INSTITUTO FEDERAL DO ESPÍRITO SANTO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE CONTROLE E AUTOMAÇÃO

GABRIEL LORENZONI BENZ

SISTEMA DE APOIO À DETECÇÃO DE FRAUDES EM *E-COMMERCE*

SERRA
2017

GABRIEL LORENZONI BENZ

SISTEMA DE APOIO À DETECÇÃO DE FRAUDES EM *E-COMMERCE*

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Controle e Automação do Instituto Federal do Espírito Santo, como parte dos requisitos necessários para a obtenção do título de Mestre em Engenharia de Controle de Automação.

Orientador: Prof. Dr. Leandro Colombi Resendo
Coorientador: Prof. Dr. Marcelo Eduardo Vieira Segatto

SERRA
2017

Dados Internacionais de Catalogação na Publicação (CIP)

B479s
2017

Benz, Gabriel Lorenzoni
Sistema de apoio à detecção de fraudes em *E-Commerce* / Gabriel Lorenzoni Benz. - 2017.
100 f.; il.; 30 cm

Orientador: Prof. Dr. Leandro Colombi Resendo.
Coorientador: Prof. Dr. Marcelo Eduardo Vieira Segatto.
Dissertação (mestrado) - Instituto Federal do Espírito Santo, Programa de Pós-graduação em Engenharia de Controle de Automação, 2017.

1. Comércio eletrônico. 2. Sistemas especialistas (Computação). 3. Internet. I. Resendo, Leandro Colombi. II. Segatto, Marcelo Eduardo Vieira. III. Instituto Federal do Espírito Santo. IV. Título.

CDD 384.33

Bibliotecária Rogeria Gomes Belchior - CRB6/ES 417

MINISTÉRIO DA EDUCAÇÃO

INSTITUTO FEDERAL DO ESPÍRITO SANTO

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE CONTROLE E AUTOMAÇÃO

GABRIEL LORENZONI BENZ

SISTEMA DE APOIO À DETECÇÃO DE FRAUDES EM E-COMMERCE

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Controle e Automação do Instituto Federal do Espírito Santo, como requisito parcial para obtenção de título de Mestre em Engenharia de Controle e Automação.

Aprovado em 06 de fevereiro de 2018

COMISSÃO EXAMINADORA

Prof. Dr. Leandro Colombi Resendo
Instituto Federal do Espírito Santo
Orientador

Prof. Dr. Daniel Cruz Cavalieri
Instituto Federal do Espírito Santo
Membro Interno

Prof. Dr. Marcelo Eduardo Vieira Segatto
Universidade Federal do Espírito Santo
Membro externo

Prof. Dr. Flávio Miguel Varejão
Universidade Federal do Espírito Santo
Membro externo

RESUMO

O comércio eletrônico torna-se cada vez mais popular à medida que a sociedade se move em direção a uma economia digital. Contudo, essa popularidade tem sido acompanhada por um aumento de práticas fraudulentas. Para evitar fraudes, as empresas de *e-commerce* investem uma quantidade significativa de dinheiro e tempo em soluções de terceiros que visam reduzir perdas. Comumente, por serem serviços de pagamento, essas soluções têm acesso apenas a um subconjunto reduzido de informações e tentam identificar transações fraudulentas analisando, por exemplo, o perfil do cliente e os dados de pagamento dos pedidos. Neste processo de análise convencional, as informações específicas de *e-commerce*, como o comportamento de navegação do cliente, seus dados cadastrais, seu histórico de compras, dentre outras são ignoradas. O objetivo deste trabalho é, portanto, propor um sistema de detecção de fraudes que utiliza tanto informações de pagamento, quanto informações de domínio do *e-commerce* para classificar transações como legítimas ou fraudulentas. Nesse trabalho, destacamos duas contribuições: em primeiro lugar, mostra a importância do conhecimento do domínio na identificação de fraudes; em segundo lugar, o sistema proposto apresentado neste trabalho pretende facilitar o projeto sistemático de novos sistemas anti-fraude. No trabalho de detecção de fraudes, foram utilizados dados reais de um *e-commerce* de bebidas, modelados em três classificadores: baseado em regras (algoritmo RIPPER), árvore de decisão (algoritmo C4.5) e *naive bayes*. Adicionalmente, os resultados dos classificadores são comparados e o sistema de detecção de fraudes é proposto, apresentando desempenho excelente em relação à área sob a curva ROC.

Palavras-chave: Detecção de fraude. Comércio Eletrônico. Sistemas especialistas. Árvore de Decisão.

ABSTRACT

Electronic commerce (e-commerce) has become more popular as society moves deeper towards a digital economy. Although, this popularity has been accompanied by an increase in fraudulent practices. In order to prevent frauds, e-commerce merchants invest significant amount of money and time in 3rd party solutions that aim to reduce losses. Commonly, as payment services, these solutions have access only to a reduced subset of information and try to identify fraudulent transactions by analyzing, for example, customer profile and payment method data. In this conventional analysis process, e-commerce specific information, such as customer search and browser behavior, promotion usage, etc. are ignored. The objective of this work is, therefore, to propose a fraud detection system at the e-commerce side that uses both payment information and e-commerce domain information to classify transactions as legitimate or fraudulent. The contributions of this research are twofold. Firstly, it shows the importance of domain knowledge at identifying frauds. Secondly, the proposed system intends to facilitate the systematic design of new anti-fraud systems. In order to detect fraudulent transactions, real data was used from *e-commerce* beverage store, modeled on three classifiers: rule-based (RIPPER algorithm), decision tree (C4.5 algorithm) and *naive* bayes. In addition, the results of the classifiers are compared and the fraud detection system is proposed, presenting excellent performance in relation to the area under the ROC curve.

Keywords: Fraud detection. E-commerce. Expert systems. Decision Tree.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS

ACP - Análise de Componentes Principais

CRISP-DM - *Cross Industry Standard Process for Data Mining*

JDK - *Java Developer Kit*

PCA - *Principal component analysis*

RIPPER - Repeated Incremental Pruning to Produce Error Reduction

REP - *Reduced Error Pruning*

ROC - *Receiver Operating Characteristic*

SPF - Sistemas de Prevenção de Fraudes

SDF - Sistemas de Detecção de Fraudes

SVM - *Support Vector Machines*

SUMÁRIO

1 INTRODUÇÃO

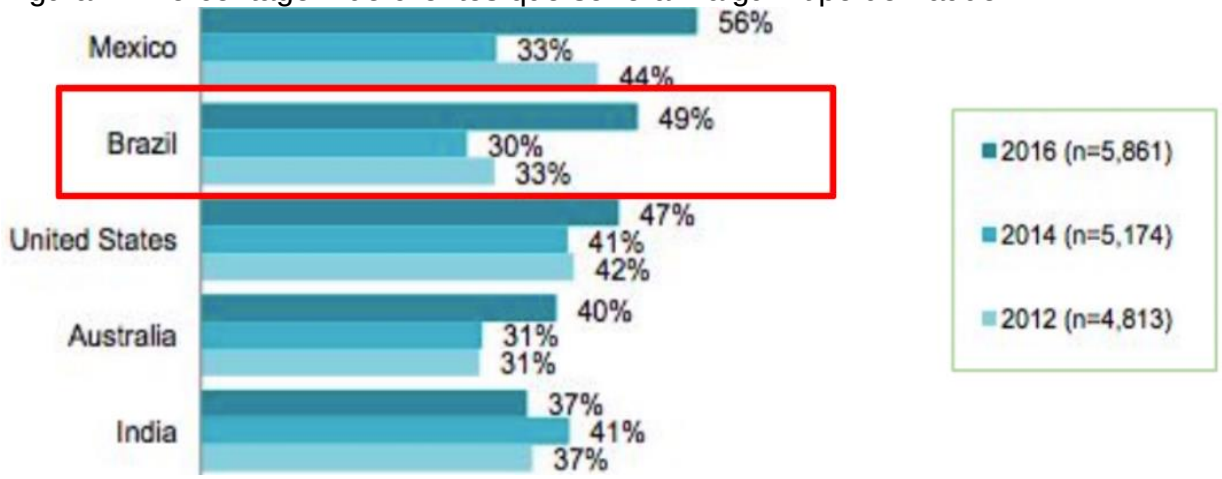
O crescimento significativo do *e-commerce* (comércio eletrônico) está mudando a maneira como as empresas abordam suas vendas e estratégias de marketing. Pesquisas da Cybersource (2016) indicavam uma movimentação de U$66,7 bilhões em e-commerces da América Latina em 2016, sendo o Brasil responsável por 42% dessa movimentação. Os números continuam promissores para 2017: Webshoppers - Ebit (2017) estima que o crescimento nominal seja de 12% e a Abcomm (2017) de 11% em relação ao ano anterior.

No entanto, a popularidade do comércio eletrônico está simultaneamente provocando um crescimento de atividades fraudulentas. Obviamente, tanto usuários legítimos como fraudadores usam o *e-commerce*, mas a natureza anônima das compras online não promove um ambiente ideal para que as transações ocorram (MASSA; VALVERDE, 2014).

Entende-se por fraude qualquer crime de lucro que use o ato de enganar como seu principal modo de operação, utilizando maneiras ilegais de tirar dinheiro da vítima através de força, truque ou roubo (WELLS, 2017). No contexto do comércio eletrônico, esses atos podem estar relacionados com roubos de cartão de crédito, roubos de contas de usuários, não identificação de compras realizadas por parte dos usuários, utilização dos sites de maneira indevida, prática de cartel de preços feitos pelos comerciantes, entre outros.

Em março de 2016, a *ACI Universal Payments*, em conjunto com instituições financeiras, varejistas e processadores de pagamento, conduziram pesquisas de mercado sobre fraude online em 20 países. Essas pesquisas evidenciaram o Brasil como um dos países com mais casos de fraude relatados pelos usuários, conforme mostra a Figura 1.

Figura 1 - Porcentagem de clientes que sofreram algum tipo de fraude.



Fonte: ACI Universal Payments (2016).

Santiago et al. (2015) afirmam que, nas transações eletrônicas realizadas pela *Internet*, não é possível realizar algumas verificações usuais do mundo físico, como por exemplo a verificação da assinatura do cliente, a análise visual do comportamento do suposto comprador, a identificação visual com fotografia, entre outras.

Como resultado, o comércio eletrônico se torna também um ambiente com uma alta incidência de fraudes, gerando prejuízos bilionários. Segundo Cybersource (2016), as perdas por motivos de fraudes no *e-commerce* atingiram uma média de 1,4% da receita total das empresas na América Latina e 1,5% no Brasil.

Coelho et al. (2006) citam que os custos do estabelecimento com fraudes incluem: perda de mercadorias (incluindo frete e embalagem), perdas com taxas bancárias, risco de cancelamento de contratos com as administradoras dos cartões, perda de faturamento pela rejeição de pedidos e custo elevado de uma equipe de análise de risco. Além disso, os autores ainda afirmam que o estabelecimento pode sofrer com a perda de confiança por parte dos clientes e com perda de reputação da loja no mercado.

Obviamente, os usuários de cartão de crédito - tanto os portadores de cartão como os comerciantes que recebem pagamentos via cartão - também acabam pagando por essa perda através de taxas e tarifas maiores e reduções de benefícios (CHAN;

STOLFO, 1998). Dessa forma, é do interesse tanto dos usuários como dos bancos a redução do uso ilícito dos cartões.

Diante desse cenário, fica evidente a necessidade de estratégias para prevenção e detecção de fraudes em transações via *Internet*. Dentre os principais desafios no desenvolvimento de um sistema de detecção de fraude para o comércio eletrônico, Pozzolo et al. (2015) e Santiago (2014) citam: i) a natureza evolutiva das fraudes (*concept drift*[1]); ii) a grande quantidade de dados que devem ser processados em um curto período de tempo; iii) o grande desbalanceamento dos dados, e; iv) a existência de ruídos nos dados.

Alguns comerciantes tentam se prevenir dessas ameaças por meio de uma rigorosa avaliação manual das tentativas de compra, porém, essa prática tende a ser cara e desvia o foco da empresa para atividades de aspectos operacionais ao invés de concentrar os esforços em atividades de geração de receita.

Obviamente, a necessidade de analisar uma grande quantidade de dados em um curto período de tempo, somada à mudança constante no padrão de identificação de fraude, torna a análise estritamente manual inviável, uma vez que alguns tipos de fraude só podem ser identificados através de análises que correlacionam diversas evidências (SANTIAGO et al. 2015).

Fica evidente, portanto, a necessidade de estruturar um sistema que utiliza técnicas computacionais e fornece ferramentas que auxiliam no combate às fraudes de maneira automatizada, atendendo aos requisitos de tempo de resposta a um custo economicamente viável.

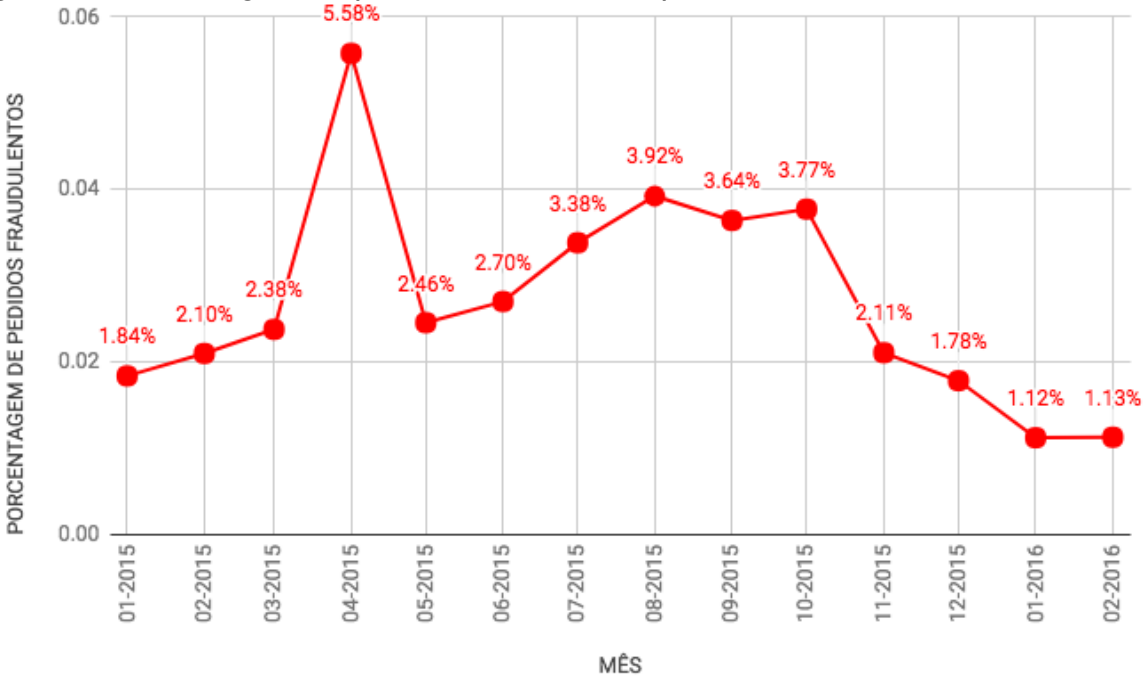
Os impactos negativos gerados pelas fraudes, somados às exigências desse novo tipo de consumidor online, de obter uma compra segura e rápida; bem como a necessidade de concorrer com os mercados globais e a busca por uma constante eficiência operacional, têm feito com que os comércios latino-americanos se adaptem para oferecer soluções que unam gerenciamento e processamento de pagamentos, segurança e prevenção à fraude (CYBERSOURCE, 2016).

[1] Segundo Yanxia et al. (2016), *concept-drift* refere-se aos dados cujas características podem mudar arbitrariamente ao longo do tempo.

Alternativamente, as empresas de *e-commerce* estão adotando soluções terceirizadas de combate à fraude, a fim de reduzir as perdas. Essas soluções podem ser contratadas e integradas às lojas separadamente ou já integradas ao intermediador de pagamentos online utilizado pela empresa.

A empresa de *e-commerce* foco do presente estudo adotou uma solução antifraude terceirizada em outubro de 2015 e o resultado pode ser visto na Figura 2, que apresenta a porcentagem de pedidos fraudulentos realizados por mês, desde janeiro de 2015.

Figura 2 - Porcentagem de pedidos fraudulentos por mês.



Fonte: Próprio autor.

Portanto, fica evidenciado que, de fato, essas soluções terceirizadas tendem a reduzir a quantidade de fraudes. No entanto, por serem serviços de pagamento, esses sistemas são limitados, pois utilizam somente informações relacionadas ao pagamento do pedido para classificar as transações. Além disso, essas soluções tendem a ser genéricas, a fim de atender a maior quantidade possível de clientes. Adicionalmente, por muitas vezes as empresas de *e-commerce* preferem que a

gestão e a inteligência do processo de combate à fraude fiquem sob sua responsabilidade.

A tarefa de análise e identificação de transações fraudulentas, portanto, pode ser vista como um problema computacional de classificação, sendo então aplicáveis técnicas de classificação, aprendizado de máquina e mineração de dados (SANTIAGO et al. 2015).

As informações de pagamento são fundamentais para a análise de fraude, contudo, há uma série de outras informações que também podem ajudar no processo de classificação, como por exemplo, os dados completos do perfil do cliente, o histórico de endereços utilizados, o histórico de produtos que geralmente são comprados pelo cliente, as formas de pagamento comumente utilizadas, a utilização de promoções, o histórico da navegação online do cliente no *e-commerce*, o histórico das buscas realizadas pelo cliente, etc. Essas informações são de domínio do *e-commerce* e, apesar de poderem ajudar no processo de identificação de fraudes, são totalmente ignoradas pelos analisadores de fraude convencionais.

Este trabalho, portanto, propõe uma abordagem abrangente para tratar o problema com fraudes em *e-commerces*. Inicialmente, são abordadas técnicas e ferramentas computacionais para detecção das fraudes. Em seguida, um sistema de apoio à detecção de fraudes é desenvolvido, utilizando informações de domínio do e-commerce em uma base real de dados.

Todas as etapas do desenvolvimento desse sistema, desde o entendimento do negócio, o entendimento dos dados, a modelagem e a avaliação dos classificadores utilizados até a implantação do sistema são detalhadas ao longo dos capítulos seguintes.

1.1 OBJETIVO

O principal objetivo deste trabalho é propor um sistema de apoio à detecção de fraudes em transações online, que aplica técnicas de classificação às informações

de domínio do e-commerce para identificar se uma determinada transação é fraudulenta ou não.

Portanto, diferentemente dos sistemas convencionais, o sistema proposto utiliza informações extras, só existentes no domínio do *e-commerce* que, quando somadas às informações de pagamento (comumente utilizadas pelos sistemas convencionais), podem ajudar a melhorar o desempenho do classificador de fraudes.

Além disso, o presente trabalho possui os seguintes objetivos específicos:

a) Elaborar um estudo das principais técnicas utilizadas na detecção de fraudes, apresentado na seção 2.

b) Aplicar as técnicas mais adequadas para o problema estudado no desenvolvimento do sistema de apoio à detecção de fraudes.

c) Aplicar e comparar diferentes técnicas de classificação a fim de extrair melhores resultados.

d) Validar o sistema desenvolvido em um cenário real. No caso deste trabalho, foram utilizados dados disponibilizados por uma empresa brasileira de *e-commerce* de bebidas.

Para o desenvolvimento do sistema de apoio a detecção de fraudes, foram utilizados três classificadores: baseado em regras (utilizando o algoritmo RIPPER), árvore de decisão (utilizando o algoritmo C4.5) e, por fim, *naïve bayes*.

Ao longo do trabalho, os classificadores foram comparados, suas vantagens e desvantagens foram analisadas.

1.2 CONTRIBUIÇÃO TECNOLÓGICA ESPERADA

A principal contribuição tecnológica deste trabalho é auxiliar a detecção de pedidos fraudulentos por meio do desenvolvimento de um sistema que automatiza a classificação de pedidos de um *e-commerce* quanto a sua legitimidade, de forma

inteligível e seja fácil de atualizar e incorporar às soluções de detecção de fraude já utilizadas.

Para alcançar este objetivo, foi desenvolvido um sistema *web* que encapsula o conhecimento de especialistas de detecção de fraudes em um classificador baseado em regras para classificar os pedidos de um *e-commerce* em relação a sua legitimidade.

1.3 ORGANIZAÇÃO DA DISSERTAÇÃO

O restante deste trabalho é organizado da seguinte forma: no Capítulo 2 é apresentada a revisão bibliográfica, abrangendo os conceitos de fraude, o funcionamento de uma transação online com cartão de crédito, a diferença entre os sistemas de prevenção de fraude e os sistemas de detecção de fraude, os classificadores utilizados no trabalho e as características do problema de fraude. No Capítulo 3 é apresentada a metodologia utilizada, os materiais e métodos que conduziram o estudo. No Capítulo 4, o experimento é detalhado em cada fase da metodologia utilizada. Os resultados são apresentados no Capítulo 5 e, por fim, a conclusão é descrita no Capítulo 6.

2 REVISÃO BIBLIOGRÁFICA

Neste capítulo são apresentados os aspectos teóricos relevantes à elaboração de um sistema de detecção de fraude.

São abordadas a definição de fraude e suas variações, o funcionamento da transação com cartão de crédito em um *e-commerce*, são discutidas as principais técnicas de prevenção e detecção de fraudes encontradas na literatura. Também são abordados os classificadores utilizados no trabalho e as principais métricas para avaliar seus desempenhos.

2.1 DEFINIÇÃO DE FRAUDE

Segundo Wells (2017), a fraude, em seu sentido mais amplo, abrange qualquer crime de lucro que use o ato de enganar como seu principal modo de operação, utilizando maneiras ilegais de tirar dinheiro da vítima através de força, truque ou roubo.

Wells (2017) também cita que todas as fraudes envolvem algum ato de enganar, mas nem todo engano é uma fraude. Segundo o autor, quatro elementos gerais devem estar presentes para que exista fraude: i) uma declaração falsa; ii) conhecimento de que a declaração era falsa quando foi proferida; iii) confiança da vítima na declaração falsa; e iv) danos resultantes da confiança da vítima na declaração falsa.

Quando aplicado ao contexto de cartões, o conceito de fraude ocorre no momento em que o portador do cartão informa ao emissor o não reconhecimento de transações descritas em sua fatura, iniciando o processo de disputa de *chargeback*.

Segundo Oliveira (2016), *chargeback* (ou estorno) é o instrumento criado pelas bandeiras de cartões para efetuar a reversão de transações financeiras que já foram realizadas, mas por quaisquer motivos, não se concretizaram integralmente. Logo,

trata-se de um processo utilizado para reverter uma transação financeira que custeou uma transação comercial mal sucedida ou, em casos específicos, para reverter a própria transação financeira quando ela é mal sucedida, por exemplo, devido a erros sistêmicos.

No contexto *online*, existem vários tipos de fraude relatados na literatura. Alguns autores, como Felipe Junior et *al.* (2012) e Santiago et al. (2015) categorizam os tipos de fraudes em: i) fraudes relacionadas a cartão de crédito; ii) fraudes realizadas por comerciantes; e iii) fraudes na *internet*. As seguintes subseções descrevem os diferentes tipos de fraude relacionadas ao meio de pagamento (cartão de crédito).

2.1.1 Fraude de proposta

Fraudes de proposta acontecem quando o fraudador tenta criar uma pessoa fictícia ou usa dados roubados de alguma pessoa para abrir uma conta de cartão de crédito. Na maioria das vezes, o endereço apresentado diverge do endereço do cliente honesto, que muitas vezes nunca entrou em contato com a instituição (GADI, 2008).

Gadi (2008) ainda cita que modelos de *Fraud Application* e cruzamento de informações com *bureaus*[2] internos e externos, como a Serasa, são altamente eficazes para detecção.

2.1.2 Perda ou roubo

2 Bureau: é um estabelecimento - geralmente um departamento ou divisão de uma repartição pública - onde se realizam trabalhos administrativos e outros negócios relacionados com a atividade da empresa.

Segundo Felipe Junior et *al.* (2012), fraudes de perda ou roubo são exemplificadas em casos em que o cliente perde seu cartão ou tem seu cartão roubado e alguma pessoa não autorizada tenta realizar transações com ele.

Felipe Junior et *al.* (2012) ainda afirmam que, como normalmente o fraudador desconhece o limite de crédito disponível do cliente, o que se observa é uma sequência de transações de valores pequenos e algumas transações negadas de valores maiores que o disponível, até que se descubra alguma transação que caiba no limite de crédito.

2.1.3 Aquisição de conta

Segundo Bhatla et al. (2003), este tipo de fraude ocorre quando um fraudador obtém ilegalmente informações pessoais válidas dos clientes. Dessa forma, o fraudador assume o controle de uma conta legítima fornecendo o número da conta do cliente ou o número do cartão.

Os autores ainda citam que há casos em que o fraudador, em seguida, entra em contato com o emissor do cartão como sendo o titular para pedir que as correspondências sejam redirecionadas para um novo endereço. O fraudador relata a perda do cartão e pede que um substituto seja enviado para o novo endereço.

2.1.4 Cartões falsos

A criação de cartões falsificados juntamente com cartões perdidos/roubados representam a maior ameaça em fraudes de cartão de crédito. Fraudadores estão constantemente encontrando meios novos e mais inovadores para criar cartões falsificados. Algumas das técnicas utilizadas para a criação de cartões falsos são: i) adulteração da faixa magnética do cartão; ii) clonagem (*skimming*); iii) sites e/ou

algoritmos geradores de números de cartão de crédito (FELIPE JUNIOR et al., 2012).

2.2 TRANSAÇÃO COM CARTÃO DE CRÉDITO

A proposta do presente trabalho é sugerir um sistema de fraude abrangente, independente do meio de pagamento utilizado no pedido: boleto, cartão de crédito, cartão de débito, cupom de desconto, etc. Contudo, entender o funcionamento da transação com cartão de crédito, método de pagamento mais popular na empresa de estudo deste trabalho, é importante para a compreensão de conceitos fundamentais relacionados à fraude.

Segundo Santiago (2014), as transações com cartões de crédito são operações complexas e possuem cinco partes envolvidas em todo o processo: o portador, o estabelecimento, o adquirente, a bandeira e o emissor. Todas essas partes interagem para a autorização e execução de uma transação quando um portador executa uma compra em um estabelecimento comercial. Cada uma delas é detalhada a seguir:

a) Portador: é o(a) responsável, dono(a) do cartão. Pode ser uma pessoa ou uma empresa, ou seja, são aqueles dispostos a realizarem pagamentos com cartões e, para isso, se tornam clientes dos emissores.

b) Estabelecimento: são lojistas, prestadores de serviço, autônomos ou estabelecimentos comerciais que desejam receber pagamentos com cartões. No caso de transações *online* (como exemplo, as transações feitas em um *e-commerce*), os estabelecimentos geralmente realizam as operações acessando serviços *web* (*web services*) fornecidos pela adquirente.

c) Adquirente: também conhecido como credenciadoras, são empresas que fazem a comunicação entre o estabelecimento e a bandeira. Em estabelecimentos físicos, o adquirente é responsável pelo aluguel e manutenção dos equipamentos (máquinas de cartão) utilizados. Em

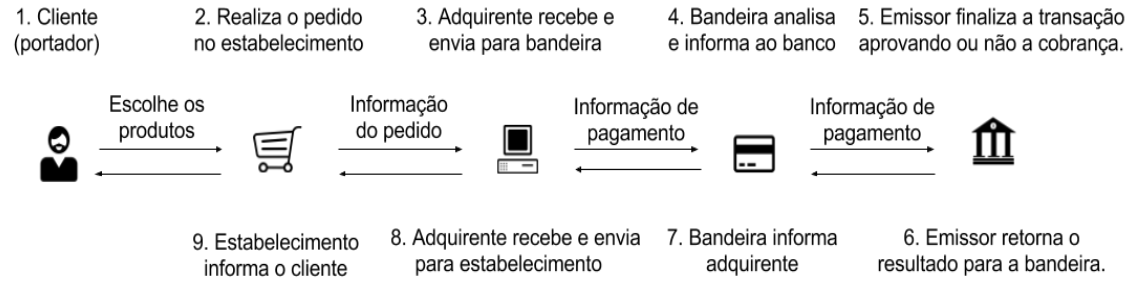
ambientes *online*, o adquirente é responsável pela integração entre o *e-commerce* e a bandeira do cartão. Exemplos de adquirentes no Brasil: Redecard e Cielo.

d) Bandeira: responsável pela comunicação entre o adquirente e o emissor do cartão de crédito. Segundo Oliveira (2016), suas funções são: criar as regras de operação, manter uma rede global de comunicação e, através de políticas e ações de marketing, manter a dinâmica do mercado, ou seja, convencer mais pessoas a realizarem pagamentos com cartões e mais estabelecimentos a aceitarem pagamentos com esse instrumento. As principais fontes de receitas das bandeiras são as taxas cobradas dos emissores e credenciadoras, além de multas aplicadas aos outros membros por descumprimento das regras. Exemplos: Visa, MasterCard, American Express.

e) Emissor: instituições financeiras, em geral bancos, que emitem e administram o cartão de crédito. Também são chamadas de administradoras do cartão. Segundo Oliveira (2016), as maiores fontes de receitas dos emissores são as taxas cobradas das transações, os juros provenientes do financiamento rotativo dos portadores, as anuidades e os serviços agregados como seguros. Apesar dos bancos não serem as únicas instituições a assumirem o papel de emissores, em geral, são eles que desempenham essa função.

Portanto, como mostra a Figura 3, no processo de compra *online*, o cliente (portador) informa os dados do cartão ao *e-commerce* (estabelecimento), que se comunica com um adquirente. O adquirente envia a transação para a bandeira que, por fim, envia para a administradora do cartão (emissor). O emissor, então, decide se a transação será aprovada ou não. Essa informação será retornada por toda cadeia até chegar no estabelecimento, que repassa a informação para o cliente (FELIPE JUNIOR et al., 2012).

Figura 3 - Transação com cartão de crédito em *e-commerce*



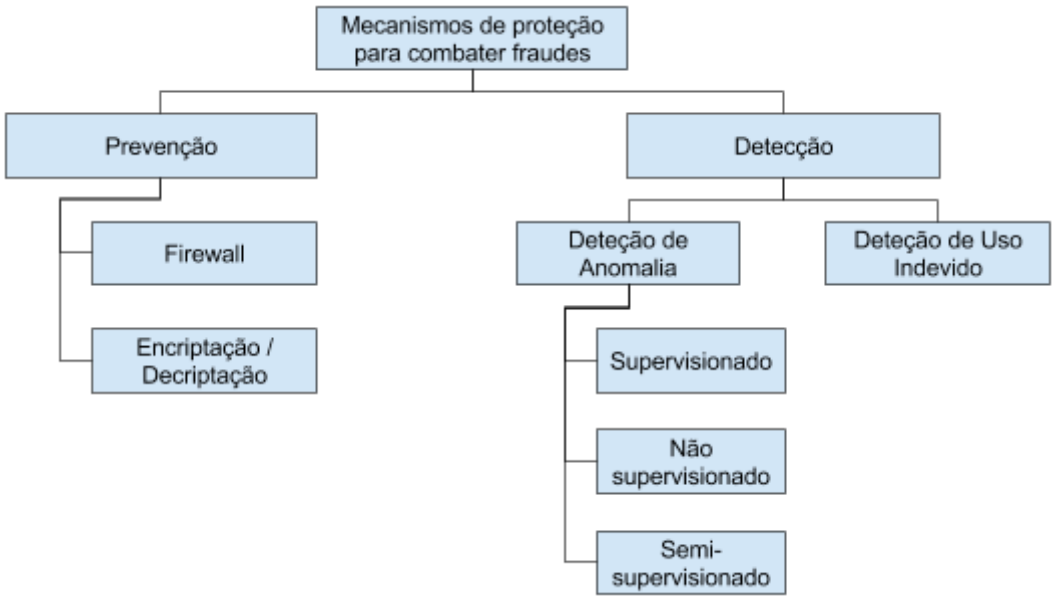
Fonte: Elaborado pelo autor.

2.3 PROTEÇÃO CONTRA FRAUDE

A fraude digital está aumentando consideravelmente com a evolução da vendas em *e-commerce*. Como resultado, a luta contra a fraude tornou-se uma questão importante a ser explorada (MAGALLA, 2013).

A Figura 4 apresenta os principais mecanismos de detecção e prevenção utilizados para combater a fraude, conforme citado por Abdallah et *al.* (2016). As subseções seguintes exploram mais sobre esses mecanismos.

Figura 4 - Principais mecanismos de proteção e detecção de fraudes.



Fonte: Abdallah et al. (2016).

2.3.1 Sistemas de prevenção de fraude

Um sistema de prevenção de fraudes (SPF) é a primeira camada de proteção para assegurar os sistemas tecnológicos contra a fraude. O objetivo desta fase é impedir a ocorrência de fraudes em primeiro lugar.

Nessa fase, esse mecanismo é responsável por restringir, suprimir, desestruturar, destruir, controlar, remover ou impedir a ocorrência de ataques cibernéticos em sistemas computacionais (*hardware* e *software*), redes ou dados (ABDALLAH et *al.* 2016).

Algoritmos de criptografia que são aplicados para decodificar dados são exemplos desses mecanismos. Outro exemplo é o *firewall*, que é uma forma de bloqueio entre a rede interna privada e as redes externas. O *firewall* não só ajuda a proteger os sistemas de acesso não autorizado, mas também permite que uma organização proponha uma política de segurança de rede sobre o fluxo de tráfego entre sua rede e a *Internet* (MAGALLA, 2013).

No entanto, essa camada nem sempre é eficiente e forte (BELO; VIEIRA, 2011). Há situações em que a camada de prevenção pode ser violada por fraudadores. Nesses casos, a proteção deve ser feita na camada de detecção de fraude.

2.3.2 Sistemas de detecção de fraude

Os sistemas de detecção de fraude (SDF) compõem a camada subsequente de proteção, que também é objeto de estudo deste trabalho.

Um sistema de detecção de fraude tenta descobrir e identificar atividades fraudulentas em um ambiente computacional e relatar a um administrador de sistema (BEHDAD et *al.*,2012).

Para melhorar o desempenho da detecção de fraude, esses sistemas geralmente integram uma vasta gama de técnicas de mineração de dados (AKHILOMEN, 2013) (DESAI; DESHMUKH, 2013) (SARAVANAN et *al.*, 2014).

2.4 DETECÇÃO DE USO INDEVIDO

Na abordagem de detecção de uso indevido, os comportamentos fraudulentos são identificados com base em um histórico de transações fraudulentas. Para cada pedido fraudulento, portanto, são identificados comportamentos fraudulentos com base em seus dados. Como exemplo, se um pedido fraudulento for realizado às 3h, tiver 15 produtos, utilizar um determinado cartão de crédito como forma de pagamento, sendo que o cliente tenha sido cadastrado há menos de 1 hora no *site*, esse comportamento é então identificado como fraudulento. De forma análoga, para os pedidos legítimos, são identificados comportamentos considerados normais.

Segundo Hand e Crowder (2012), esse tipo de abordagem utiliza classificadores baseados em regras, estatísticas ou métodos heurísticos para revelar a ocorrência de transações suspeitas específicas.

A detecção de uso indevido é, portanto, um sistema especialista, considerado como um mecanismo de detecção simples e rápido, mas possui uma grande limitação porque não é possível detectar todos os diferentes tipos de fraudes, porque essa abordagem procura apenas padrões previamente conhecidos (WEI et al, 2012).

2.5 DETECÇÃO DE FRAUDE BASEADO EM ANOMALIAS

A abordagem de detecção de anomalias (ou *outliers*) baseia-se em análises de perfil comportamental. Nesse tipo de abordagem, o padrão comportamental de cada indivíduo é modelado, e qualquer desvio do padrão é monitorado (JYOTHSNA et *al.*, 2011).

Os sistemas baseados em anomalias têm o potencial de detectar novas fraudes. Portanto, é o tipo mais comum utilizado na literatura (SUN et *al.*, 2006). Segundo Akhilomen (2013), esses sistemas podem ser categorizados em três tipos: sistemas supervisionados, semi-supervisionados e não supervisionados.

2.5.1 Aprendizado supervisionado

Segundo Abdallah et al. (2016), as técnicas de aprendizagem supervisionada exigem um conjunto de dados que tenha sido previamente rotulado como "fraude" e "não fraude" e envolve o treinamento de um classificador. Esta é a abordagem de aprendizagem mais comum.

Em outras palavras, métodos supervisionados examinam transações previamente classificadas a fim de determinar futuras transações fraudulentas, ou seja, os modelos que utilizam esse tipo de aprendizado possuem seus dados já classificados. Dessa forma, métodos supervisionados precisam de dados históricos para fazer as classificações.

Os métodos de aprendizado abrangem muitos algoritmos, incluindo:

a) algoritmos de classificação: como redes neurais artificiais, k-vizinhos mais próximos, árvores de decisão, regressão logística, redes bayesianas e máquinas de vetor de suporte (SVM, do inglês, *Support Vector Machines*); e

b) algoritmos de regressão: como regressão linear, regressão simples e regressão logística.

Para o problema de detecção de fraudes em cartão de crédito, os métodos mais comuns apresentados na literatura são: sistemas baseados em regras, árvores de decisão, redes neurais, SVM, redes bayesianas e regressão logística, sendo os dois primeiros utilizados no presente trabalho.

No entanto, há críticas relacionadas à aprendizagem supervisionada na literatura. Abdallah et al. (2016) e Phua et al. (2010) citam as seguintes desvantagens:

a) dificuldade de coletar supervisão ou classes: principalmente quando há um enorme volume de dados de entrada, é proibitivamente caro, se não impossível, classificar todos eles;

b) além de difícil de coletar, a classificação dos dados de treinamento pode estar incorreta, às vezes é extremamente difícil encontrar classes distintas, há incertezas e ambiguidades na definição de classes;

c) na maioria das vezes vezes, pessoas precisam analisar manualmente cada transação. Isso demanda tempo, além de expor dados muitas vezes sensíveis dos usuários. Essas limitações podem obstruir as implementações das abordagens de aprendizagem supervisionada em alguns casos.

2.5.2 Aprendizado semi-supervisionado

A aprendizagem semi-supervisionada situa-se entre a aprendizagem supervisionada e não supervisionada: são algoritmos capazes de aprender a partir de exemplos rotulados e não rotulados e são bastente úteis quando apenas um pequeno número de exemplos rotulados encontra-se disponível.

O aprendizado semi-supervisionado pode ser utilizado tanto em tarefas de classificação, em que os exemplos rotulados são utilizados no processo de classificação, quanto em tarefas de *clustering*, sendo os exemplos rotulados responsáveis por auxiliar o processo de formação de *clusters*.

Segundo Santiago (2014), o método de aprendizado semi-supervisionado mais popular utiliza técnicas de *clustering* e detecção de *outliers*. O autor ainda cita que primeiro aplicam-se técnicas de *clustering* sobre os dados pré-existentes para modelar os agrupamentos naturais (ou *clusters*), e a cada transação nova aplicam-se técnicas de detecção de *outliers* sobre essa transação em relação aos *clusters* predefinidos.

Um *outlier*, no contexto de detecção de fraudes, é uma transação com um alto grau de discrepância em relação a todo o conjunto de transações (por exemplo, um pedido com mais de 50 produtos, em um conjunto em que todos os pedidos possuem em média 2 produtos).

Portanto, a vantagem da aprendizagem semi-supervisionada em relação à aprendizagem supervisionada é que é possível chegar a desempenhos melhores utilizando amostras rotuladas ou não, com poucas amostras já rotuladas (ABDALLAH et al., 2016).

2.5.3 Aprendizado não supervisionado

As técnicas de aprendizagem não supervisionadas, também conhecidas como aprendizado por observação e descoberta, detectam fraudes em um conjunto de dados de teste não rotulados sob o pressuposto de que a maioria das instâncias no conjunto de dados é não fraude (ABDALLAH et al. 2016).

Ao contrário da técnica supervisionada, "sem supervisão" significa que não há uma classificação prévia (rótulo de classe) para a construção do modelo. Dois algoritmos clássicos simples empregados na aprendizagem não supervisionada são: i) algoritmos de *clustering*, como técnicas do tipo *K-means* e ii) algoritmos de redução dimensional como PCA (*Principal Component Analysis*, ou ACP, Análise de Componentes Principais).

Bolton e Hand (2001) citam que o principal benefício do uso de dessa abordagem é que ela não se baseia em identificação precisa para os dados do rótulo, que muitas vezes são escassos ou inexistentes.

2.6 CLASSIFICADORES

No contexto de aprendizagem de máquina, há uma série de problemas estudados que são denominados como problemas de classificação.

Nesse tipo de problema, o objetivo é organizar os objetos em categorias, ou classes, previamente definidas. Assim, tem-se (i) um conjunto de classes, (ii) um conjunto de instâncias e (iii) um classificador. A partir disso, a classificação é a atividade, realizada pelo classificador, de atribuir uma classe a cada uma das instâncias (OLIVEIRA, 2016).

Oliveira (2016) ainda afirma que, para executar a tarefa de classificação, é preciso construir o classificador (ou método de classificação) que será responsável por fazer as atribuições entre os elementos do conjunto de amostras e os elementos do conjunto das classes.

Para se construir o classificador, é necessário dispor de um conjunto de treinamento, ou seja, um conjunto de amostras cujas classes sejam previamente conhecidas. Após seu treinamento, o classificador é exposto a uma série de amostras cujas classes são desconhecidas para ele prever as classes dessas amostras.

No contexto de transações de *e-commerce*, o conjunto de classes é formado por "transação legítima" e "transação fraudulenta" e o conjunto de instâncias, ou amostras, é formado pelas informações a respeito da transação em si, que neste trabalho envolvem não só informações de pagamento, mas também as informações de domínio do *e-commerce*.

O conjunto de treinamento é formado por instâncias que foram previamente classificadas como fraudulentas ou legítimas e, por fim, o classificador deve atribuir uma e somente uma classe para cada nova transação.

Na literatura, são encontrados diversos classificadores relacionados ao problema de fraude, sendo os encontrados mais frequentemente apresentados na Tabela1.

Tabela 1 - Relação dos classificadores utilizados nos trabalhos de referência.

Classificador	Referências
Árvores de Decisão	Ramos (2014); Felipe Junior et al. (2012); Hilas e Sahalos (2007); Queiroga (2005); Lemos et al. (2005); Chan e Stolfo (1999).
Sistema baseado em regras	Oliveira (2016); Santiago (2014); Felipe Junior et al. (2012); Chan e Stolfo (1999).
Naive Bayes	Souza (2014); Felipe Junior et al. (2012).
Algoritmo Genético	Duman e Ozcelik (2011).
K-Vizinhos mais próximos (k-NN)	Queiroga (2005).
Redes Neurais	Akhilomen (2013); Caldeira et al. (2012); Queiroga (2005); Lemos et al. (2005).
Máquinas de Vetores de suporte (SVM)	Santiago (2014); Hejazi e Singh (2012); Qibei e Chunhua (2011).

Fonte: Elaborado pelo autor.

2.6.1 Sistemas baseados em regras

Segundo Oliveira (2016), sistemas baseados em regras são algoritmos de aprendizagem supervisionada cujo classificador é constituído por um conjunto de regras, cada uma com o formato:

Se *determinada condição é verdadeira*

Então *faça determinada ação.*

Conforme citado em Felipe Junior et al. (2012), formalmente, as regras para o modelo são representadas na form $R = (r_1 \wedge r_2 \wedge r_3 \ldots \wedge r_k)$a, em que R é conhecido como o conjunto de regras e os ri's são as regras de classificação. Cada regra pode ser expressa da seguinte maneira:

$$r_i: (Condição) \rightarrow y_i \quad (1)$$

O lado que contém a condição é chamado de antecedente da regra, o lado que contém a ação é chamado consequente da regra.

$$Condição_i = (A_1\ op\ v_1) \wedge (A_2\ op\ v_2) \ldots \wedge (A_k\ op\ v_k) \quad (2)$$

Em que:

(A_1, v_1): é um par atributo-valor.

op: é um operador lógico escolhido no conjunto: $\{=, \neq, <, >, \leq, \geq\}$.

Cada teste de atributo $\{A_1\ op\ v_1\}$ é conhecido como um conjunto. O lado direito da regra é chamado de consequência da regra, que contém a classe y_i prevista. Conceitualmente, uma regra r cobre um registro x se os atributos de x satisfizerem as condições expressas no antecedente de r.

Felipe Junior et. al. (2012) ainda afirmam que a qualidade de uma regra de classificação pode ser estabelecida utilizando-se duas medidas: cobertura e precisão. Dado um conjunto de dados D e uma regra de classificação $r: A \rightarrow y$, a cobertura é definida como a fração de registros em D que satisfazem a regra r. A precisão, por sua vez, é uma medida de fator de confiança e é definida como a fração de registros disparados por r cujos rótulos de classe sejam iguais a y. Portanto, tem-se:

$$Cobertura(r) = \frac{|A|}{|D|} \quad (3)$$

$$Precisão(r) = \frac{|A \cap y|}{|A|} \quad (4)$$

Em que:

$|A|$: é o número de registros que satisfazem ao antecedente da regra.

$|A \cap y|$: é o número de registros que satisfazem tanto o antecedente quanto o consequente da regra.

$|D|$: quantidade total de registros.

Segundo Tan et al. (2014), um conjunto de regras pode ter duas propriedades:

a) Conjunto de regras mutuamente excludentes: quando não houver nenhum registro que é coberto por mais de uma regra. Em outras palavras, cada instância dispara uma, e apenas uma, regra.

b) Conjunto de regras completo: em que cada um dos registros submetidos ao classificador dispara ao menos uma regra. Um ponto importante é que muitos classificadores baseados em regras não possuem tais propriedades. Neste caso, se um conjunto de regras não for completo uma regra padrão deve ser adicionada e disparada quando todas as outras falham. A classe atribuída por essa regra padrão é conhecida como classe padrão que geralmente é a classe majoritária.

Caso as regras do classificador não forem mutuamente excludentes, Felipe Junior et al. (2012) citam que há duas formas de resolver o problema:

a) Regras Ordenadas: nesta forma, as regras são ordenadas de forma decrescente de prioridade, que pode ser determinada de diversas maneiras, como por exemplo, utilizando as medidas de precisão e cobertura, ou levar em consideração a ordem que foi gerada em relação às outras regras.

b) Regras não Ordenadas: nesta forma, o classificador permite que um registro dispare múltiplas regras de classificação e considera cada uma como um voto - que pode ser ponderado ou não - para uma determinada classe. O registro recebe a classe que tiver o maior número de votos.

A construção do classificador baseado em regras passa a ser, portanto, a determinação de quais regras são participantes do conjunto de regras. Oliveira (2016) cita que a extração das regras que subsidiam o classificador pode ser dividida em dois grupos de métodos:

a) Métodos Diretos: extraem as regras diretamente dos dados; e

b) Métodos Indiretos: primeiramente utilizam outros métodos de classificação para, posteriormente, converter os resultados desses métodos em regras. Classificadores frequentemente utilizados nos métodos indiretos são as árvores de decisão e os algoritmos genéticos.

2.6.1.1 O algoritmo RIPPER

No presente trabalho, foi utilizado o algoritmo RIPPER (*Repeated Incremental Pruning to Produce Error Reduction*) como implementação do classificador baseado em regras.

Segundo Oliveira (2016), um algoritmo de cobertura sequencial começa com uma lista de regras vazia e, depois, é utilizada uma função para extrair a regra que cubra o máximo de amostras de positivas. Ao verificar que a regra formulada atingiu seu objetivo (de acordo com um critério de parada), a função retorna o controle para o modulo principal. Neste ponto, os registros cobertos pela regra são removidos do conjunto de treinamento e a regra é adicionada à lista de regras. A função de extração de regras é novamente chamada e o processo todo se estende até que todas as amostras do conjunto de treinamento tenham sido analisadas.

Oliveira (2016) ainda cita que o algoritmo RIPPER é a evolução de outro algoritmo para indução de regras chamado REP (*Reduced Error Pruning*). Além de utilizar a técnica de cobertura sequencial para gerar as regras, o RIPPER utiliza um conjunto de dados de validação para executar a poda das regras e, quando há duas classes no conjunto de treinamento, o algoritmo define a classe com maior número de amostras como a classe padrão (no caso do presente trabalho, as transações legítimas) e descobre regras para detectar a classe que tenha um número menor de amostras (no caso, as transações fraudulentas).

Dessa forma, o RIPPER é adequado para conjuntos de treinamento desbalanceados ou com ruídos. Além disso, a complexidade computacional do algoritmo é linear ao número de amostras de treinamento.

O RIPPER é, portanto, um algoritmo que utiliza regras ordenadas, que repete uma poda incremental para a redução de erro. Implementar a técnica REP é uma vantagem do algoritmo que se mostra útil para evitar o ajuste excessivo (*overfitting*).

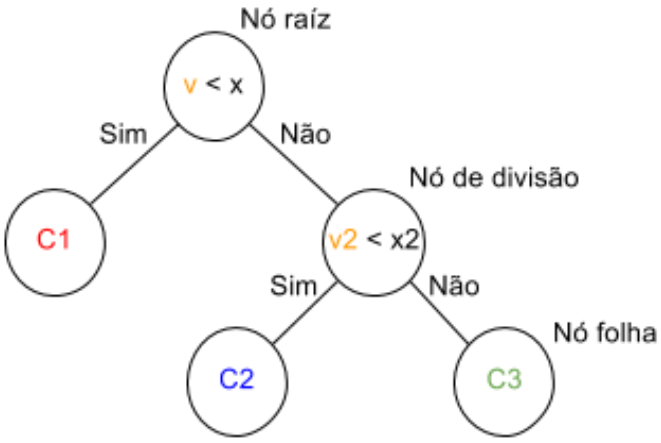
O algoritmo é composto por duas fases: uma de construção das regras (que envolve uma fase de crescimento das regras através da medida de cobertura) e outra de otimização (poda) das regras. Tais fases são destacadas no algoritmo apresentado no Apêndice A.

2.6.2 Árvores de Decisão

Um classificador baseado em árvore de decisão é aquele que expressa uma aproximação para os valores resultantes de uma função-alvo cuja imagem é formada por valores discretos e, para representar tal aproximação, utiliza-se uma estrutura de árvore (MITCHELL, 1997).

Segundo Ramos (2014), o classificador árvore de decisão usa a estratégia "dividir para conquistar" na solução de um problema de decisão, ou seja, um problema complexo é dividido em problemas mais simples, aos quais é aplicada a mesma estratégia, de forma recursiva. As soluções dos subproblemas são combinadas na forma de uma árvore para gerar uma solução do problema complexo, como mostra a Figura 5.

Figura 5 - Exemplo de estrutura de Árvore de Decisão.



Fonte: Elaborado pelo autor.

Segundo Facelli et al. (2011), conceitualmente, uma árvore de decisão é um grafo acíclico direcionado em que cada nó:

a) um nó de divisão: possui dois ou mais sucessor e contém um teste condicional baseado nos valores dos atributos; ou

b) um nó folha: não possui sucessores e contém os valores dos rótulos.

Lemos et al. (2005) citam as seguintes vantagens da utilização de Árvores de Decisão:

a) não assumem nenhuma distribuição particular para os dados;

b) as características ou os atributos podem ser categóricos (qualitativos) ou numéricos (quantitativos);

c) pode-se construir modelos para qualquer função desde que o número de exemplos de treinamento seja suficiente;

d) elevado grau de compreensão.

2.6.2.1 Indução de Árvores de Decisão

O processo de construção de uma árvore a partir de uma amostra de treinamento é conhecido como indução da árvore. O algoritmo abaixo foi adaptado de Ramos (2014) e descreve como funciona o processo de indução, começando com uma árvore vazia.

a) Se todos os objetos de treinamento no nó corrente n pertencem à categoria c, crie um nó folha com a classe c. Caso contrário, avalie cada um dos possíveis testes condicionais t pertencentes ao conjunto T dos possíveis testes, usando uma função heurística.

b) Escolha o melhor teste condicional t como teste do nó corrente n.

c) Crie um nó sucessor para cada resultado distinto do teste t e particione os dados de treinamento entre os nós sucessores usando o teste t.

d) Um nó n é considerado puro se todos os objetos de treinamento em n pertencem à mesma classe. Sendo assim, repita os passos anteriores em todos os nós impuros.

Ramos (2014) ainda cita que as condições de parada do crescimento da árvore frequentemente citadas na literatura são: (i) quando todas as instâncias de treinamento pertencem a uma mesma classe; (ii) quando a árvore chega à altura máxima; (iii) caso aconteça a divisão do nó, o número de objetos em um ou mais nós sucessores é menor que um determinado limite inferior.

2.6.2.2 Seleção de atributos

O ponto principal da classificação utilizando árvores de decisão é a seleção dos atributos que serão considerados, ou seja, no processo de construção da árvore, faz-se necessária a escolha correta do atributo preditivo, que definirá o sucesso do algoritmo de indução.

A ideia é identificar o atributo que trará a pureza mais rapidamente, ou seja, que particionará a base de dados de forma que cada ramo tenha apenas registros

relacionados a uma classe. Dessa forma, pequenas variações nos dados utilizados pelo treinamento podem causar seleções de atributos diferentes em cada ponto de escolha da árvore. O efeito pode ser significativo, pois as escolhas de atributos afetam todas as sub-árvores descendentes (PEDROSO et al. 2013).

Segundo Felipe Junior et al (2012), no processo de indução de uma árvore de decisão, a cada iteração, um algoritmo de aprendizagem deve selecionar um atributo (condição de teste do atributo) que melhor divida os registros e, para isso, são utilizadas algumas métricas, que são calculadas em termos da distribuição da classe dos registros antes e depois da divisão.

Os autores ainda citam que as métricas utilizadas para selecionar a melhor divisão são muitas vezes baseadas no grau de impureza dos nós filhos. Dessa forma, quanto menor o grau de impresa, mais distorcida é a distribuição das classes, isto é, um nó com uma distribuição de classe (0;1) posui impureza zero. Por outro lado, um nó com distribuição (0,5;0,5) possui maior impureza.

As métricas de imprureza podem ser classificadas de acordo com sua entropia ou utilizando a medida de impureza gini, descritas a seguir:

$$Entropia(t) = -\sum_{i-0}^{c-1} p\left(i/t\right) \log_2 p\left(i/t\right) \quad (5)$$

$$Gini(t) = 1 - \sum_{i-0}^{c-1} \left[p\left(i/t\right)\right]^2 \quad (6)$$

Em que:

$p\left(i/t\right)$: é a fração de registros que pertencem à classe i de um nó t

Portanto, a entropia possui uma escala que varia entra 0 e 1, enquanto que a medida de impureza Gini varia antre 0 e 0,5. Ambas chegam ao seu valor máximo quando não se consegue diferenciar uma classe de outra e chegam ao seu valor mínimo quando todos os registros pertencem à mesma classe. Logo, a cada iteração, o algoritmo irá selecionar o atributo que possua o menor valor obtido pela métrica.

2.6.2.3 O algoritmo C4.5

O algoritmo C4.5, referenciado no Apêndice B, foi desenvolvido por J. R. Quinlan, na década de 90, é é considerado até os dias atuais um algoritmo de referência para o desenvolvimento e análise de novos algoritmos de classificação.

Segundo Carvalho (2014), assim como o ID3 (algoritmo também criado por Quinlan em na década de 80), o C4.5 realiza a construção da árvore por meio de um algoritmo guloso, utilizando a estratégia de “dividir e conquistar” - o algoritmo, portanto, indica a melhor escolha de forma local, com a expecativa que estas escolhas levem à melhor escolha global, também chamada de ótimo global.

No caso de árvores de decisão, o algoritmo C4.5 trabalha com a melhor escolha local, utilizando o conceito de entropia de informação, a cada nó da árvore sem reconsiderar suas escolhas prévias, trabalhando dessa forma, de modo recursivo a cada nó até aque a árvore esteja montada.

Em cada nó da árvore, portanto, C4.5 escolhe o atributo dos dados que mais efetivamente divide seu conjunto de amostras em subconjuntos rotulados em uma classe ou outra. O critério de divisão é o ganho de informação normalizado. Logo, o atributo com maior ganho de informação normalizado é escolhido para tomar a decisão. Esse passo é repetido recursivamente em sublistas menores.

2.6.3 *Naive Bayes*

O *Naive Bayes* é um método de aprendizagem de máquina supervisionado, baseado no Teorema de *Bayes*, que é considerado “ingênuo” porque assume que os atributos são independentes (ZAREAPOOR et al., 2012).

Dessa forma, ao supor que X seja um conjunto de atributos e Y seja a variável de classe, se esta tiver um relacionamento não determinístico com os atributos, isto é, se os atributos de X forem independentes, então, podem-se tratar Xe Y como

variáveis aleatórias e capturar seu relacionamento utilizando probabilisticamente $P\,(Y|X)$. Esta probabilidade condicional também é conhecida como probabilidade posterior de Y, em oposição à sua probabilidade anterior, $P(Y)$.

Sendo assim, conforme citado por Felipe Junior et al. (2012), durante a fase de treinamento, descobrem-se as probabilidades posteriores $P\,(Y|X)$ para cada combinação de X e Y baseada em informações coletadas a partir dos dados de treinamento. A partir dessas probabilidades, portanto, um registro de teste X' pode ser classificado encontrando-se a classe Y'que maximize a probabilidade posterior, $P\,(Y'|X')$.

Logo, assumindo que a classe $y_j \in Y$ esteja relacionada com as n variáveis que representam as instâncias, tem-se:

$$P(Y|X_1\,,X_2\,,X_3\,,\ldots,X_n\,) = \frac{P(Y)\;P(Y|X_1\,,X_2\,,X_3\,,\ldots,X_n\,)}{P(X_1\,,X_2\,,X_3\,,\ldots,X_n\,)} \tag{7}$$

A equação pode ser simplificada, considerando-se a independência entre os atributos (característica do *Naïve Bayes*), e assim podendo-se utilizar a definição de probabilidade condicional. Além disso, o denominador pode ser ignorado, já que não depende de Y. O resultado da simplificação é dado por:

$$P(Y|X_1\,,X_2\,,X_3\,,\ldots,X_n\,) = P(Y)\prod_{i=1}^{n} P(X_i|Y) \tag{8}$$

Por fim, a classificação é feita atribuindo-se a clase $y_j \in Y$ com maior probabilidade à instância de entrada.

2.7 CARACTERÍSTICAS DO PROBLEMA DA DETECÇÃO DE FRAUDE

Nesta seção são descritas as principais características do problema de detecção de fraudes sob o ponto de vista das técnicas computacionais.

2.7.1 *Concept drift* - Variação das fraudes

Uma característica inerente ao problema de detecção de fraudes é a variação constante na relação entre transações legítimas e fraudulentas.

Os fraudadores tendem a evoluir suas técnicas e as fraudes tendem a mudar com o tempo, adaptando-se às reações do sistema. Tão logo os fraudadores percebem que certo tipo de comportamento fraudulento pode ser detectado, eles irão adaptar suas estratégias e tentar outras (BOLTON; HAND, 2001) (PHUA et al., 2010). Com isso, os sistemas de detecção de fraude precisam ser adaptativos e reavaliados constantemente (DELAMAIRE et al., 2009). Logo, visto que os fraudadores aprendem o funcionamento dos sistemas de detecção de fraude e os padrões de fraude evoluem ao longo do tempo, alguns classificadores serão mais relevantes que outros em um dado momento (CHAN; STOLFO, 1998) (CHAN et al., 1999).

Porém, é importante ressaltar que fraudadores novos surgem a todo o momento e estes não terão conhecimento dos métodos de detecção de fraude que tiveram sucesso no passado. Isso significa que os padrões de detecção de fraudes previamente aplicados precisam ser utilizados constantemente em conjunto com suas respectivas evoluções (BOLTON; HAND, 2001).

No contexto do *e-commerce*, outra característica presente é a variação do padrão de consumo, que tende a ser sazonal: durante épocas do ano, há picos de consumo que alteram a frequência e o *ticket* médio[3] dos pedidos, dificultando a identificação de eventuais anomalias.

Portanto, é necessário o uso de algoritmos de aprendizado adaptativo para lidar com a questão da evolução da fraude. Os algoritmos de aprendizagem adaptativa podem ser vistos como algoritmos avançados de aprendizagem incremental que são capazes de atualizar o modelo de detecção para transmissão de dados em evolução ao longo do tempo (GAMA et al., 2013).

[3] Ticket médio é o termo utilizado para identificar o valor médio gasto por cada cliente. Ele é calculado a partir da soma total de venda em um período de tempo e dividido pelo número de pedidos realizados no mesmo período (RODRIGUES; LOPES, 2015).

2.7.2 Classificação desbalanceada

Uma característica do problema de detecção de fraudes em transações eletrônicas é o grande desbalanceamento dos dados, uma vez que a quantidade de transações fraudulentas é muito menor que a de transações legítimas (BHATTACHARYYA et al., 2011).

Alguns autores da literatura, como Abdallah et al. (2016), consideram este o mais crítico dos problemas da detecção de fraudes. De acordo com pesquisas feitas pela Cybersource (2016), as fraudes representam cerca de 1% das transações (1,4% na América Latina, 0,8% na Europa). Dessa forma, um classificador poderia obter uma assertividade grande se simplesmente classificasse todas as transações como legítimas (CHAN et al., 1999).

2.7.3 Ruído

Segundo Bhattacharyya et al. (2011), o ruído pode surgir, por exemplo, através de transações fraudulentas que não foram detectadas, e portanto erroneamente tratadas como legítimas durante o treinamento do classificador; ou falsos possitivos, isto é, quando uma compra legítima é classificada como fraude.

Segundo Weiss (2004), o ruído nos dados afetará qualquer sistema de mineração de dados, mas traz um prejuízo maior para aqueles que precisam lidar com dados desbalanceados, impactando principalmente as classes mais raras.

2.7.4 Quantidade de dados a serem analisados

No contexto de cartões de crédito, Chan et al. (1999) apontam que milhões de transações são processadas todos os dias. Analisar quantidades tão enormes de

transações requer técnicas altamente competentes e bem dimensionadas, além de exigir um poder de computação considerável.

Este problema também é citado por Hilas e Sahalos (2007), que mencionam que a grande escala e as dimensões elevadas do conjunto de dados de fraude somada à presença de vários recursos, atributos, entradas e variáveis tornam extremamente difícil e complicado o processo de mineração de dados e detecção de fraude.

Além disso, no contexto do *e-commerce*, essa grande quantidade de informação precisa ser processada em um curto período de tempo, de modo que o cliente tenha a resposta da transação - seja ela positiva ou não - preferencialmente ainda em seu tempo de navegação na loja *online*.

2.7.5 Indisponibilidade de um conjunto de dados real

O desenvolvimento acadêmico na área de detecção de fraudes têm sido seriamente impactado de forma negativa pela severa limitação no compartilhamento de dados.

Muitos autores, dentre eles, Santiago (2015), Jha et al. (2012), Qibei e Chunhua, (2011) e Ngai et al. (2011) citam que uma característica associada à detecção de fraudes é a indisponibilidade do conjunto de dados reais em que os pesquisadores podem ter acesso para trabalhar em suas pesquisas. Dentre os motivos, é citado que bancos e instituições financeiras, que geralmente são o foco de pesquisa e responsáveis pelas informações, não estão prontos para revelar dados sensíveis da transação do cliente devido a razões de privacidade.

Além disso, muitas vezes os resultados das pesquisas desenvolvidos na indústria ou em parceria com esta não são abertos ao público (BHATTACHARYYA et al., 2011).

2.8 ANÁLISE SENSÍVEL AO CUSTO

Conforme citado em Oliveira (2016), Gadi et al. (2010) e Chan e Stolfo (1998), a distribuição das classes no conjunto de treinamento pode afetar o desempenho do classificador. Os autores ainda sugerem que, para o caso de classes desbalanceadas, os modelos de classificação terão melhores resultados ao se alterar a proporção de pedidos fraudulentos e pedidos legítimos na amostragem.

Portanto, conforme estudos apresentados principalmente em Chan e Stolfo (1998), também citado em Oliveira (2016), conclui-se que, dada a característica de classes desbalanceadas do problema de fraude, não é recomendável manter os dados amostrados com mesma proporção de classes da base de dados. Dessa forma, a amostragem deve ter uma distribuição de pedidos fraudulentos e pedidos legítimos diferente da proporção da base de dados, que no caso do presente trabalho, é de 1,68%.

Para fazer a estratificação da base, portanto, ao invés de fazer uma amostragem aleatória simples, optou-se por seguir o processo apresentado em Oliveira (2016), que apresenta uma análise sensível ao custo, proposto inicialmente por Elkan (2001).

Para isso, nesta fase do projeto, foram definidas as seguintes métricas com os especialistas de domínio:

a) o custo aproximado do falso-positivo (C_{FP}) é de aproximadamente 7% do valor da transação;

b) o custo aproximado do falso-negativo (C_{FN}) é de aproximadamente 1,3% do valor da transação;

c) analogamente ao proposto em Oliveira (2016), os custos aproximados de verdadeiro-positivo (C_{VP}) e verdadeiro-negativo (C_{VN}) são considerados como zero.

A Tabela 2 apresenta matriz de benefícios utilizada no trabalho.

Tabela 2 - Matriz de benefícios.

	Custo pedido de Fraude	**Custo pedido Legítimo**
Aprovada	-1,3 * (valorTransacao)	0
Negada	0	-0,07 * (valorTransacao)

Fonte: Elaborado pelo autor.

Assim sendo, utiliza-se o teorema apresentado em Elkan (2001) para se chegar à proporção adequada de amostras negativas, que no presente trabalho, são representadas pelas transações legítimas.

$$\frac{C_{FP}}{C_{FN}} = \frac{-0{,}07 * (valorTransacao)}{-1{,}3\ (valorTransacao)} = 5{,}38\% \quad (9)$$

Com isso, considera-se um limiar a priori de classificação para transações fraudulentas igual a 5,38%.

2.9 MÉTRICAS PARA AVALIAÇÃO DE CLASSIFICADORES

Muitas medidas de avaliação de desempenho foram definidas para sistemas de classificação. Dentre elas temos taxa de cobertura, taxa de acertos e área sob a curva ROC (DUMAN; OZCELIK, 2011).

Para os sistemas de classificação de fraude, em cartão de crédito em específico, uma métrica bastante utilizada é a taxa mensal de *chargeback* (BHATLA et al., 2003). Porém, para o problema em questão, a melhor solução não necessariamente é aquela que detecta muitas fraudes, mas sim aquela que detecta fraudes possivelmente em quantidade menor, mas com um prejuízo em potencial maior (DUMAN; OZCELIK, 2011).

Assim como citado frequentemente na literatura (ZAREAPPOR; SHAMSOLMOALI, 2015; SANTIAGO et al. 2014; SEEJA; ZAREAPPOR, 2014; POZZOLO et al. 2014),

o desempenho do classificador proposto neste trabalho é avaliado em termos de 5 métricas de classificação relevantes para a detecção de fraudes, descritos nas subseções seguintes.

Com exceção da área sob a curva ROC, as métricas são determinadas através da matriz de confusão, exibida na tabela 3, em que “positivo” corresponde aos casos de fraude e “negativo” aos casos de não-fraude:

Tabela 3 - Matriz de Confusão.

	Previsão Positiva	**Previsão Negativa**
Caso Positivo	Verdadeiro Positivo (VP)	Falso Negativo (FN)
Caso Negativo	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Fonte: Elaborado pelo autor.

Em que:

a) VP: quantidade de pedidos fraudulentos classificados corretamente como fraudulentos;

b) FN: quantidade de pedidos legítimos classificados incorretamente como fraudulentos;

c) FP: quantidade de pedidos fraudulentos classificados incorretamente como legítimos;

d) VN: quantidade de pedidos legítimos classificados corretamente como legítimos.

2.9.1 Sensibilidade (cobertura ou taxa de detecção de fraude)

Representa a taxa de pedidos fraudulentos que são acertadamente classificados como fraudes, ou seja, é a razão entre as transações corretamente classificadas como fraudulentas e todas as transações fraudulentas.

$$S = \frac{VP}{VP+FN} \quad (10)$$

2.9.2 Precisão

Representa a medida do quão é exata a classificação para amostras positivas (pedidos fraudulentos), ou seja, é a razão entre as transações corretamente classificadas como fraudulentas e todas as transaçòes classificadas como fraudulentas.

$$P = \frac{VP}{VP+FP} \quad (11)$$

2.9.3 Taxa de alarme falso

É a taxa de pedidos verdadeiramente legítimos que são classificados como fraudes. Pode ser entendido como:

$$TAF = 1 - E \quad (12)$$

Em que:

> E: representa especificidade, ou seja, é a razão entre as transações corretamente classificadas como legítimas e todas as transações realmente legítimas.

Dessa forma, tem-se:

$$E = \frac{VN}{VP+FN} \quad (13)$$

Logo:

$$TAF = 1 - \frac{VN}{VP+FN} \quad (14)$$

2.9.4 Medida F

Representa a media harmonica entre a precisão e a sensibilidade (ou cobertura), representada por:

$$Medida\ F = \frac{2*P*S}{P+S} \quad (15)$$

Em que:

P: representa a precisão;

S: representa a sensibilidade (ou cobertura).

2.9.5 Área sob a curva ROC

Segundo Oliveira (2016), o termo ROC, acrônimo de *Receiver Operating Characteristic*, é usado para designar a relação entre taxa de acerto e a taxa de falsos alarmes em um canal com ruídos. Trata-se, portanto, de um gráfico no qual o eixo das ordenadas é dado pelo índice de verdadeiros positivos, enquanto o eixo das abcissas é dado pelo índice de falsos positivos.

Logo, a área sob essa curva (AUC, do inglês *area under curve*, ou AROC) é tida como uma medida de qualidade do classificador, pois quanto maior a área, melhor o desempenho do classificador. Hosmer et al. (2013) definem o poder de classificação de um modelo como mostra a Tabela 4.

Tabela 4 - Poder de classificação de um modelo dado pela AROC.

Valor AROC	Poder de Classificação
$AROC = 0{,}5$	Inexistente
$0{,}7 \leq AROC \leq 0{,}8$	Aceitável
$0{,}8 \leq AROC \leq 0{,}9$	Muito bom
$AROC \geq 0{,}9$	Excelente

Fonte: Elaborado pelo autor.

É válido ressaltar que métricas como desempenho (acurácia) e taxa de erro, comuns em problemas análise de classificadores, foram desconsideradas no presente trabalho por serem consideradas métricas tendenciosas no caso de dados desbalanceados. Como as fraudes representam cerca de 1% dos pedidos, um sistema que simplesmente classifica pedidos como legítimos em todas as ocasiões teria um desempenho de 99% e, não por isso, é um bom classificador.

2.10 EFICIÊNCIA ECONÔMICA

É importante ressaltar que, para o problema de detecção de fraudes, a melhor solução não é necessariamente aquela que detecta muitas fraudes, mas sim aquela que detecta fraudes com prejuízo em potencial maior (DUMAN; OZCELIK, 2011).

Cada transação em um *e-commerce* tem um valor financeiro específico e, por isso, possui um prejuízo em potencial próprio. Dessa forma, o custo da classificação errada de uma fraude varia para cada transação (CHAN et al., 1999).

Para minimizar o custo total da fraude, Bhatla et al. (2003) sugerem um balanço ótimo entre a redução do número de fraudes e o custo do processo de análise manual. É importante lembrar que, uma eventual classificação errada de uma transação implica, em, além dos gastos financeiros, outros gastos mais difíceis de calcular, relacionados à insatisfação do cliente devido a eventuais demoras de análise, podendo gerar impressões de mau serviço prestado.

No presente trabalho, foi feita uma simplificação do custo do processo de detecção de fraudes para estimar a eficiência econômica do classificador utilizado. Essa simplificação abstrai o custo do processo de análise manual por transação, bem como o custo da impressão de um mau serviço prestado no caso de transações legítimas serem identificadas.

Trata-se, portanto, de uma adaptação do conceito de eficiência econômica proposta por Caldeira et al. (2012), também utilizada por Santiago (2015). O cálculo é feito baseado em 4 situações possíveis, conforme detalhado na tabela 5.

Tabela 5 - Cálculo da eficiência financeira

Situação	**Contexto**	**Impacto**	**Cálculo**
1	Transação legítima e o classificador identifica como legítima	Ganho	Lucro * Valor pedido
2	Transação legítima e o classificador identifica como fraude	Perda	Lucro * Valor pedido
3	Transação fraudulenta e o classificador identifica como fraude	Nulo	
4	Transação fraudulenta e o classificador identifica como legítima	Perda	(1-Lucro) * Valor pedido

Fonte: Elaborado pelo autor.

3 METODOLOGIA

O objetivo do sistema de apoio à detecção de fraudes em *e-commerce* é identificar transações fraudulentas dentre milhares de transações legítimas em um curto período para análise, minimizando esforço operacional e perdas financeiras.

É válido ressaltar que toda ação posterior à identificação da fraude, como estorno de transações, repatriação de valores ou eventual reporte aos órgãos de segurança pública não pertencem ao escopo deste trabalho.

3.1 MATERIAIS

Para a elaboração do sistema, foi utilizada uma base de dados transacionais composta por 3.191.889 transações realizadas em um *e-commerce* entre janeiro de 2015 e fevereiro de 2016.

Cada transação desta base de dados possui uma classificação (rótulo) binária que a identifica como fraudulenta ou genuína. Portanto, todo pedido aprovado pela instituição financeira que posteriormente foi contestado, isto é, houve *chargeback*, foi rotulado como fraude. Os demais pedidos aprovados pela instituição financeira foram classificados como genuínos.

É importante ressaltar que a maioria dos sistemas de detecção de fraude de transações em *e-commerce* considera apenas as informações relativas ao pagamento para identificar se um pedido é fraudulento ou não.

No presente trabalho, além das informações de pagamento, foram coletadas várias de informações de domínio do e-commerce que se mostraram importantes para a identificação das fraudes, como informações de perfil do cliente, tentativas de acesso ao *e-commerce*, histórico de pedidos realizados, histórico de produtos comprados, endereços cadastrados, entre outros.

O MER (Modelo Entidade Relacionamento) do sistema, descrevendo os objetos envolvidos no domínio de negócios, com suas características e como elas se relacionam entre si, é apresentado no Apêndice C.

3.1.1 Equipamentos e ferramentas

Devido ao volume de dados utilizado no experimento, a utilização de equipamentos e ferramentas robustos se mostrou fundamental para o sucesso do trabalho. Por isso, descrevê-los é importante para o entendimento de como o trabalho foi realizado.

Os classificadores utilizados no trabalho foram executados utilizando o programa *Weka (The Waikato Environment for Knowledge Analysis)*, que é uma ferramenta desenvolvida na linguagem de programação Java e consiste em uma coleção de algoritmos de aprendizado de máquina para tarefas de mineração de dados.

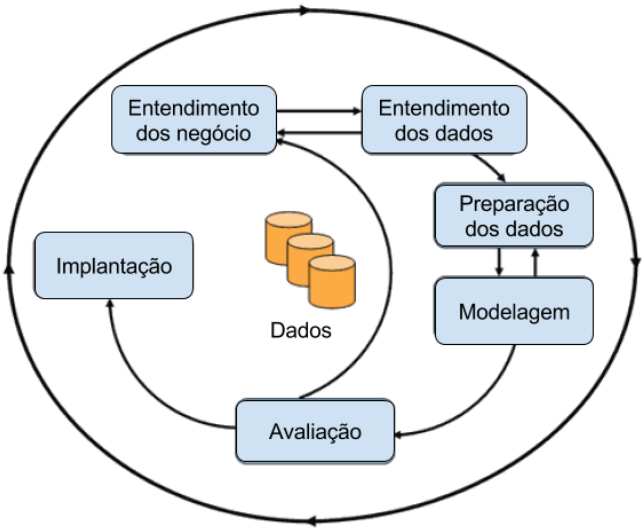
Os experimentos foram executados em um computador, com processador 2 GHz *Intel Core* i5, 16 GB de Memória RAM 1867 MHz LPDDR3. Por fim, foi desenvolvido um sistema de apoio à detecção de fraudes representado como uma aplicação *web*, ou seja, projetado para ser utilizado através de um navegador. Esse sistema foi desenvolvido na linguagem de programação Java (JDK 8), utilizando principalmente o *framework Spring MVC* versão 3.1.4. Para repositório de dados, foi utilizado o *MySQL Community Server*, versão 5.7.20.

3.2 MÉTODO

Para a construção do sistema de apoio à detecção de fraudes, foi utilizada a metodologia de mineração de dados CRISP-DM *(Cross Industry Standard Process for Data Mining)*. Essa metodologia foi utilizada como guia deste trabalho, pois se mostrou mais adaptável às necessidades do projeto.

Dessa forma, a realização do experimento do projeto será detalhada e discutida seguindo as seis etapas do CRISP-DM, conforme exibido na Figura 6. Cada etapa é detalhada nas subseções seguintes.

Figura 6 - Fases do modelo CRISP-DM



Fonte: Adaptado de Ramos (2014).

3.2.1 Entendimento do negócio

Segundo Ramos (2014), o objetivo desta fase é compreender os objetivos do projeto em uma perspectiva de negócio, entendendo o negócio da companhia, os recursos disponíveis, os problemas que precisam ser tratados e as metas do projeto.

O autor ainda cita as seguintes tarefas dessa etapa:

a) determinar objetivos de negócio, alinhado com as necessidades da organização;

b) avaliar o contexto atual com o levantamento dos recursos disponíveis, eventuais restrições e outros fatores que podem impactar o projeto, e;

c) determinar metas de mineração, traduzindo regras e metas de negócio para uma linguagem técnica, clara e não ambígua.

3.2.2 Entendimento dos dados

Segundo Souza (2014), essa etapa tem como objetivo entender os dados disponíveis para resolução do problema. Ramos (2014) cita que a fase de entendimento dos dados é uma etapa crítica, pois pode evitar problemas inesperados durante a fase de preparação dos dados.

Esta fase é composta de:

a) coleta de dados disponíveis;

b) exploração dos dados com a identificação de atributos relevantes para uma tarefa de classificação;

c) verificação da qualidade dos dados, com a busca de inconsistências, erros de codificação e valores faltantes.

3.2.3 Preparação dos dados

Segundo Ramos (2014), esta fase compreende todas as atividades que têm como objetivo a construção do conjunto final de dados. Segundo o autor, esta é uma das etapas mais importantes do CRISP-DM e a que consome mais tempo.

Souza (2014) cita as seguintes atividades dessa etapa:

a) limpeza e tratamento dos dados;

b) seleção dos dados relevantes a serem utilizados no projeto;

c) formatação dos dados para utilização em algoritmos de aprendizado de máquina.

3.2.4 Modelagem

Nesta etapa são selecionados e aprimorados alguns modelos para atingimento dos objetivos do projeto ou, caso necessário, é feito o retorno às etapas anteriores para adequação do projeto a um determinado classificador ou algoritmo de aprendizado de máquina (RAMOS, 2014).

Nesta etapa, portanto, é feita a aplicação de técnicas de modelagem para atingir os objetivos especificados na fase de entendimento do negócio.

3.2.5 Avaliação

Uma vez determinado o modelo (ou os modelos), é preciso analisar se ele atingiu um desempenho satisfatório para alcance dos objetivos de negócio. Esta análise é feita na etapa de avaliação (RAMOS, 2014). Nesta etapa também é feita uma revisão do processo para determinar se alguma tarefa deve ser revista, assim, o projeto poderá progredir para a etapa de implantação ou retornar para fases anteriores.

3.2.6 Implantação

Segundo Souza (2014), esta é a fase final do processo, em que é feita a organização do conhecimento gerado de forma que seja possível utilizá-lo dentro da organização. O autor ainda afirma que a implantação pode ser desde uma solução

simples, como um relatório, ou soluções mais robustas, como um conjunto de dados integrados a sistemas ou a outro processo de *data mining*.

3.3 CLASSIFICADORES UTILIZADOS

Os classificadores e algoritmos discutidos ao longo das seções anteriores apresentam um considerável arcabouço de técnicas possíveis para serem utilizadas em tarefas de classificação.

No presente trabalho, a escolha do classificador foi feita com base em diversos fatores, entre eles, seu desempenho técnico. É importante ressaltar que, na prática, o problema de detecção de fraudes apresenta um requisito importante: como a classificação errada implica em perdas financeiras, é fundamental que o classificador escolhido, além de ter um desempenho altamente satisfatório, também permita que a classificação realizada possa ser interpretada com clareza.

Portanto, para a construção do sistema, foi escolhida a técnica de classificação baseada em regras (utilizando o algoritmo RIPPER, conforme discutido no capítulo 2), devido às seguintes vantagens:

a) Eficiência e flexibilidade: sua complexidade de tempo é linear em função do número de objetos e atributos. Além disso, trata-se de um método não paramétrico, ou seja, não assume nenhuma distribuição para os dados, por isso, não sofre tanto impacto com dados desbalanceados;

b) Possuem boa interpretabilidade: classificadores baseados em regras emulam a estratégia de tomada de decisão de especialistas humanos, logo, apresentam uma alta capacidade de explanação do processo de inferência. Todas as decisões são baseadas nos valores dos atributos usados para descrever o problema, e, portanto, são mais fáceis de interpretar que os pesos numéricos das conexões entre os nós de uma rede neural, por exemplo;

c) Robustez: em relação a valores atípicos e ruídos, são invariantes a transformações monotônicas dos atributos. Dessa forma, utilizar xi ou $\log xi$ como i-ésimo atributo de entrada produz regras com a mesma estrutura;

d) Seleção de atributos: o algoritmo de classificação baseado em regras possui embutido o processo de seleção de atributos relevantes. Esta seleção produz modelos que tendem a ser bastante robustos contra a adição de atributos irrelevantes e redundantes.

Por outro lado, a utilização desse classificador implica em algumas desvantagens, como:

a) Alta dependência de especialistas de domínio: o conhecimento inicial do modelo geralmente é obtido com o conhecimento de especialistas. A aquisição desse conhecimento tende a ser custosa e, por muitas vezes, o algoritmo atua em um domínio restrito, podendo apresentar dificuldades em sua generalização;

b) Capacidade de aprendizado: em geral, essa estratégia não apresenta capacidade de modificar automaticamente sua base de conhecimento, cabendo ao projetista realizar a manutenção e a revisão do sistema;

c) Dificuldade para tratar dados ausentes, que apesar de ser uma desvantagem, não impactou no presente trabalho, pois não foram utilizados dados com essa característica.

3.3.1 Comparação com outros classificadores

Adicionalmente à classificação baseada em regras, para comparação dos resultados, foram utilizados os classificadores Árvore de Decisão e *Naive Bayes*.

O algoritmo utilizado para o classificador Árvore de Decisão foi o C4.5, por conseguir lidar também com atributos contínuos e ser considerado, em muitos casos segundo Phua et at. (2010), melhor que os algoritmos CART e ID3.

Os objetivos dessas comparações são diversos:

a) A utilização da estratégia de árvore de decisão possui vantagens e desvantagens semelhantes, porém, as regras não são modulares como no classificador baseado em regras. Ou seja, no classificador de árvore de decisão, os atributos possuem uma relação de dependência, diferentemente do classificador baseado em regras, em que cada regra pode ser interpretada isoladamente;

b) Devido à capacidade de seleção de atributos, a utilização de árvores de decisão pode ajudar a elucidar novas descobertas e também validar a importância dos atributos utilizados no modelo, que foram sugeridos pelos especialistas;

c) Por fim, a comparação com o classificador *Naive Bayes*, tem como objetivo comparar as estratégias adotadas com um modelo mais tradicional de classificação. O *Naive Bayes*, diferentemente das abordagens de árvore de decisão e baseado em regras, considera que a relação entre a classe e o atributo é não determinística. Por considerar os atributos independentes, é um algoritmo que possui um baixo custo de processamento. Apesar de ser um classificador largamente utilizado e de simples implementação, não possui uma interpretabilidade tão boa quanto os classificadores baseados em regras e árvores de decisão, por isso, uma eventual superioridade no desempenho dessa estratégia pode levantar discussões a respeito de interpretatividade em detrimento ao desempenho.

4 EXPERIMENTO

O estudo de caso apresentado neste trabalho utiliza dados reais do *e-commerce* Wine.com.br, que, de acordo com seu site, é o maior e-commerce de bebidas da América Latina.

Para combater o problema da fraude, a empresa possui uma equipe de analistas especialistas em detectar pedidos fraudulentos. Além disso, em outubro de 2015, a empresa também implantou um sistema anti-fraudes, integrado à adquirente, para reduzir ainda mais os casos de *chargeback*.

Neste capítulo, portanto, são detalhadas todas as tarefas executadas para a geração do sistema de apoio à detecção de fraudes em *e-commerce*, seguindo o modelo de referência CRISP-DM.

4.1 ENTENDIMENTO DO NEGÓCIO

A fase de compreensão e entendimento do negócio consistiu em entender as características do problema de fraude em *e-commerce* e a importância de combatê-la, bem como conhecer a fundo o funcionamento de uma transação online de cartões e as técnicas de prevenção e detecção de fraude.

Nesta fase foi observada ainda a grande dependência que os estabelecimentos têm em relação aos especialistas que fazem análises de fraude de forma manual. Além desse processo ser altamente custoso, é extremamente ineficiente dada a quantidade de informações que é preciso analisar para fazer uma análise de fraude assertiva em um curto período de tempo. Ficou evidente nesta fase, portanto, a necessidade de aplicar técnicas computacionais para a detecção de fraudes.

4.1.1 Aquisição do conhecimento

Os analistas responsáveis por avaliar se um pedido é fraudulento ou genuíno possuem um enorme conhecimento de detecção de fraudes. Um procedimento comum de um analista de fraude envolve os seguintes passos:

a) análise manual dos dados do cliente e seu histórico de compras;

b) análise manual do histórico de navegação do cliente;

c) análise manual do pedido, horário de compra, produtos comprados;

d) análise manual da forma de pagamento utilizada;

e) análise manual do histórico de fraudes detectadas anteriormente.

Cada análise pode ser conduzida com detalhes maiores ou menores dependendo da experiência do analista e da informação disponível sobre cada pedido. Por ser extremamente custoso, esse procedimento é feito em apenas uma amostragem de pedidos, coletados aleatoriamente durante o expediente normal dos analistas.

Para coletar as informações de como é feita a tomada de decisão no processo de análise de fraude, foi adotado um método de engenharia de conhecimento, que incluiu diferentes técnicas para permitir a aquisição de conhecimento dos analistas.

O procedimento de aquisição de conhecimento utilizado no presente trabalho seguiu as seguintes etapas:

a) Entrevistas pessoais: esse tipo de entrevista consistiu em uma série de perguntas a respeito do domínio do problema, envolvendo tópicos específicos e genéricos. O conjunto de perguntas foi preparado antecipadamente, com a finalidade de evitar questionamentos ambíguos ou pouco objetivos;

b) Entrevistas estruturadas: essa técnica consistiu em uma série de entrevistas que foram guiadas por objetivos que permitiam que os analistas sintetizassem seus conhecimentos a respeito de informações específicas. Por fim, esta técnica se mostrou um dos mais eficientes métodos de aquisição de

conhecimento, porém, requer um longo tempo de preparação das perguntas e das dinâmicas;

c) Observação dos analistas: durante o processo de aquisição do conhecimento, os analistas foram observados por 20h, distribuídos em 10 dias úteis;

d) Análise de protocolo: depois de observar o trabalho dos analistas, uma série de possíveis cenários foram estabelecidos e, para um deles, os especialistas expuseram a solução do processo. Nesta fase, alguns exemplos de pedidos foram criados e, para cada exemplo, foi registrada a forma com que cada analista interpreta os dados e infere a classificação do mesmo.

Este processo teve como principal objetivo a compreensão do domínio do problema, elucidando o processo de detecção de fraudes. Ao final desta etapa, foi gerado um documento - demonstrado na Tabela 6 - descrevendo o processo de análise e, para cada informação analisada, atribuiu-se um peso que demonstra a importância da informação para a tomada de decisão. Esta estratégia foi a que se mostrou mais eficaz para entendimento da importância de cada informação no processo de análise das fraudes na opinião dos especialistas.

Tabela 6 - Descrição de análises e importâncias segundo os especialistas

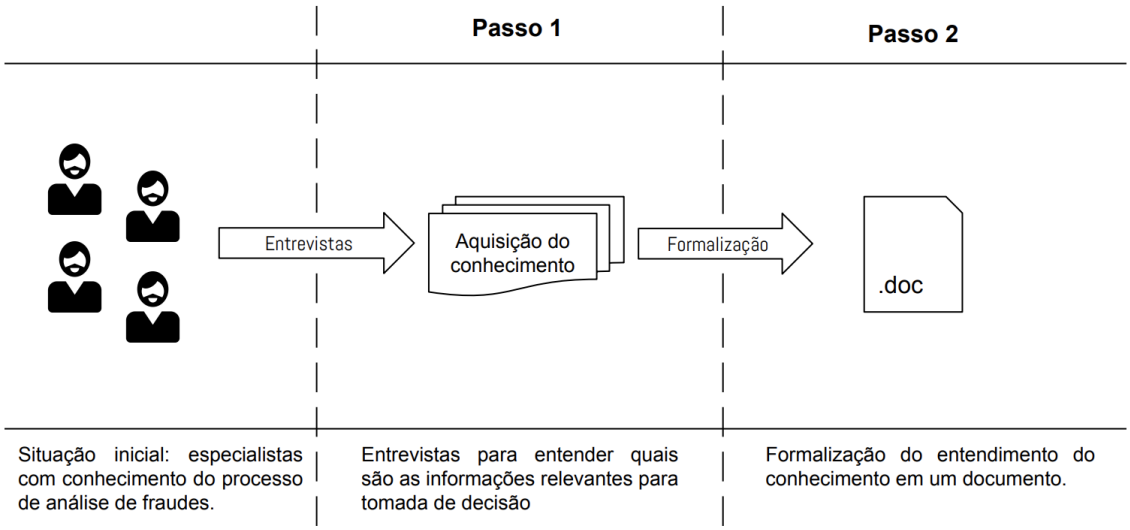
Tipo	Análise	Importância
Pedido	Loja (*site*) em que o pedido foi realizado	75
	Endereço IP, quantidade de *chargebacks* com IP	100
	Total (valor do pedido)	90
	Data e hora (valores discretizados)	70
Cliente	Tempo de criação da conta na loja	10
	Ticket médio (gasto médio, soma do total dos pedidos dividido pela quantidade)	95
	Quantidade de compras realizadas	85
	Proporção da quantidade de pedidos cancelados e pedidos completos	80
	Data da última compra do comprador	20
	Quantidade de compras realizadas nas últimas 24h	35
	Proporção da quantidade de pedidos cancelados e pedidos completos nas últimas 24h	30
Produtos	Quantidade de produtos comprados	25
	Proporção de quantidade e tipos de produtos comprados	20
Endereço de entrega	Quantidade de clientes diferentes que utilizaram o mesmo endereço do pedido	60
	Quantidade de pedidos enviados pelo cliente para o endereço	55
	Quantidade de endereços diferentes utilizados pelo cliente	50
Forma de pagamento	Quantidade de pedidos que o cliente realizou com a mesma forma de pagamento	45
	Quantidade de formas de pagamento diferentes utilizadas pelo cliente	40

Fonte: Elaborado pelo autor.

Após formalizada, a documentação foi enviada aos especialistas para aprovação. Além disso, as conclusões extraídas de cada entrevista foram verificadas por todos os analistas envolvidos no processo. Dessa forma, a representação do conhecimento foi revisada e aprovada por todos os especialistas previamente entrevistados.

O processo de entendimento do negócio é exibido na Figura 7.

Figura 7 - Processo de entendimento do negócio.



Fonte: Elaborado pelo autor.

4.2 ENTENDIMENTO DOS DADOS

A fase de análise dos dados originais consistiu nas seguintes fases:

a) Análise prévia dos dados originais e seleção de principais informações;

b) Aplicação de transformações (filtros e criação de novas variáveis derivadas);

c) Adequação das bases de dados para modelagem do classificador.

Portanto, a primeira atividade desta etapa consistiu em analisar as informações originais armazenadas no banco de dados da plataforma de *e-commerce* e no banco do sistema de ERP[4] da empresa estudada.

Das informações desses dois sistemas, foram selecionados os campos e tabelas que estavam dentro do escopo de interesse do trabalho, ou seja, foram selecionadas todas as tabelas e campos relacionados a dados de pedido, perfil do cliente, produtos, formas de pagamento e endereço de entrega.

4.2.1 Dados

As informações foram coletadas considerando informações dos pedidos realizados entre janeiro de 2015 e fevereiro de 2016. Ao todo, o conjunto de dados possui 139 variáveis originais disponíveis e 3.191.889 transações, sendo 53.040 fraudes.

A Tabela 7 detalha análise da quantidade de variáveis originais disponíveis e de registros coletados relacionados a cada entidade do domínio do *e-commerce*. Os dados coletados e as relações entre as entidades de pedido, cliente, endereço de entrega, forma de pagamento e produtos são mostrados no Apêndice C.

Tabela 7 - Descritivo quantitativo dos dados.

Tipo (entidade)	Quantidade de variáveis originais disponíveis	Quantidade de registros
Cliente	10	462.151
Pedido	16	3.191.889
Endereço de entrega	11	3.191.889
Forma de pagamento	12	5.433.952
Produto	90	6.949

Fonte: Elaborado pelo autor.

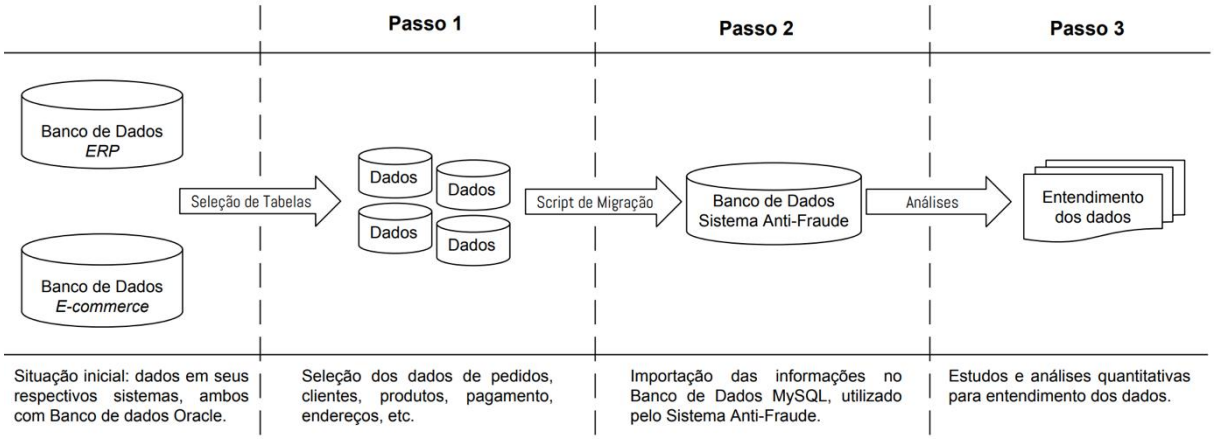
[4] O acrônimo ERP do inglês *enterprise resource planning,* significa planejamento de recursos empresariais e refere-se aos sistemas e pacotes de software utilizados pelas organizações para gerenciar as atividades comerciais do dia-a-dia, como contabilidade, compras, gerenciamento de projetos e fabricação.

A quantidade de registros difere de acordo com a entidade porque um único cliente pode realizar um ou mais pedidos e esse pedido terá sempre um endereço de entrega, com uma ou mais formas de pagamento (caso o cliente deseja fazer o pagamento com mais de uma forma de pagamento, exemplo: dois cartões de crédito, ou parte em boleto e parte em cartão de crédito, ou até pagar parte do pedido utilizando cupons de desconto).

Obviamente, nem todas as variáveis originais disponíveis serão relevantes para o processo de classificação. Porém, neste ponto do trabalho, a preocupação é que a base de dados esteja íntegra e que contenha as informações necessárias para o processo de classificação.

Todo o processo de entendimento dos dados é mostrado na Figura 8.

Figura 8 - Processo de entendimento dos dados.



Fonte: Elaborado pelo autor.

4.3 PREPARAÇÃO DOS DADOS

Após a coleta e o entendimento dos dados, foi feita a preparação dos dados com o objetivo de deixá-los prontos para a modelagem dos classificadores. O trabalho consistiu em duas etapas, que são detalhadas nas subseções seguintes: (i) representação do conhecimento dos especialistas em atributos, e; (ii) geração das informações no formato esperado para a classificação.

4.3.1 Representação do conhecimento dos especialistas em atributos

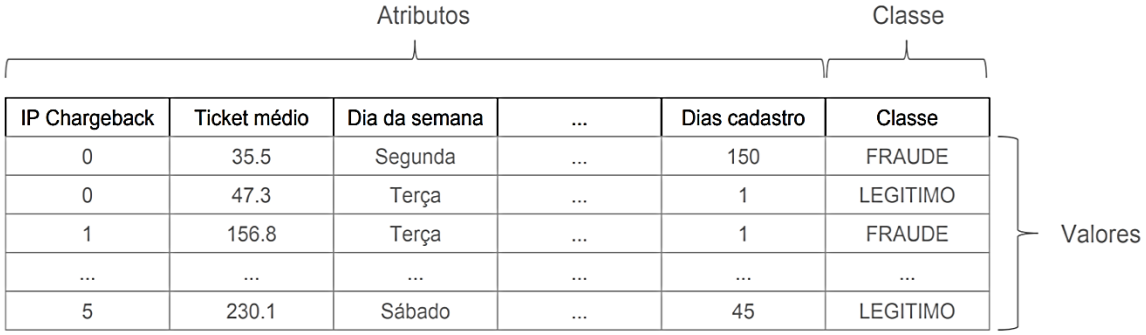
Na fase de conhecimento de negócio, foi desenvolvida uma documentação com todas as informações importantes para a classificação do pedido em conjunto com os especialistas de domínio. A partir dessa documentação, foi possível inferir as informações (representados pelos atributos do pedido, cliente, produto, endereço de entrega, forma de pagamento) utilizadas classificar o pedido como fraudulento ou legítimo.

4.3.2 Preparação dos arquivos para classificação

Posteriormente à preparação de todas as variáveis necessárias para a modelagem, foram gerados os arquivos necessários para a classificação.

Para isso, foi feita uma consulta no banco de dados do sistema antifraude proposto, considerando todas as informações sugeridas pelos especialistas. Essa consulta foi exportada no formato CSV, conforme exibido na Figura 9.

Figura 9 - Modelo do arquivo CSV gerado para classificação no Weka.



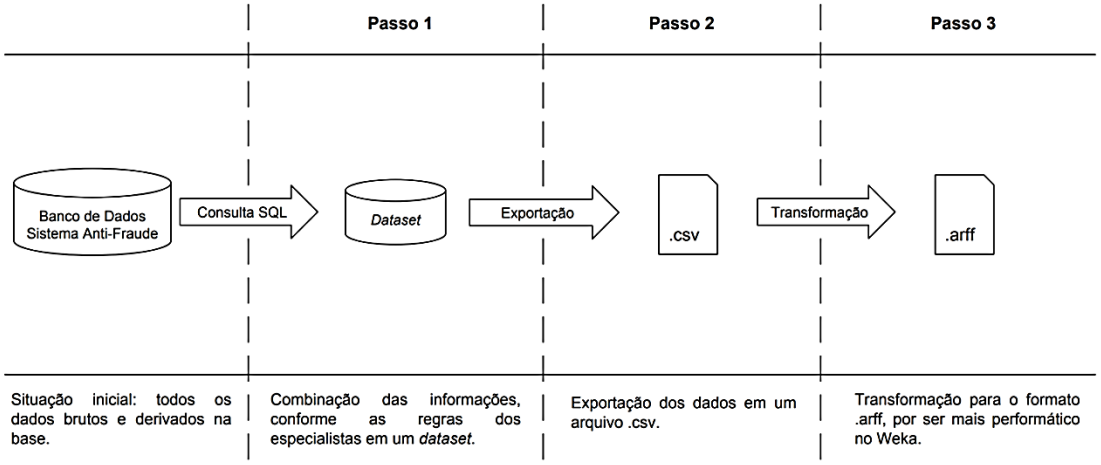
Fonte: Elaborado pelo autor.

Por motivos de desempenho do Weka, foi necessário fazer a transformação do arquivo .CSV para. ARFF. Dessa forma, seguindo as recomendações da ferramenta,

o classificador consegue apresentar os resultados mais rapidamente pois consegue fazer o carregamento das informações de forma gradual, o que não acontece caso seja utilizado o arquivo .CSV. A conversão dos arquivos foi feita utilizando o próprio software Weka.

O processo completo da fase de preparação dos dados elaborada no projeto é detalhado na Figura 10.

Figura 10 - Processo de preparação dos dados.



Fonte: Elaborado pelo autor.

4.4 MODELAGEM

Na fase de modelagem, foram feitos: (i) separação dos conjuntos de para cada experimento; (ii) a amostragem; e por fim, (iii) a classificação dos pedidos. Ambas as etapas são descritas nas próximas subseções.

4.4.1 Separação dos conjuntos de cada experimento

Nesta etapa, os atributos foram ordenados de acordo com sua importância para o processo de classificação.

A partir da ordenação dos atributos de acordo com a importância determinada pelos especialistas, os conjuntos para classificação foram separados da seguinte forma:

a) Primeiro conjunto: 10 atributos mais importantes;

b) Segundo conjunto: 15 atributos mais importantes;

c) Terceiro conjunto: Todos os 20 atributos sugeridos pelos especialistas.

Essa proposta de seleção dos atributos tem como objetivos: (i) entender se os especialistas estão de fato analisando os atributos mais importantes; (ii) possibilitar o treinamento mais rápido dos classificadores, uma vez que utilizar todos os atributos seria mais custoso computacionalmente; e (iii) facilitar a compreensão dos resultados.

Os atributos selecionados pelos especialistas são posteriormente comparados com os sugeridos pela árvore de decisão e pelo classificador baseado em regras.

4.4.2 Amostragem

Ainda na etapa de conhecimento dos dados, foi feito o processo de amostragem, ou seja, dentre todas as transações a serem classificadas, foi feita uma seleção de algumas transações para servirem de base para a modelagem do sistema de detecção de fraudes.

Esse processo é importante, pois a partir da análise de uma quantidade menor de registros é possível fazer uma prova de conceito do modelo desenvolvido em um tempo menor. Posteriormente os dados amostrados são validados.

Como descrito no Capítulo 2, uma das características do problema de detecção de fraudes é o fenômeno de classes desbalanceadas. Isso significa que há relativamente poucas transações fraudulentas e muitas transações legítimas.

No caso do presente trabalho, a base de dados possui 53.040 pedidos fraudulentos e 3.138.849 pedidos legítimos, ou seja, pouco mais de 59 pedidos legítimos para

cada fraude, que representa uma proporção de apenas 1,68% de pedidos fraudulentos em toda base de dados.

Conforme citado no Capítulo 2, dada a característica de classes desbalanceadas do problema de fraude, não é recomendável manter os dados amostrados com mesma proporção de classes da base de dados. Dessa forma, a amostragem deve ter uma distribuição de pedidos fraudulentos e pedidos legítimos diferente da proporção da base de dados, que no caso do presente trabalho, é de 1,68%.

Para fazer a estratificação da base, portanto, ao invés de fazer uma amostragem aleatória simples, optou-se por seguir o processo apresentado em Oliveira (2016), com uma análise sensível ao custo, proposto inicialmente por Elkan (2001), também mencionado no Capítulo 2.

Tem-se como quantidade de transações legítimas da amostragem, portanto:

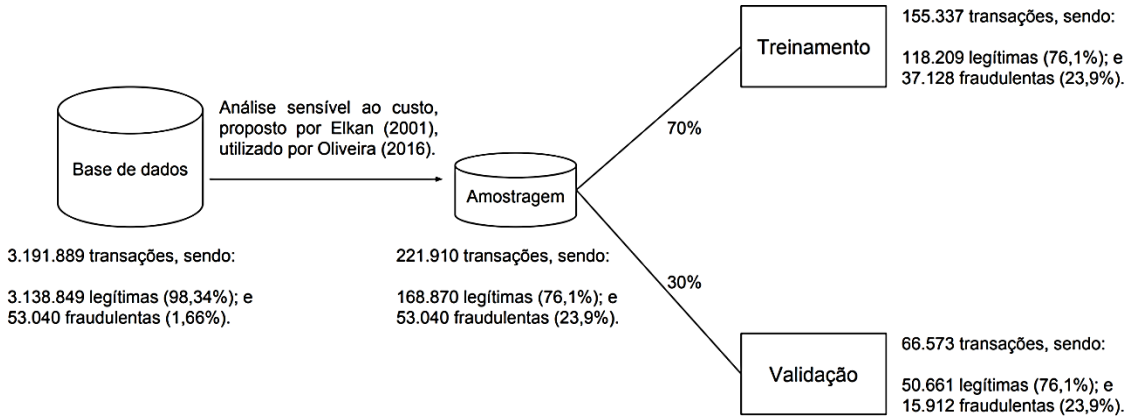
$$Quantidade\ transações\ legítimas\ da\ amostragem = Transações\ legítimas\ da\ base * \frac{C_{FP}}{C_{FN}}$$

$$Quantidade\ transações\ legítimas\ da\ amostragem = 3.138.849\ *\ 0{,}0538$$

$$Quantidade\ transações\ legítimas\ da\ amostragem = 168.870$$

Depois de determinar a quantidade de transações de cada classe no processo de amostragem, a amostra foi separada duas bases, sendo uma de treinamento, que representa 70% da quantidade total de transações da amostra; e a outra base de validação, que representa 30%, conforme exibido na Figura 11.

Figura 11 - Processo de amostragem de dados.



Fonte: Elaborado pelo autor.

Após a definição da composição das bases de treinamento e validação, bem como a respectiva quantidade de transações para cada classe, fez-se uma coleta aleatória na base de dados de acordo com a classificação da transação.

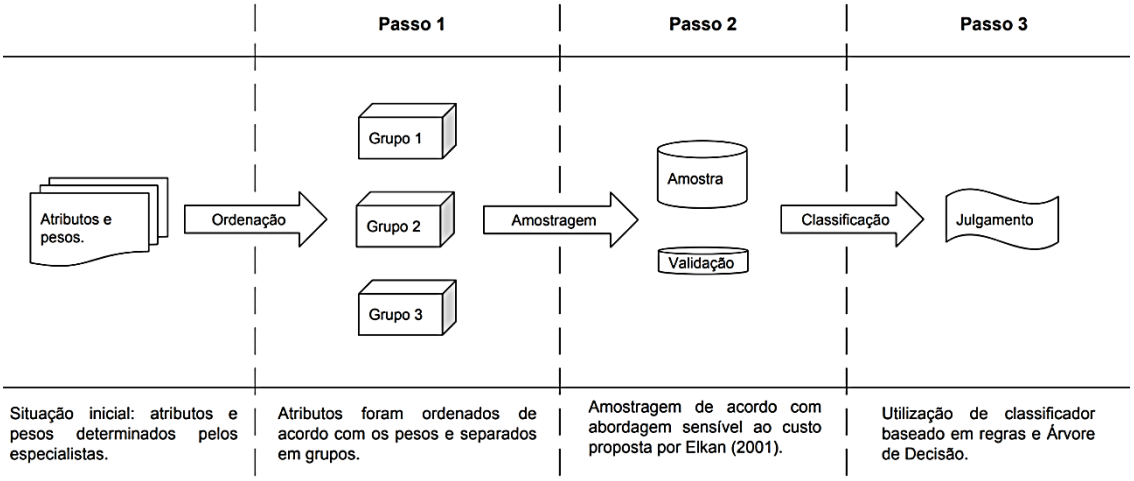
4.4.3 Classificação

A última etapa do processo de modelagem consistiu no processo de classificação dos registros. Os classificadores escolhidos neste processo foram: (i) classificador baseado em regras, utilizando o algoritmo RIPPER; (ii) árvore de decisão (algoritmo C4.5), e; (iii) *Naive Bayes*.

O Capítulo 2 descreve com detalhes as características de cada classificador utilizado. Os motivos que levaram à escolha dos classificadores, bem como suas vantagens e desvantagens são discutidas também no Capítulo 3.

O processo completo da modelagem realizada no presente trabalho é mostrado na figura 12.

Figura 12 - Processo de modelagem.



Fonte: Elaborado pelo autor.

A fim de facilitar a comparação entre os classificadores, toda classificação foi realizada adotando os parâmetros padrões do Weka, ou seja, não foi feito nenhum processo de otimização (*tuning*) dos classificadores baseados em regra, árvore de decisão e *naïve bayes*.

4.5 AVALIAÇÃO

Após o processo de modelagem, foi realizado processo de treinamento dos classificadores em todos os três conjuntos (com 10, 15 e 20 atributos respectivamente). Todos os conjuntos, portanto, foram testados, utilizando a mesma estratégia de amostragem, começando pela classificação baseada em regras, seguido da classificação por árvore de decisão e, por fim, utilizando Naive Bayes.

Posteriormente, cada grupo foi submetido às amostras de validação, que, conforme definido na fase de modelagem, equivalem a 30% da amostra original, contendo 22.910 transações, sendo 168.870 legítimas (76,1%) e 53.040 fraudulentas (23,9%).

O objetivo desta fase é conseguir inferir a capacidade de generalização dos modelos utilizados, já que as amostras de validação não foram utilizadas durante o processo de treinamento.

Assim sendo, se o classificador apresentar métricas de desempenho satisfatórias em relação à amostra de validação, pode-se inferir que a capacidade de generalização do modelo é boa e o trabalho de modelagem foi bem-sucedido.

A figura 13 apresenta as matrizes de confusão para cada classificação das amostras de validação de cada um dos conjuntos de dados utilizados no trabalho.

Figura 13 - Matrizes de confusão do processo da amostra de validação.

Conjunto 1, com 10 atributos.

Classificador Baseados em Regras		
Classificada	Real	
	Fraude	Legítimo
Fraude	9.802	6.025
Legítmo	3.475	47.271

Árvore de Decisão		
Classificada	Real	
	Fraude	Legítimo
Fraude	11.061	4.766
Legítmo	3.427	47.319

Naive Bayes		
Classificada	Real	
	Fraude	Legítimo
Fraude	2.751	13.076
Legítmo	1.870	48.876

Conjunto 2, com 15 atributos.

Classificador Baseados em Regras		
Classificada	Real	
	Fraude	Legítimo
Fraude	12.027	3.800
Legítmo	2.368	48.378

Árvore de Decisão		
Classificada	Real	
	Fraude	Legítimo
Fraude	12.749	3.078
Legítmo	2.252	48.494

Naive Bayes		
Classificada	Real	
	Fraude	Legítimo
Fraude	5.041	10.786
Legítmo	2.604	48.142

Conjunto 3, com 20 atributos.

Classificador Baseados em Regras		
Classificada	Real	
	Fraude	Legítimo
Fraude	12.426	3.401
Legítmo	2.197	48.549

Árvore de Decisão		
Classificada	Real	
	Fraude	Legítimo
Fraude	13.015	2.812
Legítmo	2.279	48.467

Naive Bayes		
Classificada	Real	
	Fraude	Legítimo
Fraude	4.951	10.876
Legítmo	2.615	48.131

Fonte: Elaborado pelo autor.

A matriz de confusão confronta a quantidade de amostras reais - ou seja, realmente pertencentes às classes fraude ou legítimo - com a quantidade de amostras que foram classificadas pelo modelo. Esses valores permitem que diversas métricas de desempenho sejam calculadas.

As métricas escolhidas para avaliar os classificadores, conforme descritas no capítulo de revisão bibliográfica, foram:

a) Cobertura ou taxa de verdadeiros-positivos;

b) Taxa de falsos positivos;

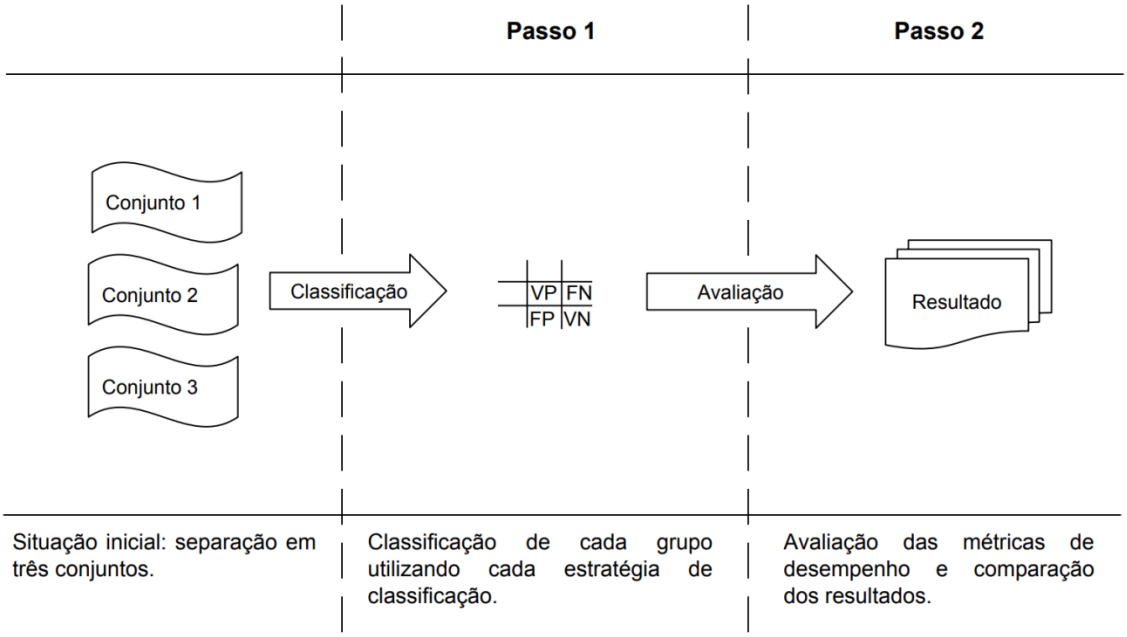
c) Precisão;

d) Medida F; e

e) Área sob a curva ROC.

A Figura 14 detalha o processo de avaliação utilizado no projeto.

Figura 14 - Fase de avaliação.



Fonte: Elaborado pelo autor.

4.6 IMPLANTAÇÃO

Depois do estudo de análise e comparação dos classificadores realizado na fase de avaliação, foi iniciada a fase de implantação, cujo objetivo é planejar como o conhecimento gerado pode ser utilizado dentro da organização.

É válido ressaltar que, no domínio do comércio eletrônico, os sistemas convencionais de detecção de fraude geralmente são fornecidos como um serviço

pelas adquirentes. Nesses casos, portanto, sempre que uma compra é feita em uma loja na *web*, os parâmetros de pagamento selecionados dessa transação são enviados para a adquirente, que é responsável por verificar se a transação é fraudulenta ou não e depois executar o pagamento. O resultado desse processo é enviado de volta ao sistema do comércio eletrônico para que o comprador possa receber o *feedback* se a compra foi aprovada ou cancelada.

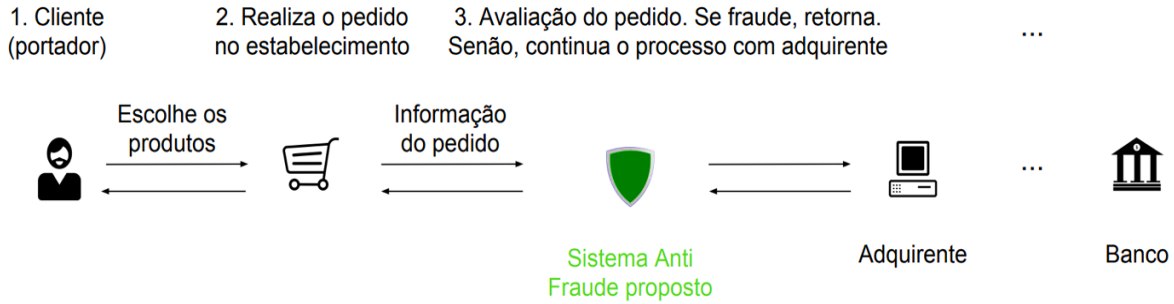
No entanto, por serem serviços de pagamento, esses sistemas convencionais comumente usam apenas parâmetros de informações restritos ao domínio de pagamento para detectar fraudes. Assim, as informações específicas do comércio eletrônico, como os produtos comprados pelo cliente, o endereço de entrega escolhido, os dados da conta e a informação do perfil do comprador, as avaliações dos produtos, o comportamento de busca e navegação, as promoções, o uso dos cupons de desconto, etc., são completamente ignorados.

O presente trabalho, no entanto, apresenta um estudo sobre detecção de fraudes que mostra a importância da utilização de informações do domínio do *e-commerce* no processo de classificação das transações.

Adicionalmente, foi desenvolvido um sistema de apoio à detecção de fraudes, utilizando o classificador baseado em regras apresentado no presente trabalho. O sistema foi desenvolvido de forma modular, para que seja fácil a adaptação do classificador utilizado. Portanto, é possível alterar o sistema para utilizar outros classificadores posteriormente.

Essa estratégia também permite que o sistema seja adicionado à estrutura do *e-commerce* tanto como sendo a única solução de proteção à fraude, conforme mostrado na Figura 15, ou então, como sendo uma segunda camada de proteção, fazendo análises posteriormente às transações já analisadas pela solução antifraude adotada pela adquirente, como detalha a Figura 16.

Figura 15 - Arquitetura com sistema proposto como camada única de proteção.



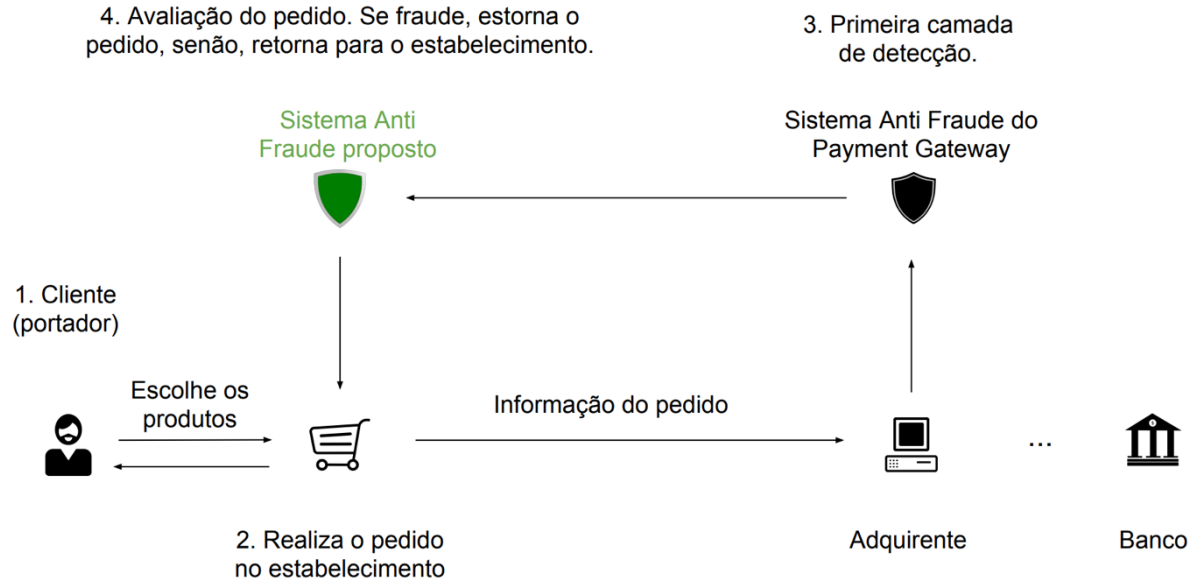
Fonte: Elaborado pelo autor.

A arquitetura apresentada na Figura 15, portanto, sugere que toda a análise de fraude fique sob responsabilidade do sistema proposto no presente trabalho. Assim sendo, após o cliente realizar o pedido, antes do estabelecimento (no caso do presente trabalho, o próprio *e-commerce*) passar as informações para a adquirente, será feita a avaliação do pedido de modo que, se foi identificado que se trata de uma fraude, o estabelecimento retorna a informação de pagamento negado para o cliente mesmo sem que essa informação tenha sido informada pela adquirente, que seria o processo normal caso o pedido fosse identificado como legítimo.

Já a arquitetura apresentada na Figura 16 sugere que o sistema proposto funcione como uma segunda camada de verificação de fraude. Neste processo, após o cliente fazer o pedido, o estabelecimento envia a informação do pedido para a adquirente, que realiza a primeira etapa de análise de fraude. Caso o pedido seja aprovado, antes da informação ser retornada para o estabelecimento, o pedido passa por uma segunda camada de validação, realizada pelo sistema proposto. Uma nova avaliação do pedido é feita e, se identificado como fraude, o pedido é estornado na adquirente, caso contrário, o pedido é aprovado. Por fim, o estabelecimento retorna o resultado desse processo para o cliente.

O Apêndice D apresenta imagens das principais telas do sistema desenvolvido.

Figura 16 - Sistema proposto como camada adicional de proteção.



Fonte: Elaborado pelo autor

5 RESULTADOS

Após a classificação dos três conjuntos de dados para cada modelo escolhido, foi feita uma comparação dos resultados obtidos. Essa comparação é suportada por uma série de indicadores de desempenho, amplamente discutidos no Capítulo 2, calculados a partir das matrizes de confusão apresentadas na fase de avaliação.

Para as métricas de cobertura, precisão, medida F e AROC, quanto maiores seus valores, melhor é o resultado final. Por outro lado, para a métrica de taxa de falso positivo, quanto menor seu valor, melhor. O comparativo completo dos indicadores de desempenho frente à amostra de validação dos conjuntos 1, 2 e 3 são exibidos nas Tabelas 8, 9 e 10, respectivamente. Em cada tabela, destaca-se em negrito a melhor métrica de cada classificador.

Tabela 8 - Comparativo dos indicadores de desempenho do Conjunto 1.

Conjunto 1, com 10 atributos			
Métrica	**Baseado em regras**	**Árvore de Decisão**	***Naive Bayes***
Sensibilidade	0.857	**0.877**	0.775
Precisão	0.852	**0.874**	0.743
Taxa de falso alarme	0.306	**0.246**	0.639
Medida F	0.853	**0.875**	0.725
AROC	0.785	**0.912**	0.776

Fonte: Elaborado pelo autor.

É importante ressaltar que a conclusão de qual classificador é o mais adequado envolve diversos aspectos, portanto, fazer essa análise tomando como base apenas uma métrica isolada não é aconselhável.

Tabela 9 - Comparativo dos indicadores de desempenho do Conjunto 2

Conjunto 2, com 15 atributos			
Métrica	**Baseado em regras**	**Árvore de Decisão**	***Naive Bayes***
Sensibilidade	0.907	**0.920**	0.799
Precisão	0.835	**0.850**	0.659
Taxa de falso alarme	0.194	**0.159**	0.532
Medida F	0.906	**0.919**	0.771
AROC	0.866	**0.942**	0.805

Fonte: Elaborado pelo autor.

Tabela 10 - Comparativo dos indicadores de desempenho do Conjunto 3

Conjunto 3, com 20 atributos			
Métrica	**Baseado em regras**	**Árvore de Decisão**	***Naive Bayes***
Sensibilidade	0.916	**0.924**	0.797
Precisão	0.914	**0.923**	0.777
Taxa de falso alarme	0.174	**0.146**	0.536
Medida F	0.915	**0.923**	0.769
AROC	0.879	**0.930**	0.807

Fonte: Elaborado pelo autor.

Na detecção de fraude, portanto, uma das métricas mais importantes é a cobertura (também referenciada como *recall* ou taxa de verdadeiros-positivos*)* ou, no caso do

presente trabalho, taxa de detecção de fraude, uma vez que a perda devida à fraude depende dessa métrica. Outra medida importante é a taxa de falso positivo, uma vez que mostra a insatisfação do cliente por falso alarme - são casos em que a transação é legal, mas o classificador "entende", erradamente, que se trata de fraude.

Em todos os conjuntos, o classificador *Naive Bayes* apresentou os resultados menos satisfatórios em quase todas as métricas, principalmente em relação à cobertura e precisão de classes fraudulentas e sua taxa de falso alarme. Apesar de haver uma série de tratativas a serem feitas para melhorar seu desempenho - como utilizar o estimador de Laplace para o problema de frequência zero, bem como revisar os atributos utilizados para garantir que eles sejam independentes - fica evidente que os demais classificadores utilizados no trabalho se adaptaram melhor à característica do problema de detecção de fraudes.

O modelo que utilizou o classificador árvore de decisão apresentou os melhores resultados em todas as métricas, inclusive, com valor AROC considerado excelente mesmo nos conjuntos 1 e 2, que possuem menor número de atributos.

Apesar desse resultado parecer excelente mesmo para os conjuntos com menos atributos, a complexidade das árvores geradas em relação à quantidade de níveis levantou suspeitas de *overfitting*. É preciso entender com mais profundidade, portanto, se o classificador se ajusta bem ao conjunto de dados observado, mas se mostra ineficaz para prever novos resultados. Essa análise pode ser feita de várias formas, tais como: (i) ajustar a medida F; (ii) adotar estratégias de poda da árvore de decisão; (iii) propor novas estratégias de amostragem para testar o classificador com dados diferentes e analisar seus resultados, entre outras.

Após a realização dos experimentos, pode-se avaliar que o bom desempenho dos classificadores em relação à medida AROC é fundamental para a discussão de algumas hipóteses em relação à metodologia aplicada e sua eficácia na resolução do problema.

Portanto, com base no bom desempenho dos classificadores, pode-se afirmar que o conhecimento dos especialistas é relevante para a detecção de fraudes e que a metodologia utilizada foi importante no processo de encapsular esse conhecimento

nos modelos de classificação. Esta afirmação é reforçada ao observar que os classificadores do presente trabalho apresentam desempenho melhor que os apresentados em Felipe Junior et al (2012), Santiago (2015) e Oliveira (2016).

Pode-se afirmar também que a eficiência econômica permite avaliar a qualidade dos resultados. Conforme sugerido por Felipe Junior (2012), geralmente as melhores eficiências econômicas estarão associadas aos melhores modelos. No entanto, essa medida pode tender para modelos que identificam casos de fraude de maiores valores. Sendo assim, é interessante que ela seja utilizada em conjunto com outras medidas de avaliação do modelo como a precisão e a cobertura, garantindo que os melhores modelos com os melhores ganhos financeiros sejam selecionados.

5.2 EFICIÊNCIA ECONÔMICA

Conforme relatado no Capítulo 2, a melhor solução não é necessariamente aquela que detecta muitas fraudes, mas sim aquela que detecta fraudes com prejuízo em potencial maior (DUMAN; OZCELIK, 2011).

Dessa forma, foi proposto um cálculo de eficiência financeira, adaptado de Caldeira et al. (2012), também utilizado por Santiago (2015). A Tabela 11 exemplifica o cálculo da eficiência financeira de uma transação cujo valor é de R$ 1.000,00 e o lucro praticado pelo estabelecimento é de 20%.

Tabela 11 - Exemplo do cálculo de eficiência financeira.

Situação	Impacto	Cálculo	Resultado
1	Ganho	Lucro * Valor pedido	Ganho de R$ 200,00.
2	Perda	Lucro * Valor pedido	Perda de R$ 200,00.
3	Nulo		Nem ganho, nem perda.
4	Perda	(1-Lucro) * Valor pedido	Perda de R$ 800,00.

Fonte: Elaborado pelo autor.

Dessa forma, cada transação foi analisada de acordo com sua situação, obtendo-se o lucro ou prejuízo para cada uma delas. Por fim, os resultados de todas as transações são somados e comparados com aqueles obtidos atualmente pela empresa, chegando assim na medida de eficiência econômica - ou eficiência financeira, definida por:

$$EE = \frac{(EE_i - EE_e)}{EE_e} \quad (16)$$

Em que:

EE_t: representa o resultado econômico obtido por meio dos experimentos com o procedimento apresentado no presente trabalho;

EE_e: representa o resultado econômico atual da empresa.

A Tabela 12 detalha a eficiência econômica obtida em cada classificador.

Tabela 12 - Eficiência econômica de cada classificador

Conjunto	**Baseado em regras**	**Árvore de Decisão**	***Naive Bayes***
Conjunto 1	-12,56%	**-8,45%**	-26,20%
Conjunto 2	1,74%	**3,15%**	-18,56%
Conjunto 3	2,78%	**3,15%**	-11,72%

Fonte: Elaborado pelo autor.

Percebe-se que há uma variação significativa nos resultados de acordo com o conjunto e o classificador utilizado. Com isso, para casos semelhantes ao considerado no presente trabalho - em que não se assume custo no processo de análise manual de classificação - pode-se afirmar em relação à economia que a taxa de falso alarme baixa em um sistema de detecção de fraudes é tão ou mais importante que a taxa de cobertura (ou detecção de fraudes).

É importante lembrar, contudo, que há uma incerteza nos resultados obtidos que são as transações não aprovadas pelo processo de análise da empresa. Uma transação não ser aprovada pela empresa não necessariamente significa que se tratava de uma fraude e, consequentemente, geraria *chargeback*, o que impede de fazer uma afirmação mais assertiva em respeito à análise em questão. Por isso, assim como sugerido por Junior et al. (2012), nos cálculos realizados, são desconsideradas as transações não aprovadas.

5.2 IMPORTÂNCIA PARA O NEGÓCIO

Além do bom desempenho, os classificadores baseados em regra e árvore de decisão apresentaram características importantes para o negócio de detecção de fraudes em *e-commerce*.

A partir da interpretação das regras utilizadas pelo modelo baseado em regras e pelo modelo de árvore de decisão (também mostrados nos Apêndices E e F), foi possível inferir alguns comportamentos que trouxeram resultados indiretos:

a) O comportamento de fraude varia de acordo com a loja em que se está fazendo o pedido. A empresa de *e-commerce* do presente estudo possui 9 lojas (ou *sites* de *e-commerce*), direcionados a públicos diferentes e com produtos diferentes. Essa inferência pode ajudar a empresa de estudo a concentrar maiores esforços de análise em lojas cujo segmento atrai mais o perfil de fraudadores.

b) Em um mesmo segmento, é possível perceber que determinados tipos de produto atraem mais consumidores fraudulentos.

c) Informações de frequência e de velocidade de compra são muito relevantes para o processo de análise de fraude, assim como a taxa de pedidos cancelados do cliente.

d) O trabalho dos especialistas pode ser ainda mais eficiente, pois há informações sugeridas como importantes para a classificação da transação, tais como a quantidade de compras, o dia da semana, a hora em que o pedido foi realizado e a quantidade de produtos comprados, não se mostraram tão relevantes para o processo de detecção de fraudes.

e) A interpretação das regras também ajudou a identificar relações pouco intuitivas entre os argumentos utilizados: por exemplo, a relação entre a quantidade de vezes que o endereço de entrega foi utilizado pelo cliente e o dia da última compra realizada. Esses padrões identificados foram fundamentais para inspirar novas análises por parte dos especialistas.

Os estudos realizados nessa etapa final foram fundamentais para a identificação de um comportamento de fraude muito difícil de tratar: o caso de roubo de conta. Neste tipo de ataque, conforme citado com mais detalhes no capítulo de referência bibliográfica, o fraudador tem acesso à conta da vítima e, com isso, consegue fazer compras utilizando dados de cartão de crédito gravados na conta do cliente invadido.

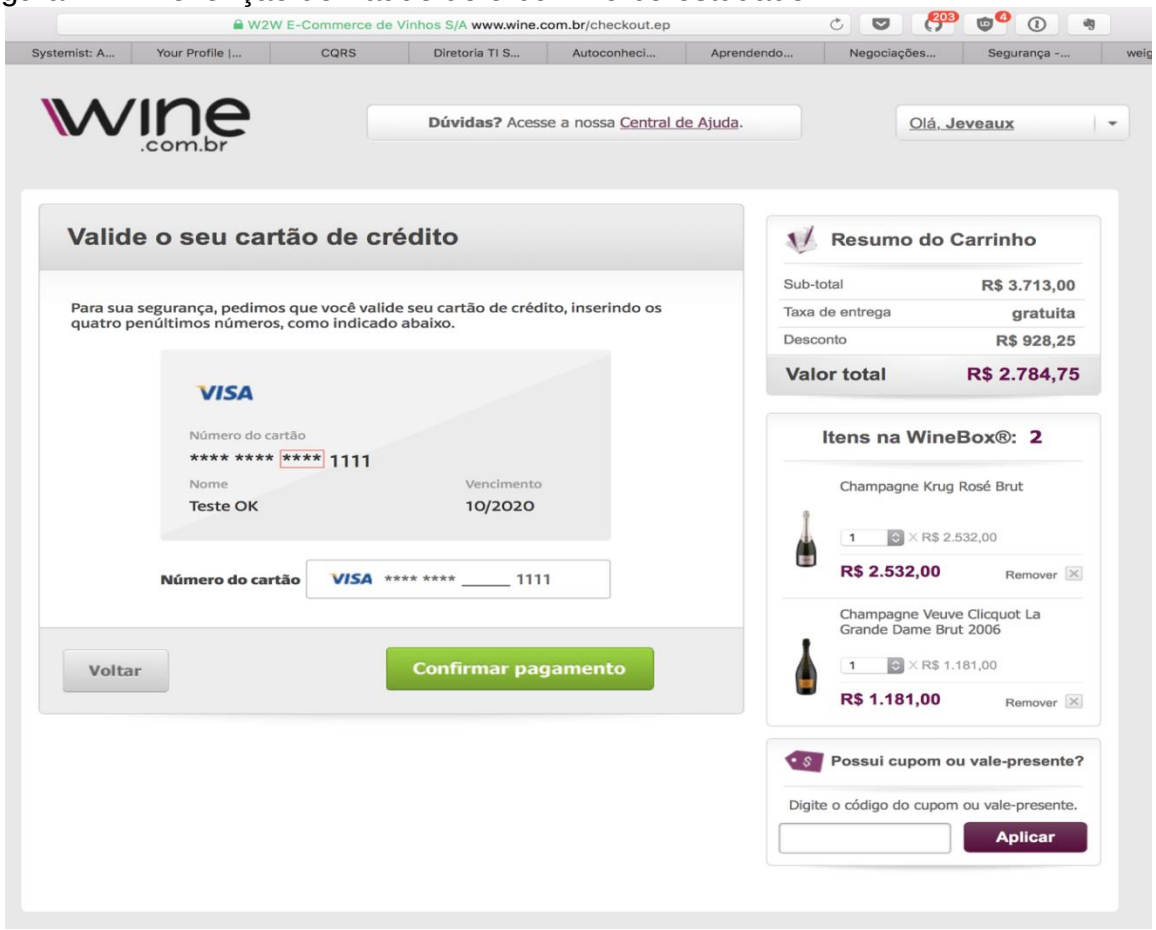
O fraudador, portanto, faz o pedido com a conta do cliente invadido utilizando seu cartão de crédito, porém, com padrões levemente diferentes, como a compra de produtos suspeitos ou utilização do endereço de entrega diferente do histórico de pedidos.

A identificação desse comportamento ajudou o *e-commerce* estudado a desenvolver uma solução de prevenção de fraude que foi implantada e incorporada ao processo de compra, conforme mostra a Figura 17.

O novo sistema de prevenção de fraudes foi desenvolvido de modo a exigir informações adicionais de segurança do cartão de crédito no momento de finalizar o pedido, caso o comportamento de compra do cliente em questão seja semelhante ao utilizado pelos invasores de contas.

Dessa forma, os invasores - apesar de terem total acesso à conta da vítima (cliente invadido) - não conseguem finalizar o pedido, uma vez que eles não têm tal informação.

Figura 17 - Prevenção de fraude do *e-commerce* estudado.



Fonte: Elaborado pelo autor.

6 CONCLUSÃO

Este trabalho descreveu o processo de desenvolvimento de um sistema de apoio à detecção de fraudes em um comércio eletrônico. Os resultados para este problema do mundo real ilustram que o uso deste tipo de sistema pode permitir a classificação precisa e inteligível de dados difíceis, principalmente ao utilizar os classificadores baseados em regras e árvore de decisão. Os resultados também mostram a importância de uma comissão de especialistas para assegurar que bons resultados sejam gerados.

Além disso, o desempenho do sistema de detecção de fraude proposto aplicado ao segundo conjunto de dados ainda é mais impressionante, pois analisa apenas supostos pedidos genuínos previamente analisados por um sistema convencional de detecção de fraude no lado do *gateway* de pagamento.

Conforme mencionado na Seção 2, os comportamentos fraudulentos mudam ao longo de um período de tempo. Isso pode degradar o desempenho do classificador de fraude. Portanto, o modelo de detecção de fraude deve ser adaptável a essas mudanças comportamentais.

No caso dos classificadores baseados em regras, essas mudanças comportamentais podem ser incorporadas pela adição de conhecimento ao atualizar o mecanismo de regras ou pela utilização de regras que utilizem técnicas de aprendizado.

A detecção de fraude é, por muitas vezes, uma importante estratégia para qualquer empresa. Por isso, espera-se que as empresas desejem ter maior controle sobre as regras e não deixar toda a responsabilidade para um sistema externo ou terceirizado. Portanto, este artigo também propõe uma nova arquitetura - mais adaptada às necessidades de negócio e de mercado - que propõe o combate à fraude em um sistema generalista, porém, podendo ser altamente customizado, que pertence à empresa de comércio eletrônico.

O método proposto também pode ser utilizado para representar visualmente e explicitamente as decisões mesmo quando a pontuação de fraude está próxima do

limiar de fraude, simplesmente interpretando as regras de fraude aplicadas a um pedido. Mesmo não especialistas foram capazes de entender a classificação após uma breve explicação. Essa característica foi fundamental para identificar fraudes de invasões de contas, que são impossíveis de detectar em sistemas antifraudes convencionais, como descrito na Seção 5.

Outra vantagem das técnicas de classificação baseada em regras e árvore de decisão é que elas requerem pouca preparação de dados em comparação com outras técnicas - que muitas vezes exigem a normalização de dados. Além disso, esses métodos se mostraram capaz de ligar com variáveis não estruturadas, trabalhando ora com dados numéricos, ora com variáveis categóricas e mostram ser úteis para modelar decisões humanas.

Apesar disso, a experiência com sistemas baseados em regras mostrou as seguintes desvantagens: a aquisição de conhecimento domínio dos especialistas - que é o processo de engenharia do conhecimento - é complexa e consome muito tempo: todas as possibilidades devem ser explicitamente enumeradas, e cabe ao desenvolvedor do sistema antifraude a capacidade de generalização do sistema. Além disso, manter o sistema atualizado e com alto poder de precisão não é fácil, uma vez que há um alto custo envolvido em sua revisão.

6.1 TRABALHOS FUTUROS

Muitas técnicas de detecção de padrões são citadas na literatura, como SVM, árvores de decisão, redes neurais, cadeias de markov, etc. Em Pozzolo et al. (2014), a combinação de diferentes estratégias foi utilizada para obter o melhor desempenho em diferentes conjuntos de dados. Dessa forma, uma maneira natural de evoluir o presente trabalho é compreender como extrair o melhor de cada técnica e integrá-las ao sistema.

Outra contribuição importante é facilitar a inclusão de feedback ao sistema. Muitos pedidos de fraude só são identificados dias após o pedido ser realizado, portanto, para evitar ruídos e melhorar o desempenho do classificador, é fundamental permitir

que uma classificação imediata possa ser alterada posteriormente. Essa característica, apesar de muito importante, não foi encontrada em nenhum trabalho estudado na revisão literária e, também por isso, pode vir a ser considerado um diferencial se implementada.

Para melhorar o processo de obtenção de conhecimento dos especialistas, técnicas de ontologia podem ser utilizadas para construir a máquina de regras do sistema e facilitar eventuais alterações e inclusões de novas regras.

O classificador de árvore de decisão apresentou suspeitas de *overfitting* no presente trabalho. Uma sugestão é explorar melhor técnicas para evitar esse comportamento, como a validação cruzada, por meio do ajuste da medida F.

Sugere-se também a otimização dos parâmetros dos classificadores para extrair o máximo de cada modelo de classificação. O presente trabalho utilizou apenas os parâmetros padrões do Weka para facilitar a comparação entre os classificadores, porém, o *tuning* (otimização) de cada classificador pode trazer resultados ainda melhores.

Outra sugestão para trabalhos futuros é a utilização de classificadores como *Random Forest*, que pode apresentar melhores desempenhos que o classificador de Árvore de Decisão.

REFERÊNCIAS

ABDALLAH, Aisha; MAAROF, Modh Aizaini; ZAINAL Anazida. Fraud detection system: a survey. **Journal of Network and Computer Applications**, 68, p. 90-113, 2016.

ACI Universal Payments. **2016 Global consumer card fraud**: where card fraud is coming from. Disponível em: <https://www.aciworldwide.com/-/media/files/collateral/trends/2016-global-consumer-card-fraud-where-card-fraud-is-coming-from.pdf>. Acesso em: mar. 2017.

AKHILOMEN, John. Data mining application for cyber credit-card fraud detection system. **Lecture Notes in Engineering and Computer Science**, p. 1537-1542, 2013.

ASSOCIAÇÃO BRASILEIRA DE COMÉRCIO ELETRÔNICO. **Relatório de Análise Macroeconômica do E-commerce 2017**. 2017. Disponível em: <http://abcomm.org/relatorio-analise-macroeconomica-ecommerce .php>. Acesso em: abr. 2017.

BARBOSA, Juliana Moreira; CARNEIRO, Tiago Garcia de Senna; TAVARES, Andrea Labrudi. **Métodos de classificação por árvores de decisão.** Ouro Preto, 2009. Trabalho de especialização (Disciplina de Projeto e análise de algoritmos) - Programa de Pós-Graduação em Ciência da Computação, Universidade Federal de Ouro Preto.

BEHDAD, Mohammad et al. Nature-inspired techniques in the context of fraud detection. **IEEE Transactions on Systems, Man, and Cybernetics**, Part C (Applications and Reviews), v. 42, n. 6, nov. 2012, p. 1273-1290, 2012.

BELO, Orlando; VIEIRA, Carlos. **Applying user signatures on fraud detection in telecommunications networks.** p. 286-299, 2011.

BHATLA, Tej Paul; PRAHU Vikram; DUA, Amit. Understanding credit card frauds. **Cards Business Review,** 2003.

BHATTACHARYYA, Siddhartha et al. Data mining for credit card fraud: a comparative study. **Decision Support Systems**, v. 50, n. 3, p. 602-613, 2011.

BOLTON, R. J.; HAND, D. J. Unsupervised profiling methods for fraud detection. **Proceedings of credit scoring credit control**, p. 235-255, 2001.

CALDEIRA, Evandro et al. **Characterizing and evaluating fraud in electronic transactions.** Eighth Latin America Web Congress.

CARVALHO, Hialo Muniz. **Aprendizado de máquina voltado para mineração de dados**: árvores de decisão. 2014. 68 f. Monografia (Graduação em Engenharia de Software) - Universidade de Brasília, Faculdade UNB Gama, Brasília, 2014.

CHAN P. K. et al. Distributed data mining in credit card fraud detection. **IEEE Intelligent Systems**, p. 67-74, 1999.

CHAN, P. K.; STOLFO, S. J. Toward scalable learning with non-uniform class and cost distributions: A case study in credit card fraud detection. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING. 4., 1998. **Proceedings...** p. 164-168, 1998.

CYBERSOURCE. **Online fraud report for latin america. Visa merchant sales & solutions 2016**. 2016. Disponível em: <https://www.cybersource.com/content/dam/cybersource/en-LAC/documents/Online_Fraud_Report_2016.pdf>. Acesso em: abr. 2017.

COELHO, L.; RAITTZ, R.; TREZUB, M. FControl®: sistema inteligente inovador para detecção de fraudes em operações de comércio eletrônico. **Gestão & Produção**, v. 13, n. 1, 2006.

DELAMAIRE, L.; ABDOU, H.; POINTON, J. Credit card fraud and detection techniques: a review. **Banks and Bank Systems**, n. 4. p. 57-68, 2009.

DESAI, Anita B.; DESHMUKH, Ravindra. Data mining techniques for fraud detection. **International Journal of Computer Science and Information Technologies**, v. 4, n. 1, p. 1-4, 2013.

DUMAN, Ekrem; OZCELIK, M. Hamdi. Detecting credit card fraud by genetic algorithm and scatter search. **Expert Systems with Applications**, v. 38, n. 10, p. 13057-13063, 2011.

FELIPE JÚNIOR, José. **Mineração de dados para detecção de fraudes em transações eletrônicas.** 2012. 110 f. Dissertação (Mestrado em Ciência da Computação) - Instituto de Ciências Exatas, Universidade Federal de Minas Gerais, Belo Horizonte, 2012.

GADI, Manoel Fernando Alonso. **Uma comparação de métodos de classificação aplicados à detecção de fraude em cartões de crédito.** 2008. 201 f. Dissertação (Mestrado em Ciência da Computação) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2008.

GAMA, João et al. A survey on concept drift adaptation. **ACM Computing Surveys**, v. 1, n. 1, Article 1, 2013.

HAND, David J.; CROWDER, Martin J. **Overcoming selectivity bias in evaluating new fraud detection systems for revolving credit operations.** Int. J. Forecast 28. 2012.

HILAS, S. Constantinos; SAHALOS, John N. **An application of decision trees for rule extraction towards telecommunications fraud detection.** In: INTERNATIONAL CONFERENCE ON KNOWLEDGE-BASED AND INTELLIGENT INFORMATION AND ENGINEERING SYSTEMS. Part of the Lecture Notes in Computer Science book series (LNCS, volume 4693), 2007.

HEJAZI, Maryamsadat; SINGH; Yashwant Prasad. Credit data fraud detection using kernel methods with support vector machine. **Journal of Advanced Computer Science and Technology Research**, n. 2, p. 35 49, 2012.

HOSMER, David W.; LEMESHOW, Stanley; STURDIVANT, Rodney X. **Applied logistic regression.** 3rd ed. [S.l.]: Wiley, 2013.

JHA, Sanjeev; GUILLEN, Montserrat; WESTLAND, J. Christopher. Employing transaction aggregation strategy to detect credit card fraud. **Expert Systems with Applications**, n. 39, p. 12650-12657, 2012.

JYOTHSNA, V.; RAMA PRASAD, V. V.; PRASAD K. A review of anomaly based intrusion detection systems. **International Journal of Computer Applications,** v. 28 n. 7, p. 26-35, 2011.

LEMOS, Eliane Prezepiorski; STEINER, Maria Teresinha Arns; NIEVOLA, Julio César. Análise de crédito bancário por meio de redes neurais e árvores de decisão: uma aplicação simples de data mining. **Revista de Administração**, São Paulo, v. 40, n. 3, p. 225-234, jul./ago./set. 2005.

MAGALLA, Asherry. **Security, prevention and detection of cyber crimes.** Tumaini University Iringa University College. Cyber Crime. 2013.

MASSA, Daniel; VALVERDE, Raul. A fraud detection system based on anomaly intrusion detection systems for E-Commerce applications. **Computer and Information Science**, v. 7, n. 2; 2014.

MITCHELL, Tom M. **Machine Learning.** [S.l.]: McGraw-Hill, 1997.

NGAI, E. W. T. et al. The application of data mining techniques in financial fraud detection: a classification framework and an academic review of literature. **Decision Support Systems,** n. 50. p. 559-569, 2011.

OLIVEIRA, Paulo Henrique Maestrello Assad. **Detecção de fraudes em cartões: um classificador baseado em regras de associação e regressão logística.** 2016. 103 f. Dissertação (Mestrado em Ciência da Computação) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2016.

PEDROSO, Louise; REVOREDO, Kate; BAIÃO, Fernanda. **Uma abordagem baseada em mineração de dados para apoio ao ciclo de vida de projetos de pesquisa e inovação.** In: SIMPÓSIO BRASILEIRO DE SISTEMAS DE INFORMAÇÃO (SBSI), 2013.

PHUA, C. et al. **Comprehensive survey of data mining-based fraud detection research.** In: INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTATION TECHNOLOGY AND AUTOMATION (ICICTA), v. 1, p. 50-53, 2010.

POZZOLO, Dal et al. Credit card fraud detection and concept-drift adaptation with delayed supervised information. In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS. **Proceedings...** IEEE, p. 1-8, 2015.

_____ et al. Learned lessons in credit card fraud detection from a practitioner perspective. **Expert Systems with Applications**, v. 41, n. 10, p. 4915-4928, 2014.

QIBEI, Li; CHUNHUA, Ju. Research on credit card fraud detection model based on class weighted support vector machine. **Journal of Convergence Information Technology**, v. 6, n. 1, jan. 2011.

QUEIROGA, Rodrigo Mendonça. **Uso de técnicas de Data Mining para detecção de fraudes em energia elétrica.** 2015. Dissertação (Metrado em Informática) - Universidade Federal do Espírito Santo, Vitória, 2015.

RAMOS, José Abílio de Paiva. **Árvores de Decisão Aplicadas à detecção de fraudes bancárias.** 2014. 54 f. Dissertação (Mestrado Profissional em Computação Aplicada) - Universidade de Brasília, Brasília, 2014.

RODRIGUES, André Iribure; LOPES, Paulo de Vitor Castilhos. Experiência para o cliente a partir de loja conceito: um estudo sobre o segmento de perfumaria e cosméticos. **Cadernos de Comunicação**, v. 19, n. 2, 2015.

SANTIAGO, Gabriel Preti. **Um processo para modelagem e aplicação de técnicas computacionais para detecção de fraudes em transações eletrônicas.** 2014.Dissertação (Mestrado em Ciência da Computação) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2014.

_____; PEREIRA, Adriano C. M., HIRATA JUNIOR, Roberto. A modeling approach for credit card fraud detection in electronic payment services. In: ACM SYMPOSIUM ON APPLIED COMPUTING. 30., 2015. **Proceedings...** p. 2328-2331, 2015.

SARAVANAN, P. et al. **Data Mining approach for subscription-fraud detection in telecommunication sector.** v. 7, n. 11, p. 515-522, 2014.

SEEJA, K. R.; ZAREAPPOR, M. FraudMiner: a novel credit card fraud detection model based on frequent itemset mining. **The Scientific World Journal**, Hindawi Publishing Corporation, 2014.

SOUZA, João Vítor Cepêda de. **Detecção de fraudes em telecomunicações com recurso a técnicas de Data Mining**. Faculdade de Engenharia, Universidade do Porto, Porto, Portugal, 16 fev. 2014. (Preparação da Dissertação)

SUN, Bo et al. Enhancing security using mobility-based anomaly detection in cellular mobile networks. **IEEE Trans. Veh. Technol**, v. 55, n. 4, p. 1385-1396, 2006.

STEINER, Maria Teresinha Arns et al. Extração de regras de classificação a partir de redes neurais para auxílio à tomada de decisão na concessão de crédito bancário. **Pesquisa Operacional**, v. 27, n. 3, p. 407-426, set./dez. 2007.

TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introduction to Data Mining.** [S.l.]: Pearson Education, 2014.

WEBSHOPPERS - E-BIT. 35. ed. 2017. Disponível em: <http://www.ebit.com.br/webshoppers>. Acesso em: abr. 2017.

WEI, Wei et al. Effective detection of sophisticated online banking fraud on extremely imbalanced data. **World Wide Web**, v. 16, n. 4, p. 449-475, 2013.

WEISS, Gary M. Mining with rarity: a unifying framework. **Sigkdd Explorations**, v. 6, n. 1, p. 7, 2004.

WELLS, Joseph T. **Corporate fraud handbook**: prevention and detection. 5. ed. [S.l.]: Wiley, 2017.

YANXIA, Lv et al. **Uncertain data stream classification with concept Drift.** In: INTERNATIONAL CONFERENCE ON ADVANCED CLOUD AND BIG DATA. 2016.

ZAREAPPOR, M.; SEEJA, R.; ALAM, A. M. Analyzing credit card: fraud detection techniques based on certain design criteria. **International Journal of Computer Application**, v. 52, n. 3, p. 35-42, 2012.

_____; SHAMSOLMOALI, P. **Application of credit card fraud detection: based on bagging ensemble classifier.** In: INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTING, COMMUNICATION & CONVERGENCE. Procedia Computer Science 48. p. 679-686, 2015.

APÊNDICE A - Classificador baseado em regras: o algoritmo RIPPER

Para o caso do problema de detecção de fraudes, de STEINER et al (2007):

Escolha uma delas como positiva e a outra como classe negativa:

1. Aprenda regras para a classe positiva;

2. A classe negativa será a classe padrão.

Construindo um conjunto de regras:

1. Use o algoritmo de cobertura sequencial:

    Encontre a melhor regra que cubra o conjunto atual de exemplos positivos; Elimine tanto os exemplos positivos quanto negativos cobertos pela regra.

2. Cada vez que uma regra é colocada no conjunto de regras, calcule o novo comprimento da descrição:

    Pare de adicionar novas regras quando o novo comprimento da descrição for d bits maior que o menor comprimento de descrição encontrado até então.

Crescendo uma regra:

1. Inicie com a regra vazia;
2. Adicione conjunções enquanto elas melhorarem o ganho de informação (*First-Order Induction Learning* ou Aprendizado por Indução de Primeira Ordem);
3. Pare quando a regra não cobrir mais exemplos negativos;

4. Pode a regra imediatamente usando o incremento da poda do erro reduzido

Medida para poda: $v = \frac{(p-n)}{(p+n)}$

Em que:

P = número de exemplos positivos cobertos pela regra no conjunto de validação; e

N = número de exemplos negativos cobertos pela regra no conjunto de validação.

Método de Poda: retire qualquer seqüência final de condições que maximize v

Otimizando o conjunto de regras:

1. Para cada regra r no conjunto de regras R:

Considere 2 regras alternativas:

Regra de substituição (r^*): cresça nova regra a partir do zero; Regra de revisão (r'): adicione conjunções para estender r.

Compare a regra r com as regras r^* e r'.

2. Escolha o conjunto de regras que minimize o *MDL* (*Minimum Description Length Principle* ou Princípio da Descrição de Mínimo Comprimento: representa o modelo da forma mais compacta possível com o máximo de informações dos dados):

Repita geração e otimização de regras para o restante dos exemplos positivos.

APÊNDICE B - Árvore de decisão: o algoritmo C4.5

O processo de indução de árvores de decisão tem a função de particionar recursivamente um conjunto de treinamento até que cada subconjunto obtido deste particionamento contenha casos de uma única classe.

Quando representada em forma de função, este processo recebe como parâmetros os exemplos e os atributos e procede da seguinte forma (BARBOSA et al. 2009):

1. Se chegar ao critério de parada, ou seja, se cada folha contiver casos de uma única classe ou não tiver como particionar mais porque os dois casos têm os mesmos valores para cada atirbuto, mas pertencem a classes diferentes

    1. Define-se a classe, ou seja, o algoritmo atribui ao nó terminal a classe mais provável dentro dos exemplos.

2. Caso contrário:

    1. Escolhe-se o atributo através de uma busca gulosa, selecionando a característica que maximiza a divisão dos dados por meio de entropia.

    2. Uma nova árvore é feita com, sendo que o melhor atributo escolhido torna-se o nó raiz da árvore;

    3. Faz-se a partição dessa nova árvore, ou seja, o algoritmo atribui a cada valor do atributo um ramo próprio. Este tipo de partição, embora permita extrair da característica todo o seu conteúdo informativo, tem como principal desvantagem a criação de um grande número de ramos muitas vezes completamente desnecessários, o que implica a formação de árvores de dimensões muitas vezes exageradas, porém, a poda, ao final do algoritmo resolve este problema).

    4. Enquanto houver partições:

        1. São escolhidos os melores exemplos da partição;

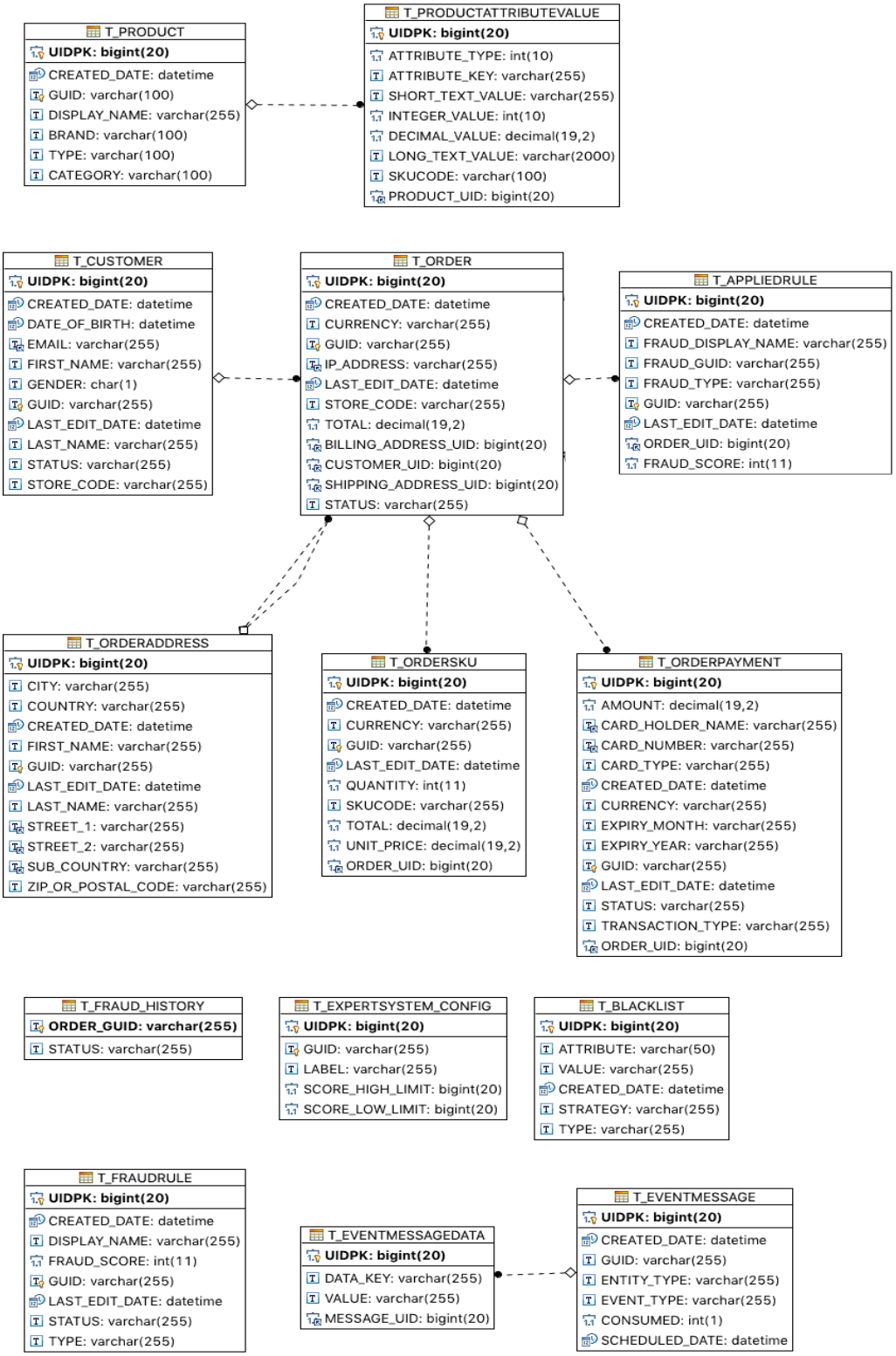
2. Inicia-se o processo de indução novamente, de forma recursive, passando como parâmetros os melhors exemploes e a nova árvore;

3. Adiciona-se o ramo à àrvore resultante do processo;

3. Faz-se a poda da árvore, ou seja, o algoritmo usa uma medida chamada taxa de erro ajustada. Esta medida incrementa cada taxa de classificação errada de cada nó para o conjunto de treinamento pela imposição de uma penalidade baseada no número de folhas da árvore. O objetivo é podar primeiro os ramos que possuem um menor poder preditivo por folha.

APÊNDICE C - Modelo Entidade Relacional do Sistema Antifraude

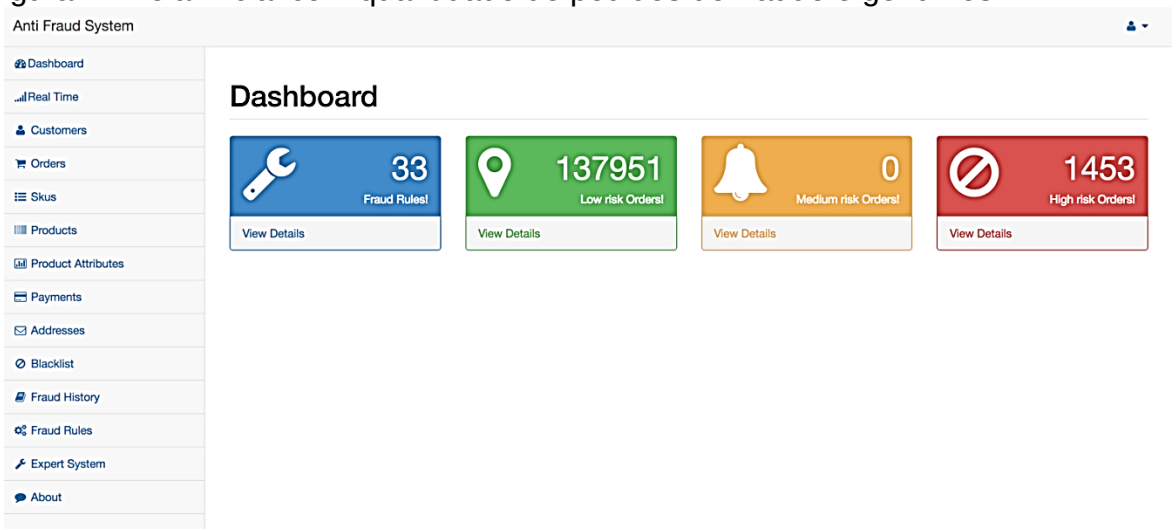
Figura 1 - Modelo Entidade Relacional do Sistema Antifraude proposto.



Fonte: Elaborado pelo autor.

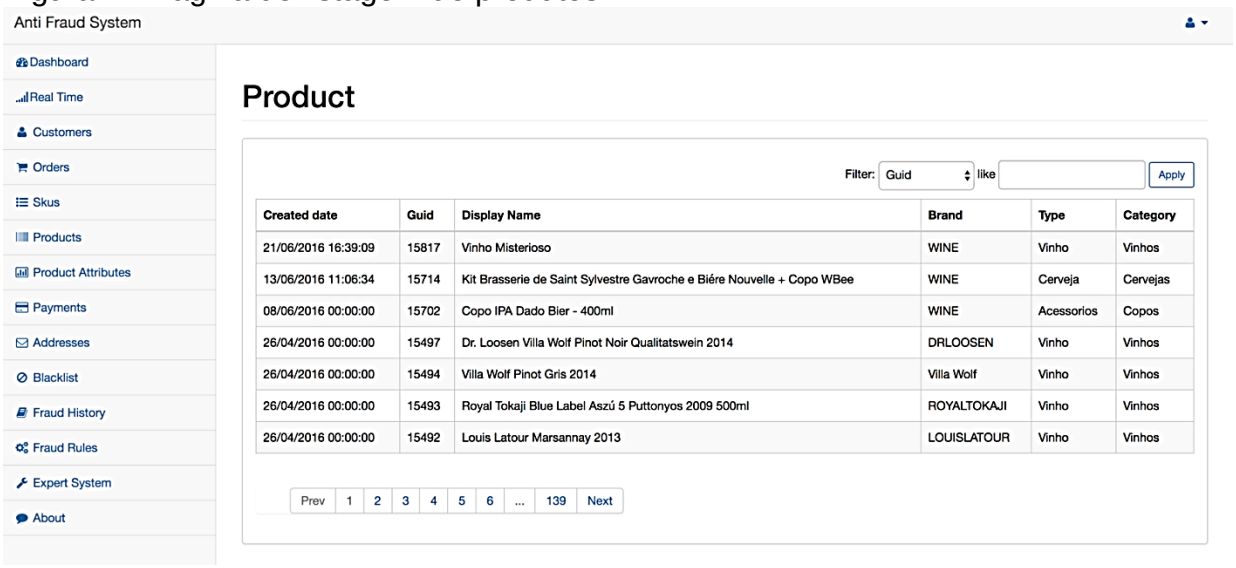
APÊNDICE D - Principais telas do Sistema Antifraude Desenvolvido

Figura 1 - Tela inicial com quantidade de pedidos de fraude e genuínos.



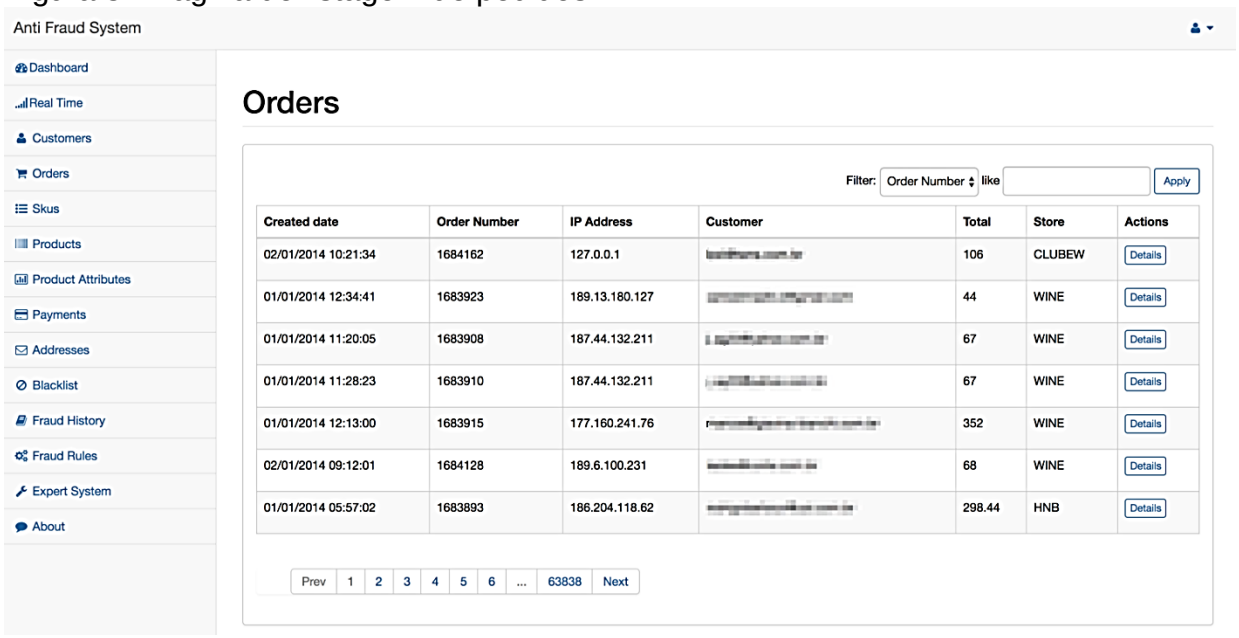
Fonte: Elaborado pelo autor.

Figura 2 - Página de listagem de produtos.



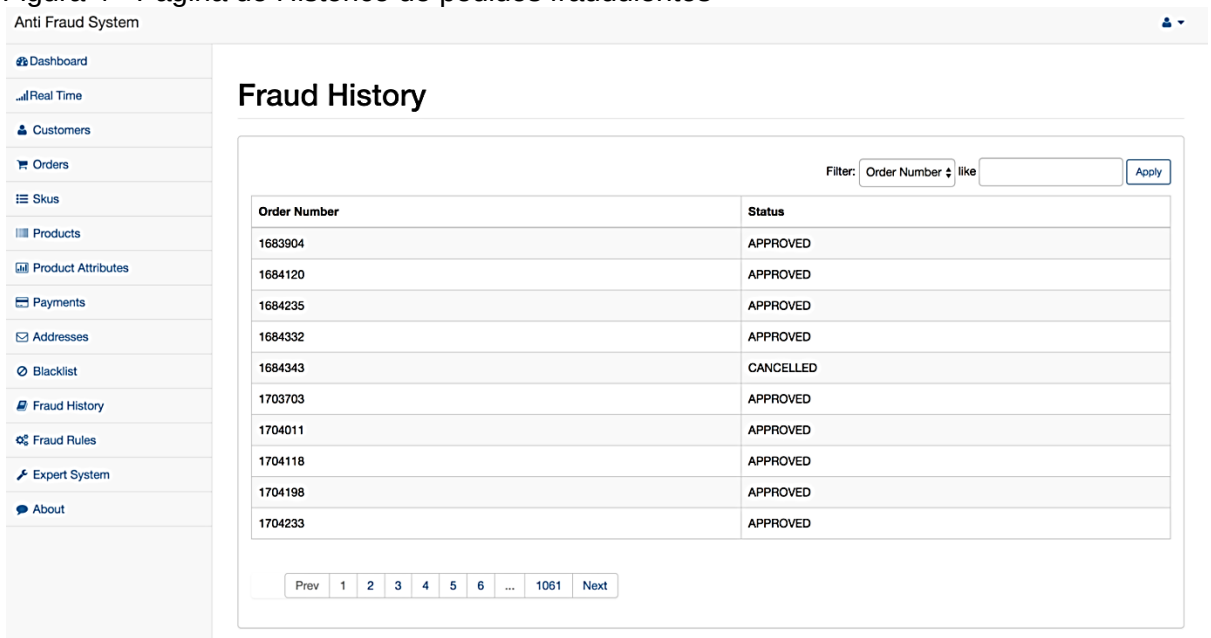
Fonte: Elaborado pelo autor.

Figura 3 - Página de listagem de pedidos.



Fonte: Elaborado pelo autor.

Figura 4 - Página do Histórico de pedidos fraudulentos



Fonte: Elaborado pelo autor.

APÊNDICE E - Exemplos de regras geradas pelo algoritmo RIPPER

Comportamento diferenciado por tipo de loja:

1. (**STORE_CODE = WINE**) and (CUSTOMER_PCT_CANCELLED_ORDER_24 >= 1.12) and (QTD_DISTINCT_CREDITCARD >= 6) and (CUSTOMER_PCT_COMPLETED_ORDER_24 <= 4.69) and (QTD_PEDIDOS_ENVIADOS_PARA_ENDERECO <= 26) and (QTD_CLIENTES_QUE_USARAM_MESMO_ENDERECO >= 2) and (ORDER_TOTAL >= 62) and (QTD_OUTROS_PRODUTOS >= 1) => CLASSE=FRAUDE (153.0/25.0)

2. (ORDER_HOUR_OF_DAY >= 11) and (CUSTOMER_PCT_COMPLETED_ORDER_24 >= 1) and (DIFF_DAYS_LAST_PURCHASE <= 123) and (QTD_ORDERS_24 >= 1) and (**STORE_CODE = HNB**) => CLASSE=FRAUDE (569.0/149.0)

Produtos diferentes com comportamentos de fraude:

1. (CUSTOMER_PCT_CANCELLED_ORDER_24 >= 1.09) and (DIFF_DAYS_LAST_PURCHASE <= 18) and (CUSTOMER_PCT_COMPLETED_ORDER_24 <= 4.69) and (CUSTOMER_AVG_TICKET >= 196.13) and (**QTD_OUTROS_PRODUTOS** >= 1) and (QTD_ORDERS_24 <= 7) => CLASSE=FRAUDE (721.0/67.0)

Informação sobre frequência de pedidos:

1. (**CUSTOMER_PCT_CANCELLED_ORDER_24 >= 1.09**) and (**QTD_ORDERS_24 >= 5**) and (**CUSTOMER_PCT_COMPLETED_ORDER_24 <= 5.58**) and (ORDER_HOUR_OF_DAY >= 17) and (ORDER_HOUR_OF_DAY <= 19) and (QTD_CLIENTES_QUE_USARAM_MESMO_ENDERECO >= 7) => CLASSE=FRAUDE (257.0/17.0)

Relações pouco intuitivas

1. (STORE_CODE = WINE) and (**DIFF_DAYS_LAST_PURCHASE <= 60**) and (CUSTOMER_PCT_COMPLETED_ORDER_24 >= 1) and (CUSTOMER_AVG_TICKET >= 459.09) and (**QTD_PEDIDOS_ENVIADOS_PARA_ENDERECO <= 2**) => CLASSE=FRAUDE (164.0/40.0)

APÊNDICE F - Exemplo da árvore gerada pelo algoritmo C4.5

Figura 1 - Exemplo da árvore gerada pelo algoritmo C4.5

```
STORE_CODE = WINE
|   CUSTOMER_PURCHASE_QTD <= 18
|   |   CUSTOMER_PCT_CANCELADOS <= 5.75
|   |   |   CUSTOMER_AVG_TICKET <= 214.05
|   |   |   |   CUSTOMER_AVG_TICKET <= 50.5
|   |   |   |   |   CUSTOMER_PURCHASE_QTD <= 1
|   |   |   |   |   |   CUSTOMER_PCT_CANCELADOS <= 0.5
|   |   |   |   |   |   |   CUSTOMER_AVG_TICKET <= 49.97
|   |   |   |   |   |   |   |   ORDER_DAY_OF_WEEK <= 2
|   |   |   |   |   |   |   |   |   ORDER_DAY_OF_WEEK <= 1
|   |   |   |   |   |   |   |   |   |   CUSTOMER_AVG_TICKET <= 39.55
|   |   |   |   |   |   |   |   |   |   |   ORDER_HOUR_OF_DAY <= 6: FRAUDE (5.0)
|   |   |   |   |   |   |   |   |   |   |   ORDER_HOUR_OF_DAY > 6: LEGITIMO (49.0/16.0)
|   |   |   |   |   |   |   |   |   |   CUSTOMER_AVG_TICKET > 39.55
|   |   |   |   |   |   |   |   |   |   |   ORDER_HOUR_OF_DAY <= 18: FRAUDE (16.0/1.0)
|   |   |   |   |   |   |   |   |   |   |   ORDER_HOUR_OF_DAY > 18
|   |   |   |   |   |   |   |   |   |   |   |   QTD_CLIENTES_QUE_USARAM_MESMO_ENDERECO <= 24: FRAUDE (7.0/2.0)
|   |   |   |   |   |   |   |   |   |   |   |   QTD_CLIENTES_QUE_USARAM_MESMO_ENDERECO > 24: LEGITIMO (2.0)
|   |   |   |   |   |   |   |   |   ORDER_DAY_OF_WEEK > 1: FRAUDE (180.0/83.0)
|   |   |   |   |   |   |   |   ORDER_DAY_OF_WEEK > 2
|   |   |   |   |   |   |   |   |   CUSTOMER_AVG_TICKET <= 32.4: LEGITIMO (519.0/178.0)
|   |   |   |   |   |   |   |   |   CUSTOMER_AVG_TICKET > 32.4
|   |   |   |   |   |   |   |   |   |   QTD_CLIENTES_QUE_USARAM_MESMO_ENDERECO <= 47
|   |   |   |   |   |   |   |   |   |   |   CUSTOMER_AVG_TICKET <= 46.2: FRAUDE (282.0/138.0)
|   |   |   |   |   |   |   |   |   |   |   CUSTOMER_AVG_TICKET > 46.2: LEGITIMO (86.0/29.0)
|   |   |   |   |   |   |   |   |   |   QTD_CLIENTES_QUE_USARAM_MESMO_ENDERECO > 47: FRAUDE (21.0/3.0)
|   |   |   |   |   |   |   |   CUSTOMER_AVG_TICKET > 49.97: FRAUDE (56.0/6.0)
|   |   |   |   |   |   |   CUSTOMER_PCT_CANCELADOS > 0.5: LEGITIMO (557.0)
|   |   |   |   |   |   CUSTOMER_PURCHASE_QTD > 1
|   |   |   |   |   |   |   CUSTOMER_AVG_TICKET <= 5.95
|   |   |   |   |   |   |   |   CUSTOMER_PCT_CANCELADOS <= 0.94: FRAUDE (46.0/17.0)
|   |   |   |   |   |   |   |   CUSTOMER_PCT_CANCELADOS > 0.94
|   |   |   |   |   |   |   |   |   QTD_CLIENTES_QUE_USARAM_MESMO_ENDERECO <= 3: LEGITIMO (167.0/10.0)
|   |   |   |   |   |   |   |   |   QTD_CLIENTES_QUE_USARAM_MESMO_ENDERECO > 3
|   |   |   |   |   |   |   |   |   |   ORDER_HOUR_OF_DAY <= 22
|   |   |   |   |   |   |   |   |   |   |   CUSTOMER_PURCHASE_QTD <= 13: LEGITIMO (43.0/2.0)
|   |   |   |   |   |   |   |   |   |   |   CUSTOMER_PURCHASE_QTD > 13
|   |   |   |   |   |   |   |   |   |   |   |   ORDER_DAY_OF_WEEK <= 5: FRAUDE (5.0)
|   |   |   |   |   |   |   |   |   |   |   |   ORDER_DAY_OF_WEEK > 5: LEGITIMO (5.0)
|   |   |   |   |   |   |   |   |   |   ORDER_HOUR_OF_DAY > 22: FRAUDE (5.0)
|   |   |   |   |   |   |   CUSTOMER_AVG_TICKET > 5.95
|   |   |   |   |   |   |   |   CUSTOMER_PCT_CANCELADOS <= 0.94
|   |   |   |   |   |   |   |   |   ORDER_TOTAL <= 51
|   |   |   |   |   |   |   |   |   |   CUSTOMER_AVG_TICKET <= 44
|   |   |   |   |   |   |   |   |   |   |   ORDER_TOTAL <= 45.01
|   |   |   |   |   |   |   |   |   |   |   |   CUSTOMER_AVG_TICKET <= 43.08
|   |   |   |   |   |   |   |   |   |   |   |   |   ORDER_TOTAL <= 39.75
|   |   |   |   |   |   |   |   |   |   |   |   |   |   CUSTOMER_AVG_TICKET <= 36.01
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   ORDER_HOUR_OF_DAY <= 10
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   ORDER_DAY_OF_WEEK <= 3: FRAUDE (185.0/21.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   ORDER_DAY_OF_WEEK > 3
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   QTD_PEDIDOS_ENVIADOS_PARA_ENDERECO <= 11
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   ORDER_TOTAL <= 18.75
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   CUSTOMER_AVG_TICKET <= 21.63
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   QTD_CLIENTES_QUE_USARAM_MESMO_ENDERECO <= 1
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   CUSTOMER_AVG_TICKET <= 14.38
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   CUSTOMER_PCT_CANCELADOS <= 0.18: FRAUDE (10.0/1.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   CUSTOMER_PCT_CANCELADOS > 0.18: LEGITIMO (31.0/11.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   CUSTOMER_AVG_TICKET > 14.38: FRAUDE (8.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   QTD_CLIENTES_QUE_USARAM_MESMO_ENDERECO > 1: FRAUDE (65.0/9.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   CUSTOMER_AVG_TICKET > 21.63
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   CUSTOMER_PCT_CANCELADOS <= 0.12
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   CUSTOMER_PURCHASE_QTD <= 2: LEGITIMO (4.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   CUSTOMER_PURCHASE_QTD > 2: FRAUDE (5.0/1.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   CUSTOMER_PCT_CANCELADOS > 0.12: LEGITIMO (7.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   ORDER_TOTAL > 18.75
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   ORDER_DAY_OF_WEEK <= 5: FRAUDE (137.0/19.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   ORDER_DAY_OF_WEEK > 5
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   ORDER_DAY_OF_WEEK <= 6
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   ORDER_HOUR_OF_DAY <= 8
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   CUSTOMER_AVG_TICKET <= 34.68: FRAUDE (35.0/2.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   CUSTOMER_AVG_TICKET > 34.68
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   ORDER_HOUR_OF_DAY <= 0: FRAUDE (6.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   ORDER_HOUR_OF_DAY > 0
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   CUSTOMER_PURCHASE_QTD <= 5: LEGITIMO (7.0/1.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   CUSTOMER_PURCHASE_QTD > 5: FRAUDE (3.0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   ORDER_HOUR_OF_DAY > 8
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   CUSTOMER_PURCHASE_QTD <= 2: FRAUDE (5.0/1.0)
```

Fonte: Elaborado pelo autor.