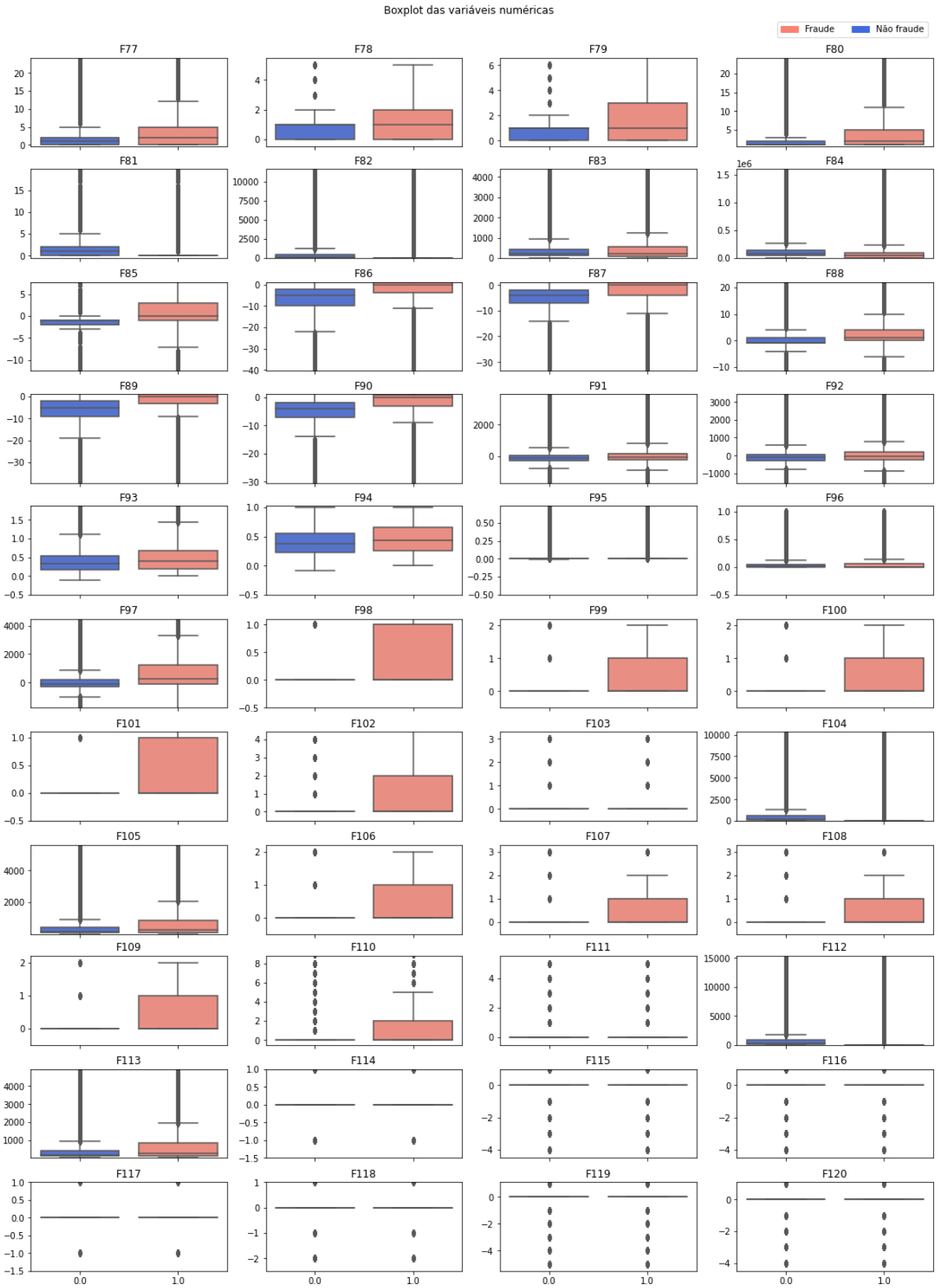

Figura 18 – *Boxplot* das variáveis numéricas encontrada na base de dados para cada uma das classes.

4.2 Comparação de modelos supervisionados

Nessa seção, faremos a comparação de performance de algoritmos de aprendizado de máquina supervisionados. A escolha pelos métodos supervisionados foi devido à grande presença

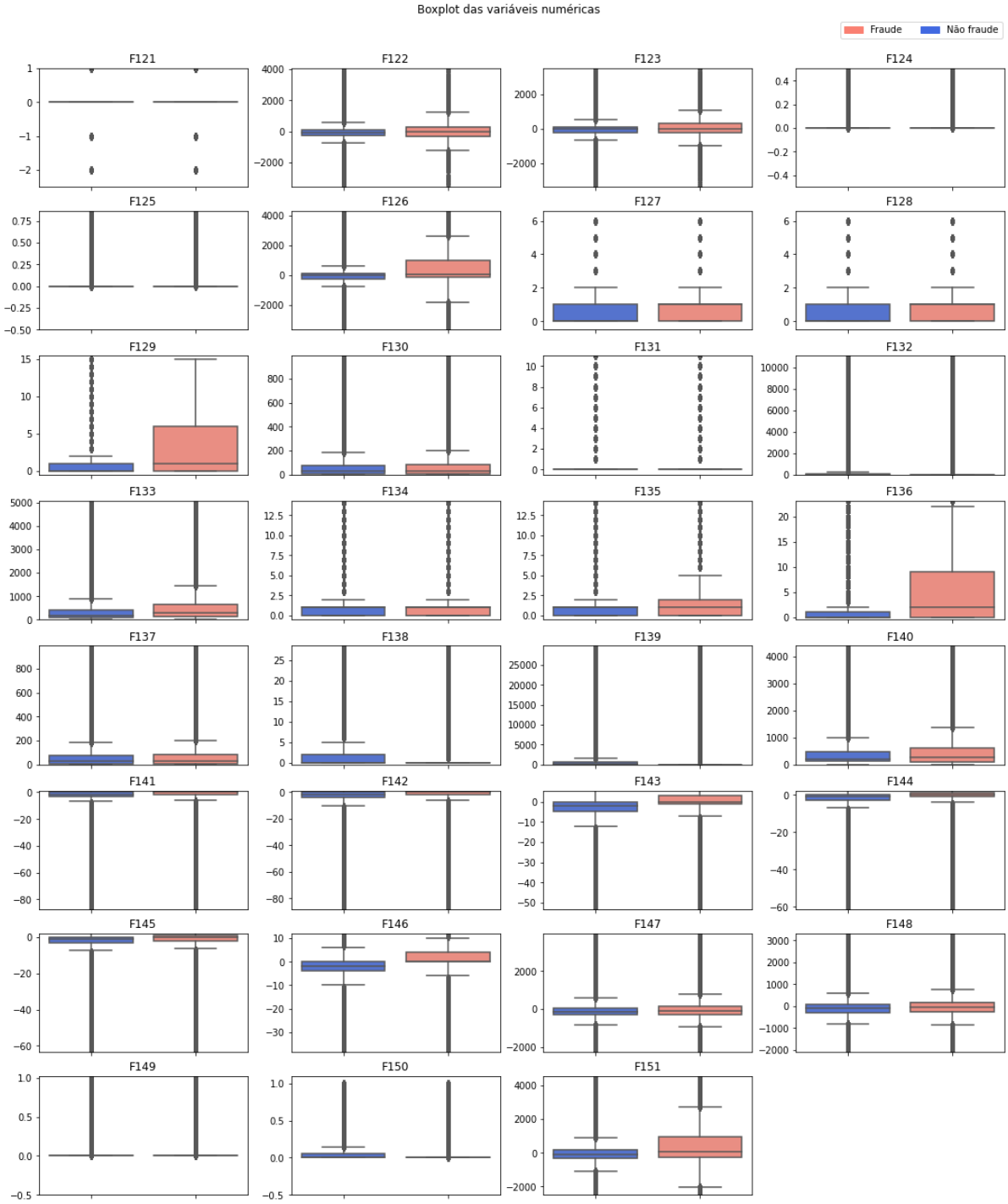

Figura 19 – *Boxplot* das variáveis numéricas encontrada na base de dados para cada uma das classes.

de variável resposta na base de dados. Nesse cenário, algoritmos supervisionados possuem performance superior aos não supervisionados (NIU; WANG; YANG, 2019) (CARCILLO *et al.*, 2019) por conseguirem aprender os padrões de fraude já conhecidos. A combinação de ambas técnicas também foi considerada por conseguir ter impacto significativo no desempenho (CARCILLO *et al.*, 2019) (Dal Pozzolo, 2015), porém, devido aos desafios da utilização de técnicas como *Peer Group Analysis* para o mapeamento de perfis de consumidores para base

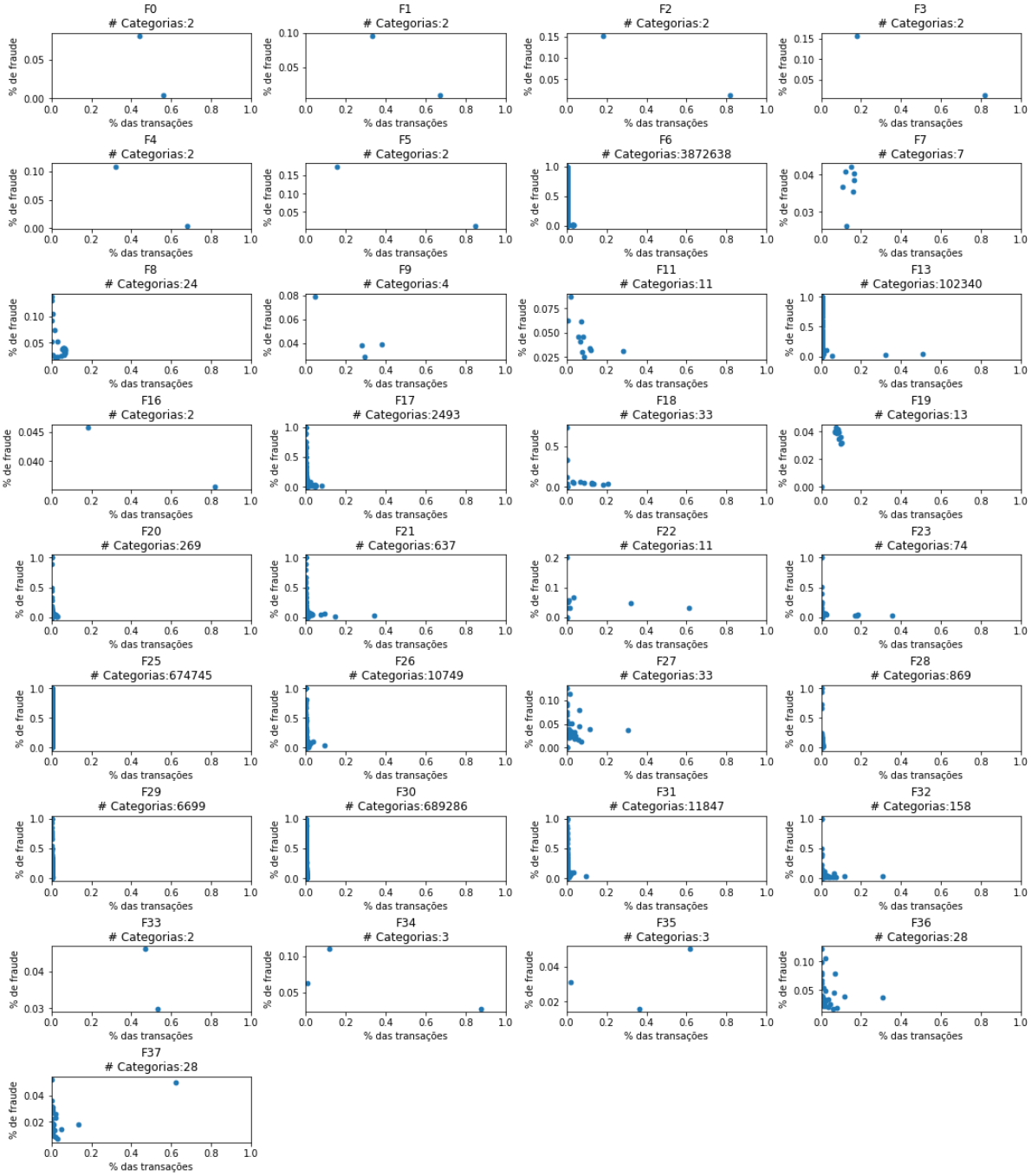

Figura 20 – Gráfico de pontos para cada categoria das variáveis. O eixo *x* representa a porcentagem do total de transações que estão na categoria e o eixo *y* é a porcentagem de fraude relativa, isto é, a proporção das transações da categoria que são fraude.

de dados grandes e a quantidade de variáveis já disponíveis para os modelos supervisionados, optamos por seguir com os métodos supervisionados e deixar a abordagem semi-supervisionada para trabalhos futuros.

Os algoritmos utilizados serão Máquina de vetores de suporte (SVM), Regressão Logística (RL), Rede Neural (RN) e Florestas Aleatórias (RF) presentes na biblioteca *Scikit Learn* da linguagem de programação *Python*, e o algoritmo *Gradient Boosting Decision Tree* (LGBM) do

framework LightGBM (MICROSOFT, 2022). Para comparação foi feita uma divisão temporal dos dados de treinamento e teste devido à natureza estocástica dos padrões de fraude. Sendo assim, a base de dados foi particionada em cinco lotes temporais de acordo com a Figura 21, onde os lotes $Lote_i$ $i \in (1,2,3,4)$ foram usados como períodos de treinamento dos modelos, que foram testados no lote imediatamente posterior, totalizando 4 treinos para cada algoritmo.

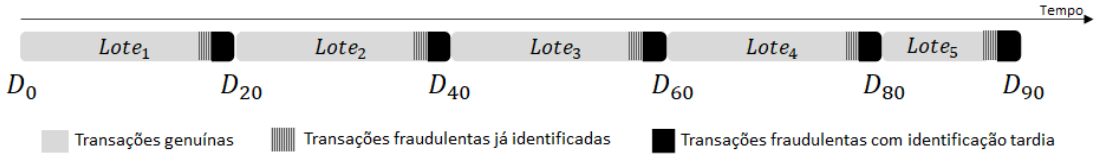

Figura 21 – Divisão da base de dados em lotes.

Além disso, para os períodos de treino, consideramos apenas as informações obtidas até sua respectiva data, de tal modo que transações com notificação de fraude posterior a data final do lote não foram consideradas como fraudulentas. Desse modo, evitamos informação futura nos treinamentos dos modelos. Para avaliação no período de teste, todas as notificações foram consideradas.

Antes do treinamento dos algoritmos, na etapa de pré-processamento de cada lote, os seguintes passos foram feitos:

- Remoção de variáveis com mais que 40% de dados faltantes.
- Atribuição de categoria específica de dados faltantes para as variáveis categóricas.
- Atribuição do valor médio para os dados faltantes das variáveis numéricas.
- Transformação das variáveis categóricas nos pesos de evidência de cada categoria.
- Normalização das variáveis numéricas.

Para encontrar os melhores hiperparâmetros de cada algoritmo, utilizamos o método *GridSearchCV* da biblioteca *sklearn*, que implementa uma busca em grade no espaço de parâmetros de interesse.

Como vimos nas seções anteriores, escolher uma boa medida para avaliação de modelos de detecção de fraudes muitas vezes não é uma tarefa trivial (ver seção 3.7.1). Dentre as métricas estudadas na seção 3, estão as que dependem das distribuições de fraude à posteriori e aquelas que são definidas a partir de um ponto de corte. Em casos específicos, a adoção de métricas sob um ponto de corte determinado pode ser interessante. Por exemplo, caso exista uma meta de 95% de aprovação nas transações da loja, é razoável que a avaliação da predição seja feita no ponto de corte associado à meta de interesse. Entretanto, para casos de uso genéricos, as métricas populacionais consideram todos possíveis pontos de corte, portanto, refletem melhor o

desempenho do classificador em diferentes cenários de uso. Por esses motivos, optamos por usar as métricas de área sob a curva roc (AucRoc) e a estatística KS (KS).

Os modelos baseados em árvores de decisão foram os que obtiveram as melhores performance, sendo que o modelo LGBM superou a performance da RF por uma pequena diferença em ambas as métricas para todos os lotes utilizados como período de teste, como podemos ver na Tabela 3.

Tabela 3 – Métricas do período de teste para cada lote.

	AucRoc					KS				
Lote	LGBM	LR	RF	RN	SVM	LGBM	LR	RF	RN	SVM
2	0,979	0,967	0,977	0,975	0,954	0,852	0,816	0,844	0,836	0,779
3	0,974	0,954	0,970	0,967	0,938	0,829	0,769	0,814	0,804	0,729
4	0,978	0,967	0,977	0,970	0,955	0,850	0,815	0,843	0,827	0,787
5	0,977	0,962	0,973	0,956	0,943	0,844	0,792	0,834	0,783	0,748

Quando avaliamos a performance média para os conjuntos de treino e teste, disponíveis na Tabela 4, a RF teve melhor performance no conjunto de treinamento para ambas as métricas, mas perdeu performance na generalização para o período de teste. Já o modelo LGBM teve uma performance um pouco pior do que a RF no conjunto de treinamento, mas conseguiu melhor generalização, alcançando as melhores médias no conjunto de dados de teste. O SVM foi o algoritmo que teve pior desempenho, pelo tempo de processamento somente foi testado o núcleo linear, outros núcleos podem melhorar a performance do classificador, mas exigem um alto tempo computacional para convergência. A média das métricas para os conjuntos de treinamentos e testes estão na tabela 4.

Tabela 4 – Média das métricas obtidas nos lotes de treino e teste.

Modelo	Treino		Teste	
	AucRoc	KS	AucRoc	KS
LGBM	0,9853	0,8762	**0,9770**	**0,8438**
RF	**0,9898**	**0,8987**	0,9743	0,8337
RN	0,9765	0.8399	0,9670	0,8124
RL	0,9730	0.8274	0,9625	0,7978
SVM	0,9608	0.7941	0,9475	0,7606

Pela tabela 3, também podemos notar que todos os algoritmos tiveram a sua menor performance no $Lote_3$. Como podemos ver pela métrica KS, a discriminação do evento no conjunto de *features* utilizado pelos algoritmos durante esse período foi pior, o que pode ser sinal de mudança no comportamento dos consumidores e dos padrões de fraude. Por exemplo, o algoritmo RF teve, em média, estatística KS menor em 2,5 unidades nesse lote. Na próxima seção, estudaremos o desempenho do método de melhor performance, LGBM, ao longo do

tempo, a fim de verificar sua estabilidade e o efeito de atualizações semanais no desempenho do algoritmo.

4.3 Modelos dinâmicos

Nesta seção, implementaremos a metodologia de aprendizado *online* proposta por (Dal Pozzolo, 2015) e descrita na Seção 3.6 a fim de entender o efeito da atualização dinâmica do modelo na performance. A metodologia foi escolhida por ser independente do algoritmo de previsão utilizado e de fácil implementação no atual sistema antifraude da loja.

Para analisar o efeito do aprendizado *online* em nossos dados, iremos comparar um cenário de modelo estático com outros baseados no aprendizado de modelos semanais. Para a criação do modelo estático, utilizaremos as três primeiras semanas da base de dados para treinamento do algoritmo LGBM, como esquematizado na Figura 22. Já nos casos dinâmicos, vamos utilizar lotes semanais para atualização do mesmo algoritmo. No cenário mais simples, representado na Figura 23, a predição da semana i é feita pelo modelo treinado no lote da semana anterior, $i-1$. Nos demais casos, a predição da semana i é realizada a partir de um *ensemble* com os modelos das $j \in 2,3,4$ semanas anteriores, a Figura 24 exemplifica o caso em que $j = 3$. Chamaremos o cenário estático de ME, o de modelos semanais de MS e os cenários com *ensembles* de E_j. Ainda, nos cenários dinâmicos, MS_i e E_{ji}, representam, respectivamente, o modelo treinado com o lote da semana $i \in [28,...,40]$ e o *ensemble* com j modelos feitos a partir de $MS_{i-j},...,MS_i$. Note que o cenário MS pode ser visto como um caso particular de *ensemble* com um único modelo, optamos pela diferenciação por motivos conceituais.

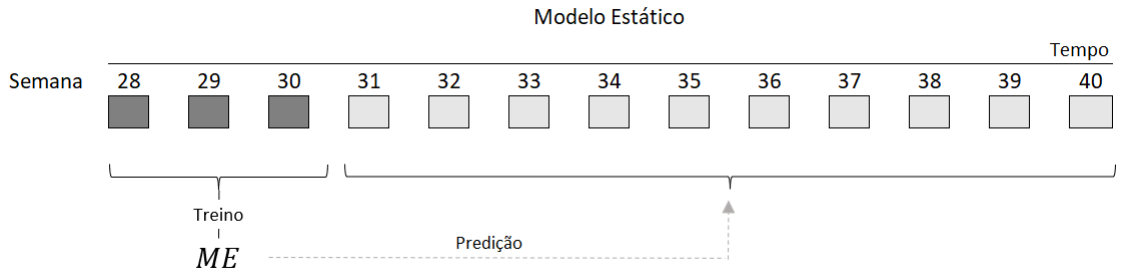

Figura 22 – Cenário Estático (ME): O modelo ME é treinado com os dados das três primeiras semanas (lotes) da base de dados e utilizado para predição dos lotes seguintes.

A escolha do lote semanal se deu para o acúmulo de uma quantidade suficientemente grande e diversa de informação da classe minoritária e pela sazonalidade semanal do comportamento de compras e das tentativas de fraudes.

Para as métricas de comparação dos cenários, usaremos a estatística KS, que se mostrou mais volátil para a comparação dos modelos do que a AucRoc no estudo anterior, e o indicador de *chargeback* obtido no percentil 90 da distribuição a posteriori, ou seja, cenário que simula uma meta de aprovação de 90% das transações. O indicador de *chargeback* foi escolhido por ser

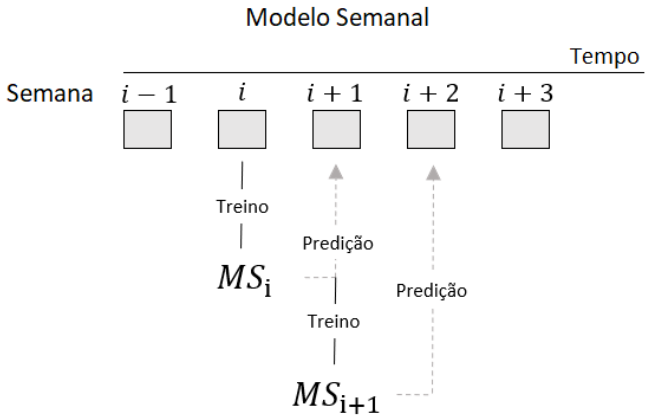

Figura 23 – Cenário Semanal (*MS*): A cada semana i o modelo MS_i é retreinado com os dados mais recentes e utilizado para predição do lote seguinte.

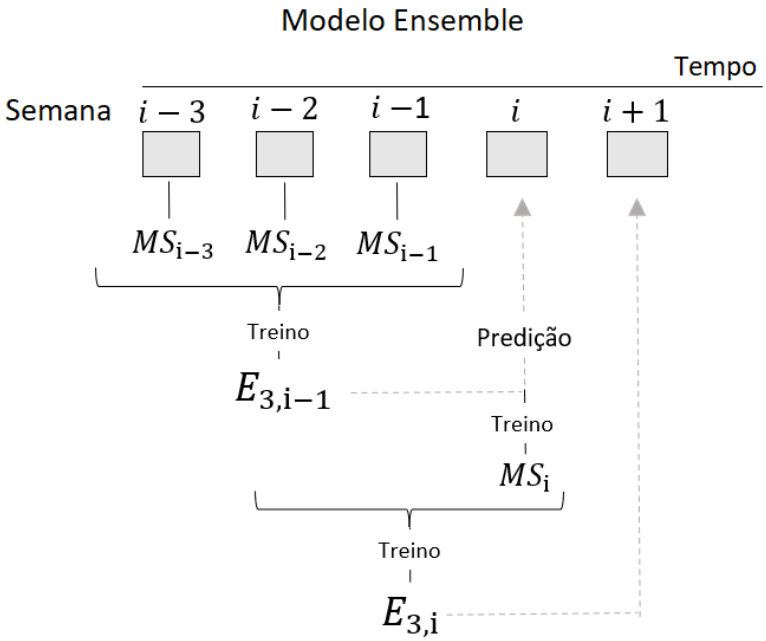

Figura 24 – Exemplo do cenário *Ensemble* (E_3): A cada semana i forma-se um novo modelo E_i com a composição dos três últimos modelos semanais MS_{i-2}, MS_{i-1} e MS_i criados para a predição da semana seguinte.

um dos principais motivadores para a criação de regras de negócio e para representação de uma métrica financeira acompanhada pelas equipes de negócio.

Considerando os resultados para a estatística KS demonstrados na Figura 25, podemos observar que o modelo estático obteve desempenho superior apenas nas duas primeiras semanas após o período de treinamento. O cenário *MS* foi melhor do que o *ME* em metade das semanas analisadas, porém superou o *ensemble* em apenas uma das semanas. Na comparação entre os modelos dinâmicos, existiram diferenças significativas apenas na comparação entre o *MS* e E_2, nos demais casos, a consideração de mais semanas não impactou significativamente

o desempenho. Na maioria das semanas, o ganho de desempenho dos modelos aprendizado *online* não é grande, ficando abaixo de 1%. Entretanto, nas semanas 34 e 35, as diferenças de desempenho se mostram maiores, atingindo até 6%.

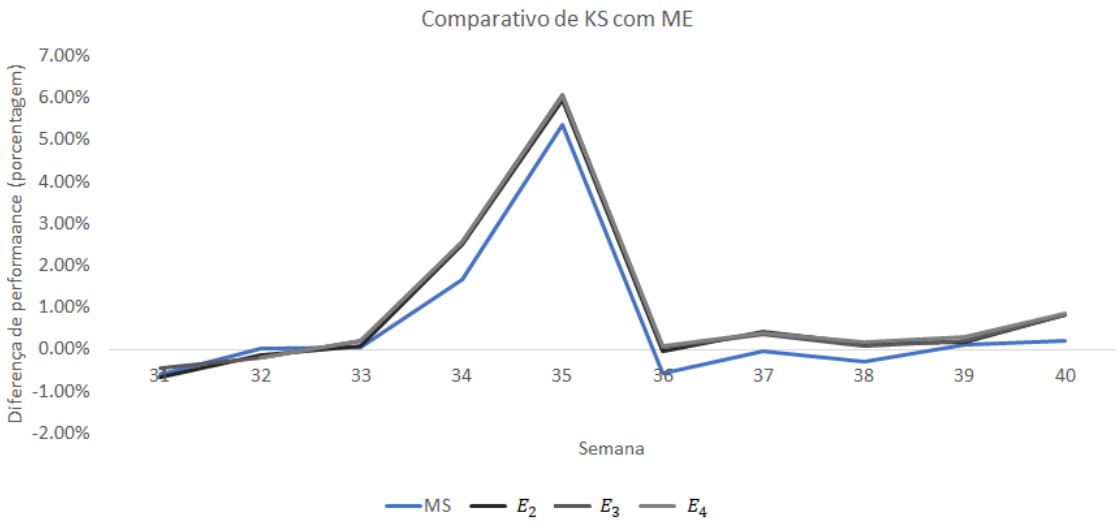

Figura 25 – Comparativo de performance da estatística KS tomando como referência o modelo estático. O cálculo da variação foi feito utilizando $(KS - KS_{ME})/KS_{ME}$.

Como podemos observar na Figura 26, as semanas que acontecem as maiores diferenças de performance pela estatística KS coincidem com a queda de performance do modelo estático, ou seja, são as semanas que o ME atingiram o mínimo de seu desempenho, o que pode estar associado com uma mudança no comportamento da fraude. No cenário de *concept drift*, os modelos dinâmicos foram menos afetados e conseguiram manter desempenho parecido com outras semanas, o que causou maiores diferenças em relação ao modelo estático.

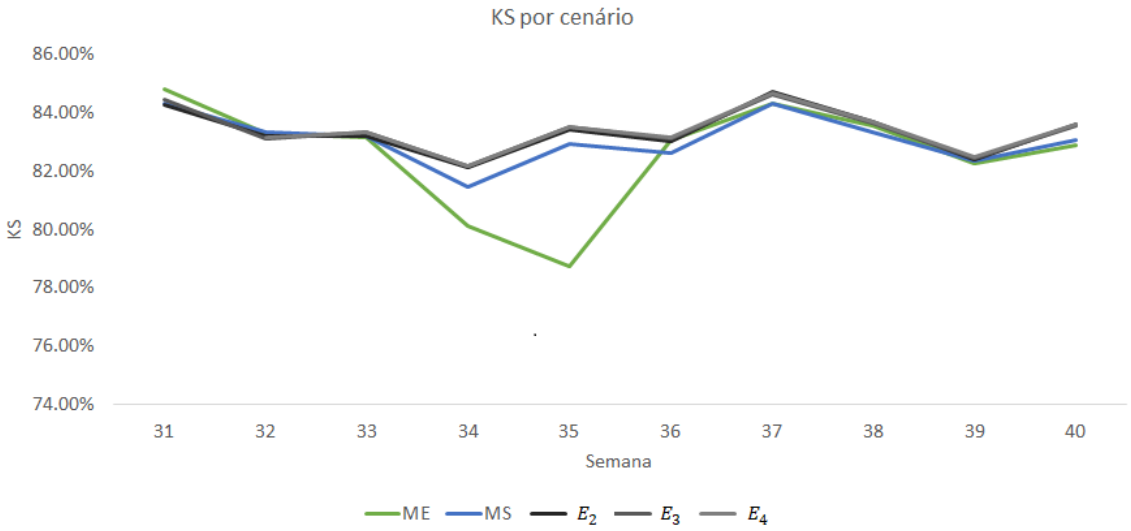

Figura 26 – Estatística KS por cenário de aprendizado

Em relação ao índice de *chargeback* para o percentil 90, as diferenças de desempenhos

em relação ao modelo estático tem um comportamento parecido, porém, o modelo de melhor performance varia mais entre as semanas. Nas semanas 34 e 35, onde temos a suspeita de *concept drift*, devido à queda de performance do modelo estático, os modelos dinâmicos tiveram uma performance até 30% melhor, enquanto que nas outras semanas as diferenças de performance ficaram, em sua maioria, em até 5%.

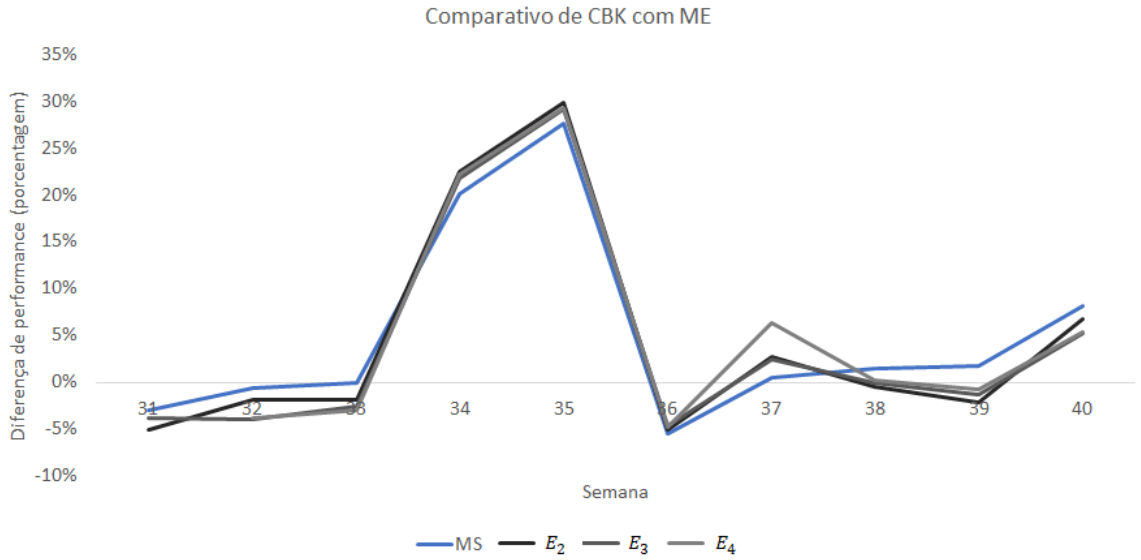

Figura 27 – Comparativo de performance do índice de *chargeback* para o percentil 90 da distribuição a posteriori tomando como referência o modelo estático. O cálculo da variação foi feito utilizando $(IC_{ME} - IC)/IC_{ME}$, onde IC é o índice de *chargeback*.

CAPÍTULO

5

CONCLUSÃO

5.1 Conclusão e trabalhos futuros

A dinâmica das tentativas de fraudes de cartão de crédito é um ambiente de elevada complexidade devido a quantidade de fatores que podem impactar sua avaliação, as mudanças no padrão de fraude e os ruídos na variável resposta. Nesse contexto, o presente trabalho utilizou-se de uma base de transações reais para avaliar a performance de diferentes algoritmos e estudar o impacto do *concept drift* e do aprendizado *online* no cenário de fraude disponível.

Na avaliação dos algoritmos, podemos observar que os métodos baseados em árvores de decisão obtiveram melhores resultados para o conjunto de transações avaliadas em um contexto de divisão temporal entre dados de treino e teste, sendo o LGBM o algoritmo com melhor performance em ambas métricas avaliadas, AucRoc e estatística KS. Esses métodos conseguiram captar melhor a relação não linear do evento no conjunto de variáveis disponíveis, enquanto que o SVM utilizando núcleo linear foi o pior em capturar a tendência não linear do evento e teve pior desempenho.

A presença de viés no fluxo de definição da variável resposta é um fator que tem grande impacto no desempenho dos modelos de prevenção à fraude. A presença da análise manual no fluxo possui grande benefício de aumentar a quantidade de fraudes encontradas, porém o resultado da análise está sujeito aos vieses conscientes e inconscientes da pessoa que a faz, principalmente em relação aos fatores socioeconômico do comprador e o padrão de fraude recente. Além disso, a falta de sinalização de fraudes de baixo valor e a reprovação de transações também possuem grande impacto. A presença de marcações incorretas tornam a seleção das transações de treino e mensuração de desempenho do classificador complexa e mudam o comportamento dos compradores, por exemplo, fazendo com que bons compradores sejam reprovados devido à uma suspeita falta de fraude.

Para avaliação do aprendizado online, aplicamos um método simplificado do proposto

no trabalho de Dal Pozzolo *et al.* (2018), com a criação de três cenários de aprendizado: modelo estático, modelos semanais e aprendizado *ensemble* utilizando-se dos modelos das últimas semanas. Pelo comportamento de performance do modelo estático, podemos perceber uma diferença entre o comportamento do *drift* entre a base de dados estudada e os dados utilizados em Dal Pozzolo *et al.* (2018). Os dados utilizados pelo autor possuem um *drift* mais constante ao longo dos lotes de dados avaliados, enquanto que os dados das transações estudadas no presente trabalho são mais estáveis nas semanas com mudanças bruscas em dois dos lotes semanais.

Em perspectiva da estatística KS, os modelos de aprendizado *online* se mostraram superiores. De forma geral, os modelos *ensembles* tiveram melhor desempenho mesmo quando comparado com os modelos semanais. Entretanto, a consideração de mais de dois modelos na construção dos *ensembles* não resultou em ganho significativo de desempenho. A diferença de performance é pequena nas semanas em que não houveram grandes variações no padrão de fraude e aumenta na presença do *concept drift*, quando o modelo estático perde performance. Na avaliação feita considerando um indicador acompanhada pela equipe de negócio, o índice de *chargeback* obtido para o corte no percentil 90, o ganho de performance dos modelos *online* ficam menos evidentes nas semanas sem *concept drift*, com os modelos alternando o posto de melhor performance durante os lotes. Nas semanas com presença de *concept drift*, os modelos de aprendizado *online* conseguem se adaptar mais rapidamente ao novo padrão de fraude e os ganhos ficam mais evidentes, chegando a uma redução nos gastos com *chargeback* entre 20% e 30%.

As diferenças de avaliação dos cenários de acordo com a métrica de performance mostram um pouco das dificuldades em se decidir uma métrica única de performance para avaliação de modelos antifraudes. Enquanto que o cenário de *ensemble* se mostra constantemente melhor nas semanas quando a avaliação é feita utilizando uma métrica populacional, o KS, o ganho em relação aos outros cenários é menos evidente utilizando uma métrica dependente de um ponto de corte. Por outro lado, o indicador de *chargeback* demonstra melhor a queda de performance do modelo estático nas semanas com alta presença de *concept drift*.

Ao longo do desenvolvimento do trabalho, identificamos algumas oportunidades de evolução na metodologia estudada que podem ser temas de pesquisas futuras. Primeiro, podemos citar a convergência das métricas utilizadas pelas áreas de negócio e as funções custo dos modelos estatísticos no aprendizado *online*. Apesar de já existirem trabalhos de utilização de métricas financeiras no treinamento de modelos para fraudes, o seu ganho no aprendizado *online* ainda é pouco claro. Outro ponto que pode ser explorado é a comparação de diferentes técnicas de aprendizado *online*. A metodologia de *ensemble* obteve resultados interessantes, mas fazem sua atualização semanalmente, enquanto que a utilização de outras técnicas pode reduzir o tempo de ajuste do sistema aos novos padrões de fraude e mantê-lo atualizado aos padrões de ataque mais recentes, principalmente em cenários de *concept drift* mais frequentes do que o encontrado na loja. Ademais, a consideração de modelos não supervisionados ou

semi supervisionados no aprendizado *online* podem ser úteis para melhorar o desempenho do classificador. Como próximos passos, em termos práticos, pretendemos continuar o estudo do desempenho para um período maior de transações e em diferentes lojas, com o objetivo de entender mais profundamente os diferentes tipos de *concept drift* presentes nos padrões de fraude, sua relação com os componentes do sistema antifraude e com a performance da metodologia.

REFERÊNCIAS

ABDALLAH, A.; MAAROF, M. A.; ZAINAL, A. Fraud detection system: A survey. **Journal of Network and Computer Applications**, v. 68, p. 90–113, 2016. ISSN 10958592. Citado nas páginas 21, 23, 30, 37, 38 e 40.

ALBERTIN, A. L. Comércio eletrônico - um estudo no setor bancário. In: . [s.n.], 1998. Disponível em: <http://www.anpad.org.br/admin/pdf/enanpad1998-ai-03.pdf>. Citado nas páginas 21 e 25.

____. Comércio eletrônico: modelo, aspectos e contribuições de sua aplicação. **Revista de Administração de Empresas**, v. 40, p. 108–115, 2000. Citado na página 25.

ALESKEROV, E.; RAO, B. : A Neural Network Based Database stern for Credit Card Fraud Detection. **Signal Processing**, p. 220–226, 1997. Citado na página 39.

B. Baesens; V. Van Vlasselaer; W. Verbeke. **Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection**. [S.l.: s.n.], 2015. v. 1. Citado nas páginas 15, 40, 42, 43, 44, 45, 46 e 47.

BHATTACHARYYA, S.; JHA, S.; THARAKUNNEL, K.; WESTLAND, J. C. Data mining for credit card fraud: A comparative study. **Decision Support Systems**, Elsevier B.V., v. 50, n. 3, p. 602–613, 2011. ISSN 01679236. Disponível em: <http://dx.doi.org/10.1016/j.dss.2010.08.008>. Citado na página 40.

BISHOP, C. M. **Pattern Recognition and Machine Learning (Information Science and Statistics)**. 1. ed. [S.l.]: Springer, 2007. ISBN 0387310738. Citado na página 41.

BOLTON, R. J.; HAND, D. J. Unsupervised Profiling Methods for Fraud Detection. **Proc. Credit Scoring and Credit Control VII**, p. 5–7, 2001. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.24.5743>. Citado na página 54.

BRYAN, T.; DUBOIS, F.; HAGEN, A.; HUGHES, M.; KOVAL, K. .; LAFRENCE, J.; WEIMAR, L.; ZIRKLE, A. Card-Not-Present Fraud around the World. **U.S. Payments Forum**, v. 1, n. March, p. 1–35, 2017. Disponível em: <https://www.uspaymentsforum.org/wp-content/uploads/2017/03/CNP-Fraud-Around-the-World-WP-FINAL-Mar-2017.pdf>. Citado na página 31.

CARCILLO, F.; Le Borgne, Y. A.; CAELEN, O.; BONTEMPI, G. Streaming active learning strategies for real-life credit card fraud detection: assessment and visualization. **International Journal of Data Science and Analytics**, v. 5, n. 4, p. 285–300, 2018. ISSN 23644168. Citado nas páginas 46 e 47.

CARCILLO, F.; Le Borgne, Y. A.; CAELEN, O.; KESSACI, Y.; OBLÉ, F.; BONTEMPI, G. Combining unsupervised and supervised learning in credit card fraud detection. **Information Sciences**, Elsevier Inc., n. xxxx, 2019. ISSN 00200255. Disponível em: <https://doi.org/10.1016/j.ins.2019.05.042>. Citado nas páginas 22, 35, 39, 40, 46, 55 e 63.

CARDOSO, S.; KAWAMOTO, M. H.; MASSUDA, E. M. Comércio Eletrônico: O Varejo Virtual Brasileiro. **Revista Cesumar – Ciências Humanas e Sociais Aplicadas**, v. 24, n. 1, p. 117, 2019. ISSN 1516-2664. Citado nas páginas 17, 25 e 26.

CLEARSALE. **Fraud Risk: What's the Global Impact?** 2018. Disponível em: <https://blog.clear.sale/fraud-risk-whats-the-global-impact>. Citado na página 31.

____. **E-commerce brasileiro sofre mais de R$ 3,6 mil em tentativas de fraude por minuto**. 2020. Disponível em: <https://www.ecommercebrasil.com.br/noticias/e-commerce-brasileiro-sofre-mais-de-r-36-mil-em-tentativas-de-fraude-por-minuto/>. Citado nas páginas 21 e 31.

Correa Bahnsen, A.; AOUADA, D.; STOJANOVIC, A.; OTTERSTEN, B. Feature engineering strategies for credit card fraud detection. **Expert Systems with Applications**, Elsevier Ltd, v. 51, p. 134–142, 2016. ISSN 09574174. Disponível em: <http://dx.doi.org/10.1016/j.eswa.2015.12.030>. Citado nas páginas 39 e 55.

CRESSEY, D. R. **Other People's Money: A Study in the Social Psychology of Embezzlement**. [S.l.]: Glencoe, 1953. Citado na página 30.

Dal Pozzolo, A. Adaptive Machine Learning for Credit Card Fraud Detection. n. December, 2015. Disponível em: <http://www.ulb.ac.be/di/map/adalpozz/pdf/Dalpozzolo2015PhD.pdf>. Citado nas páginas 39, 45, 48, 49, 63 e 67.

Dal Pozzolo, A.; BORACCHI, G.; CAELEN, O.; ALIPPI, C.; BONTEMPI, G. Credit card fraud detection: A realistic modeling and a novel learning strategy. **IEEE Transactions on Neural Networks and Learning Systems**, IEEE, v. 29, n. 8, p. 3784–3797, 2018. ISSN 21622388. Citado nas páginas 29, 53, 54 e 72.

eMarketer. **Brazil Ecommerce by Category Forecast 2022**. 2022. Disponível em: <https://www.insiderintelligence.com/content/spotlight-brazil-ecommerce-by-category-forecast-2022>. Citado na página 21.

____. **Global Ecommerce 2022**. 2022. Disponível em: <https://www.insiderintelligence.com/content/worldwide-ecommerce-forecast-update-2022>. Citado na página 21.

FANG, F.; CHEN, Y. A new approach for credit scoring by directly maximizing the Kolmogorov–Smirnov statistic. **Computational Statistics and Data Analysis**, Elsevier B.V., v. 133, p. 180–194, 2019. ISSN 01679473. Disponível em: <https://doi.org/10.1016/j.csda.2018.10.004>. Citado na página 51.

Fidel Beraldi. **Atualização dinâmica de modelo de regressão logística binária para detecção de fraudes em transações eletrônicas com cartão de crédito**. Tese (Doutorado), 2014. Citado nas páginas 15, 22, 26, 27, 29, 35, 39, 48 e 55.

GIANINI, G.; Ghemmogne Fossi, L.; MIO, C.; CAELEN, O.; BRUNIE, L.; DAMIANI, E. Managing a pool of rules for credit card fraud detection by a game theory based approach. **Future Generation Computer Systems**, v. 102, p. 549–561, 2020. ISSN 0167-739X. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0167739X18317151>. Citado na página 36.

HOI, S. C.; SAHOO, D.; LU, J.; ZHAO, P. Online learning: A comprehensive survey. **Neurocomputing**, v. 459, p. 249–289, 2021. ISSN 18728286. Citado na página 48.

HUBER, D. Forensic Accounting, Fraud Theory, and the End of the Fraud Triangle. **Journal of Theoretical Accounting Research**, v. 12, n. 2, p. 28–48, 2017. Citado na página 30.

JANSSON, M.; AXELSSON, M. Federated Learning Used to Detect Credit Card Fraud. **Lu-Cs-Ex**, 2020. Disponível em: <https://lup.lub.lu.se/student-papers/record/9024753/file/9024763.pdf>. Citado na página 23.

JHA, S.; GUILLEN, M.; Christopher Westland, J. Employing transaction aggregation strategy to detect credit card fraud. **Expert Systems with Applications**, Elsevier Ltd, v. 39, n. 16, p. 12650–12657, 2012. ISSN 09574174. Disponível em: <http://dx.doi.org/10.1016/j.eswa.2012.05.018>. Citado nas páginas 55 e 56.

KUMARI, N.; KANNAN, S.; MUTHUKUMARAVEL, A. Credit card fraud detection using Hidden Markov Model-A survey. **Middle - East Journal of Scientific Research**, v. 20, n. 6, p. 697–699, 2014. ISSN 19909233. Citado nas páginas 36, 40 e 46.

LI, J.; LIU, Y.-w.; JIA, Y.; NANDURI, J. Discriminative Data-driven Self-adaptive Fraud Control Decision System with Incomplete Information. **Decision Support Systems**, n. Junxuan Li, 2018. Disponível em: <http://arxiv.org/abs/1810.01982>. Citado nas páginas 36 e 39.

LUCAS, Y.; PORTIER, P. E.; LAPORTE, L.; HE-GUELTON, L.; CAELEN, O.; GRANITZER, M.; CALABRETTO, S. Towards automated feature engineering for credit card fraud detection using multi-perspective HMMs. **Future Generation Computer Systems**, Elsevier B.V., v. 102, p. 393–402, 2020. ISSN 0167739X. Disponível em: <https://doi.org/10.1016/j.future.2019.08.029>. Citado nas páginas 40, 55 e 56.

MICROSOFT. **LightGBM Release 3.3.2**. 2022. Disponível em: <https://lightgbm.readthedocs.io/_/downloads/en/v3.3.2/pdf/>. Citado na página 65.

MILO, T.; NOVGORODOV, S.; TAN, W.-C. Rudolf: Interactive rule refinement system for fraud detection. **Proc. VLDB Endow.**, VLDB Endowment, v. 9, n. 13, p. 1465–1468, sep 2016. ISSN 2150-8097. Disponível em: <https://doi.org/10.14778/3007263.3007285>. Citado na página 36.

MoIP Pagamentos. **Cartão de crédito corresponde a mais de 87% dos pagamentos on-line**. 2010. Disponível em: <https://www.infomoney.com.br/minhas-financas/cartao-de-credito-corresponde-a-mais-de-87-dos-pagamentos-on-line/>. Citado na página 21.

NIU, X.; WANG, L.; YANG, X. A Comparison Study of Credit Card Fraud Detection: Supervised versus Unsupervised. 2019. Disponível em: <http://arxiv.org/abs/1904.10604>. Citado nas páginas 21, 23, 39, 40, 42 e 63.

NOVAKOVIĆ, J. D.; VELJOVIĆ, A.; ILIĆ, S. S.; PAPIĆ, Ž.; MILICA, T. Evaluation of Classification Models in Machine Learning. **Theory and Applications of Mathematics & Computer Science**, v. 7, n. 1, p. Pages: 39 – 46, 2017. ISSN 2067-2764. Disponível em: <https://uav.ro/applications/se/journal/index.php/TAMCS/article/view/158>. Citado nas páginas 49 e 51.

OLIVEIRA, P. H. M. A. **Detecção de fraudes em cartões: um classificador baseado em regras de associação e regressão logística**. 117 p. Tese (Doutorado), 2016. Citado nas páginas 21, 30, 36, 39, 40 e 52.

POWERS, D. M. W.; PROCESSING, N. L. What the F- - measure doesn ' t measure .... n. April, 2015. Citado na página 51.

POZZOLO, A. D.; BORACCHI, G.; CAELEN, O.; ALIPPI, C.; BONTEMPI, G. Credit card fraud detection: A realistic modeling and a novel learning strategy. **IEEE Transactions on Neural Networks and Learning Systems**, IEEE, v. 29, n. 8, p. 3784–3797, 2018. ISSN 21622388. Citado nas páginas 22, 24, 37, 38 e 54.

SADGALI, I.; SAEL, N.; BENABBOU, F. Adaptive Model for Credit Card Fraud Detection. **International Journal of Interactive Mobile Technologies (iJIM)**, v. 14, n. 03, p. 54, 2020. ISSN 1865-7923. Citado nas páginas 22, 37, 39, 40 e 48.

SOEMERS, D. J.; BRYS, T.; DRIESSENS, K.; WINANDS, M. H.; NOWÉ, A. Adapting to concept drift in credit card transaction data streams using contextual bandits and decision trees. **32nd AAAI Conference on Artificial Intelligence, AAAI 2018**, n. 1, p. 7831–7836, 2018. Citado nas páginas 39 e 49.

SUTERIO, V. DETECÇÃO DE CARDIOPATIAS POR ELETROCARDIOGRAMA UTILIZANDO REDES NEURAIS ARTIFICIAIS. 2017. Citado nas páginas 15 e 43.

THUDUMU, S.; BRANCH, P.; JIN, J.; SINGH, J. J. A comprehensive survey of anomaly detection techniques for high dimensional big data. **Journal of Big Data**, Springer International Publishing, v. 7, n. 1, 2020. ISSN 21961115. Disponível em: <https://doi.org/10.1186/s40537-020-00320-x>. Citado na página 38.

TICKNER, P.; BUTTON, M. Deconstructing the origins of Cressey's Fraud Triangle. **Journal of Financial Crime**, v. 28, n. 3, p. 722–731, 2020. ISSN 17587239. Citado na página 30.

WHITROW, C.; HAND, D. J.; JUSZCZAK, P.; WESTON, D.; ADAMS, N. M. Transaction aggregation as a strategy for credit card fraud detection. **Data Mining and Knowledge Discovery**, v. 18, n. 1, p. 30–55, 2009. ISSN 13845810. Citado nas páginas 39, 40, 54 e 55.

Yufeng Kou; Chang-Tien Lu; Sirwongwattana, S.; Yo-Ping Huang. Survey of fraud detection techniques. In: **IEEE International Conference on Networking, Sensing and Control, 2004**. [S.l.: s.n.], 2004. v. 2, p. 749–754 Vol.2. Citado na página 22.

.

.

GLOSSÁRIO

Ataque de fraude: Conjunto de tentativas de fraudes que possuem mesmas características ou *modus operandi*..

Bandeiras: As bandeiras são instituições que criam as plataformas de pagamento. Elas são responsáveis por administrar as políticas das operações, manter a rede de comunicação global e tornar a plataforma atraente, aumentando o número de pagamentos com cartões. Sua fonte de receita é composta pelas taxas cobradas dos estabelecimentos e a aplicação de multas devido ao não cumprimento das políticas. Alguns exemplos bem conhecidos são a Visa e a MasterCard..

Big Data: Refere-se ao volume de informações que uma empresa coleta e armazena de diversas fontes, como dados de clientes, transações comerciais, correspondência por e-mail, bem como presença nas mídias sociais.

Cartão de Crédito: Cartão de crédito é uma forma de pagamento eletrônico. É um cartão de plástico que pode conter ou não um chip e apresenta na frente o nome do portador, número do cartão e data de validade (pelo menos) e, no verso, um campo para assinatura do cliente, o número de segurança (CVV2) e a tarja magnética (geralmente preta).

Chargeback: O chargeback, que em tradução livre significa estorno, ocorre quando o titular do cartão não concorda com uma transação do extrato do cartão de crédito e reclama junto ao emissor. Se o emissor entender que o titular do cartão não efetuou a compra ou se o produto não foi recebido, o valor da transação é reembolsado..

Compra legítima: Compra realizada por um consumidor que pretende realizar o pagamento..

Comércio Eletrônico: Também conhecido como e-commerce, é a compra ou venda de qualquer informação, serviço ou produto através de redes de computadores.

Concept drift: Na análise preditiva e no aprendizado de máquina, *oncept drift* significa que as relações estatísticas entre as *features* e a variável que o modelo está tentando prever mudam ao longo do tempo de maneiras imprevistas..

CVV: Código de três dígitos utilizado como validador para compras de cartão não presente, normalmente encontrado no verso do cartão de crédito..

Desacordo comercial: Acontece quando existe divergência entre a expectativa do comprador e a entrega do vendedor sobre o produto ou serviço negociado..

Detecção de fraude: Ações que procuram detectar fraudes uma vez que ela foi realizada..

Emissores: São os responsáveis por conceder crédito, emitir o cartão e manter relacionamento com o titular do cartão de crédito. Suas principais fontes de receita com o mercado de cartões são juros relacionados ao financiamento, anuidades e outros serviços agregados ao funcionamento do cartão. Normalmente, são instituições financeiras que atuam como emissores, onde os principais exemplos são os bancos..

Ensemble: Modelo de aprendizado de máquina que utiliza-se de vários modelos preditivos com objetivo de atingir um desempenho superior ao dos modelos individuais..

Escore ou Score de fraude: Pontuação que visa mensurar o risco de uma transação ser fraudulenta. Normalmente é a saída de um modelo de aprendizado de máquina..

Especialista de fraude: Pessoa com grande experiência em fraudes e com conhecimento do domínio que faz revisão manual de transações com objetivo de identificar fraudes..

Estatística KS: A estatística KS é derivada do teste de hipótese não paramétrico de igualdade de distribuições contínuas, conhecido como teste Kolmogorov-Smirnov. Ela é obtida a partir da maior distância entre as distribuições de score acumuladas das transações fraudulentas e não fraudulentas..

Feature engineering: Processo de construção ou derivação de novas *features* para modelos de aprendizado de máquina..

Features: Conjunto de variáveis, informações ou características utilizadas como entrada de um modelo de aprendizado de máquina..

Fluxo de compra: Conjunto de etapas necessárias para o consumidor realizar a compra..

Fraudador: Pessoa que comete fraude. No comércio eletrônico se refere à qualquer pessoa que age sozinha ou em grupo que realiza uma compra utilizando seus dados ou de terceiro sem intenção de realizar o pagamento..

Gateways: Funcionam como as maquininhas do mundo físico mas no ambiente online. São empresas que conectam os estabelecimentos às adquirentes, integrando a infraestrutura computacional de ambas as partes. O gateway é responsável pela captura dos dados e sua transmissão para as adquirentes..

Marketplace: Plataformas online onde pessoas físicas e jurídicas podem fazer anúncios, achar consumidores interessados e negociar produtos e serviços..

Meio de pagamento: Método utilizado para pagamento da compra, sendo os mais comuns: cartão de crédito, boleto bancário, pix e cateiras eletrônicas..

Prevenção de fraude: Ações que visam evitar que os fraudadores realizem fraudes..

Reason Code: *Reason code* de *chargeback* é um código alfanumérico de 2 a 4 dígitos fornecido pelo banco emissor envolvido em um *chargeback*, com o objetivo de identificar o motivo da disputa..

Segundo fator de autenticação: Camada extra de segurança que tem como objetivo confirmar a identidade do usuário..

Sistema Antifraude: Conjunto de ferramentas e processos que objetivam a identificação e prevenção de fraudes..

Sistema de regras: Conjunto de regras lógicas desenvolvidas por analistas de fraudes da forma: *Se* <condição> *Então* <ação>. Elas podem ser voltadas à política da loja, por exemplo, no caso de não liberar a compra de bebidas alcoólicas para pessoas com idade menor que 18 anos, ou para gerenciamento de risco, inclusive, utilizando-se de Modelos Estatísticos..

Subadquirentes: Funcionam como facilitadores, os subadquirentes mantém contato com diversas adquirentes e oferecem acesso à elas para seus clientes..

Tipo de fraude: Classificação das ocorrências de fraude de acordo com características ou *modus operandi*..

Titular: Pessoa física ou jurídica que possui o cartão de crédito registrado em seu nome..

Transação de cartão não presente: Transação com pagamento remoto, quando o cartão ou o titular não está presente fisicamente, com coleta das informações via inserção manual..

Transação de cartão presente: Transação com pagamento em ponto de venda presencial com leitura do chip magnético, normalmente com utilização de senhas..

Índice de Chargeback: Indicador calculado a partir da proporção do valor monetário de compras aprovadas que se mostraram fraudulentas, matematicamente, é a divisão do valor de fraude sofrido pelo valor total aprovado..


ICMC
USP
SÃO CARLOS