INSTITUTO FEDERAL DO ESPÍRITO SANTO
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA

FREDERICO LUIS DE AZEVEDO

DETECÇÃO DE FRAUDES DE CARTÃO DE CRÉDITO EM UMA BASE BRASILEIRA UTILIZANDO AUTOENCODER

Serra
2021

FREDERICO LUIS DE AZEVEDO

DETECÇÃO DE FRAUDES DE CARTÃO DE CRÉDITO EM UMA BASE BRASILEIRA UTILIZANDO AUTOENCODER

Dissertação apresentada ao Programa de Pós-Graduação em Computação Aplicada do Instituto Federal do Espírito Santo, Campus Serra, como requisito parcial para obtenção do título de Mestre em Computação Aplicada.

Orientador: Prof. Dr. Hilario Seibel Junior.

Serra

2021

Dados Internacionais de Catalogação na Publicação (CIP)

A994d 2021

Azevedo, Frederico Luis de
Detecção de fraudes de cartão de crédito em uma base brasileira utilizando Autoencoder / Frederico Luis de Azevedo. - 2021.
55 f.; il.; 30 cm

Orientador: Prof.Dr. Hilário Seibel Junior.
Dissertação (mestrado) - Instituto Federal do Espírito Santo, Programa de Pós-graduação em Computação Aplicada, 2021.

1. Redes neurais (Computação). 2. Fraude no cartão de crédito. 3. Aprendizado do computador. 4. Banco de dados. 5. Mineração de dados. I. Seibel Junior, Hilário. II. Instituto Federal do Espírito Santo. III. Título.

CDD 006.32

Bibliotecária Rogeria Gomes Belchior - CRB6/ES 417

FREDERICO LUIS DE AZEVEDO

DETECÇÃO DE FRAUDES DE CARTÃO DE CRÉDITO EM UMA BASE BRASILEIRA UTILIZANDO AUTOENCODER

Dissertação apresentada ao Programa de Pós-Graduação em Computação Aplicada do Instituto Federal do Espírito Santo, como requisito parcial para obtenção do título de Mestre em Computação Aplicada.

Aprovado em 20 de dezembro de 2021

COMISSÃO EXAMINADORA

Prof. Dr. Hilário Seibel Júnior
Instituto Federal do Espírito Santo
Campus Serra

Profª Dra. Kelly Assis de Souza Gazolli
Instituto Federal do Espírito Santo
Campus Serra

Profª Dra. Mariella Berger Andrade
Instituto Federal do Espírito Santo

Emitido em 20/12/2021

DECLARAÇÃO Nº 23/2021 - SER-CGEN (11.02.32.01.08.02)

(Nº do Protocolo: NÃO PROTOCOLADO)

(Assinado digitalmente em 08/02/2022 14:03)
HILARIO SEIBEL JUNIOR
PROFESSOR DO ENSINO BASICO TECNICO E TECNOLOGICO
SER-CGEN (11.02.32.01.08.02)
Matrícula: 1509954

(Assinado digitalmente em 08/02/2022 15:29)
KELLY ASSIS DE SOUZA GAZOLLI
PROFESSOR DO ENSINO BASICO TECNICO E TECNOLOGICO
SER-CGEN (11.02.32.01.08.02)
Matrícula: 1344568

(Assinado digitalmente em 09/02/2022 15:00)
MARIELLA BERGER ANDRADE
PROFESSOR DO ENSINO BASICO TECNICO E TECNOLOGICO
CEF-CGE (11.02.38.01.05)
Matrícula: 1509960

Para verificar a autenticidade deste documento entre em https://sipac.ifes.edu.br/documentos/ informando seu número: **23**, ano: **2021**, tipo: **DECLARAÇÃO**, data de emissão: **08/02/2022** e o código de verificação: **465cedf3ea**

AGRADECIMENTOS

Agradeço a Deus por ter chegado até aqui.

Agradeço a minha esposa e minha família pelo apoio e incentivo que tanto precisei para continuar a seguir em frente nos meus objetivos.

Agradeço ao meu orientador pela paciência e atenção dadas durante os meses de desenvolvimento desta dissertação.

Agradeço aos meus amigos que, de alguma forma, contribuíram para a realização deste trabalho.

RESUMO

O aumento no número de transações com cartões de crédito feitas pela *internet* nos últimos anos levou a um crescimento na quantidade de fraudes na mesma proporção. Devido ao grande volume de transações realizadas diariamente, é necessário contar com um sistema robusto para predição deste crime visando reduzir o prejuízo e aumentar a confiança de bancos e emissores de cartões. Técnicas de *Deep Learning* surgem como uma maneira de automatizar esse processo, treinando classificadores com dados de transações passadas para tentar prever fraudes futuras. Neste trabalho, foi construído um modelo *Autoencoder* e feito um ajuste de limiar para prever transações fraudulentas. Uma base de dados proprietária de transações de cartão de crédito Brasileira foi usada para treinamento e avaliação de desempenho do modelo, contendo quase 40 milhões de transações e fraudes desafiadoras, que não foram previamente detectadas e filtradas pelos sistemas de detecção de fraude já existentes da organização.

Palavras-chave: Aprendizado de máquina. Detecção de fraude em cartão de crédito. Aprendizagem profunda. Autoencoders.

ABSTRACT

The increasing number of credit-card transactions made over the internet in recent years has lead to a rise in the same proportion in the amount of fraud. Due to the large volume of web-based transactions that should be carried out daily, it is necessary to have a robust system to predict such crime to reduce loss and increase the confidence of banks and issuers. Deep Learning techniques emerge as a way to automate this process, training classifiers with data from past transactions to try to predict future frauds. In this paper, we build an Autoencoder model and perform a threshold tuning to predict fraudulent transactions. A proprietary Brazilian credit-card transaction database was used for training and performance evaluation of the model, containing almost 40 million transactions and challenging frauds, which were not previously detected by the organization's current fraud detection systems.

Keywords: Machine learning. Credit card fraud detection. Deep learning. Autoencoders.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS

AUC – Area Under the Curve

FN – False Negative

FP – False Positive

KNN – K-nearest Neighbor

LR – Logistic Regression

MCC – Matthews Correlation Coefficient

MLP – Multilayered Perceptron

MSE – Mean Squared Error

RBM – Restricted Boltzmann Machine

ROC – Receiver Operating Characteristic

RMSE – Root-Mean-Squared Error

SVM – Support Vector Machine

TN – True Negative

TP – True Positive

SUMÁRIO

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

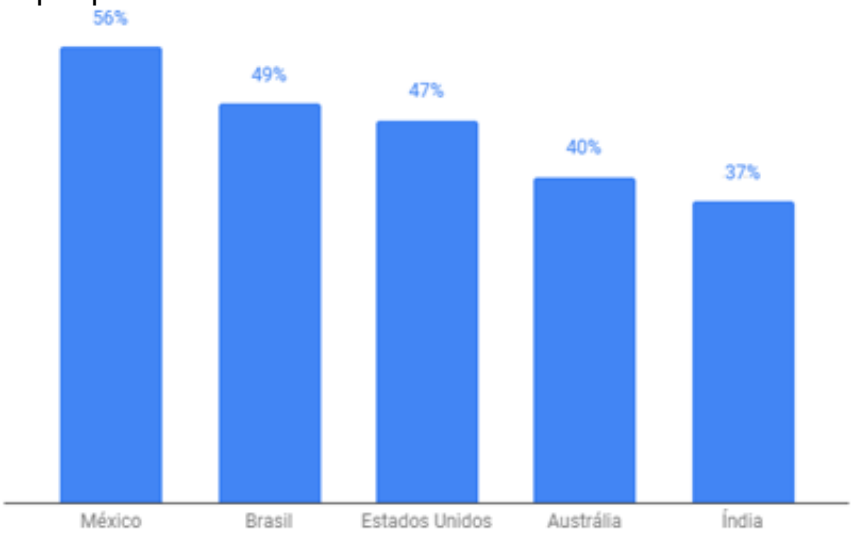
O constante crescimento de transações *online*, causado principalmente pela popularidade de *e-commerces* na última década, fez com que os casos de fraudes com cartões de crédito se tornassem cada vez mais comuns. Shakya (2018) descreve que a fraude é um tipo de crime intencional em que um fraudador se beneficia negando o direito a uma vítima ou obtendo ganhos financeiros. A definição de fraude, segundo Oxford (2021), é um ato criminoso de enganar alguém para obter dinheiro ou bens ilegalmente.

A fraude de cartão de crédito pode ser classificada em dois tipos: *Offline* e *Online*. A fraude *Offline* acontece quando um cartão é roubado e utilizado fisicamente para aquisição de algum bem, como, por exemplo, uma compra em uma loja. A fraude *Online* é feita via *internet*, telefone ou qualquer outra situação em que o dono do cartão não precisa estar presente fisicamente. Kou et al. (2004) citam que na modalidade *Online* são necessários apenas alguns dados do cartão, não havendo necessidade de digitação de senha ou assinatura.

O Brasil está em segundo lugar no *ranking* de países com mais fraudes em cartões de crédito, débito e pré-pagos de acordo com uma pesquisa divulgada por Knieff (2016). Os dados mostram que 49% dos entrevistados brasileiros sofreram algum tipo de fraude de cartão nos últimos 5 anos da publicação da pesquisa. O estudo também mostra que o Brasil é um ambiente favorável para ataques deste tipo, pois muitas empresas de comércio eletrônico não têm fortes controles para a prevenção de fraudes.

A Figura 1 apresenta o gráfico de cinco países e suas respectivas porcentagens de entrevistados que já sofreram algum tipo de fraude em cartão de crédito nos últimos cinco anos.

Figura 1 – Porcentagem de entrevistados que sofreram alguma fraude em cartão nos últimos 5 anos por país.



Fonte: Knieff (2016).

Quando bancos perdem dinheiro devido a fraude de cartão de crédito, os titulares dos cartões pagam inteiramente pela perda através de taxas de juros mais altas, taxas maiores de adesão e manutenção e benefícios reduzidos (CHAN et al., 1999). Por conta disso, implementar soluções eficazes de detecção de fraudes é de extrema importância para todas as organizações que emitem cartões de crédito ou gerenciam transações *online*, a fim de reduzir as perdas e, ao mesmo tempo, melhorar a confiança dos clientes (FIORE et al., 2017).

De acordo com Shakya (2018), o processo de detecção de fraude consiste em rotular se uma transação é legítima ou não. Sistemas automatizados para detecção de fraudes são necessários considerando o enorme volume de transações, não sendo possível para humanos verificar manualmente se cada transação é fraudulenta ou não, considerando que no ano de 2020, foram feitas 9,6 bilhões de transações com cartão de crédito no Brasil (Banco Central do Brasil, 2021).

Técnicas que envolvem aprendizado de máquina demonstram ser extremamente eficazes para enfrentar esse desafio. Duas delas em particular se destacam nesse contexto: as técnicas de aprendizado supervisionado e não supervisionado. O aprendizado supervisionado busca por padrões pré-definidos utilizando bases pré-rotuladas para encontrar novas fraudes. Já as técnicas de aprendizado não supervisionado não precisam de conhecimento prévio de transações fraudulentas e

não fraudulentas da base histórica. Pelo contrário, detectam mudanças no comportamento ou transações incomuns (KOU et al., 2004).

O maior desafio da detecção de fraude por aprendizado de máquina é o desbalanceamento de classes: a quantidade de transações legítimas é proporcionalmente mais representada quando comparada ao número de transações fraudulentas (CHAN et al., 1999). Quando os dados de treinamento de um sistema são desbalanceados desta forma, um algoritmo de aprendizado pode descartar a classe minoritária, tratando-os como ruído e classificando todos os registros como instâncias da classe majoritária (FIORE et al., 2017) ou seja, isso é equivalente a não detectar fraude alguma (CHAN et al., 1999).

Técnicas de detecção de anomalias (*Anomaly Detection*, em inglês) têm sido usadas para detecção de fraudes em bases desbalanceadas para não haver necessidade de manipular os dados de treinamento. A detecção de anomalias refere-se à tarefa de encontrar padrões nos dados que não estão de acordo com o comportamento esperado (CHANDOLA; BANERJEE; KUMAR, 2009).

Outra técnica que vem ganhando relevância no contexto de aprendizado de máquina é conhecido como Aprendizado Profundo (*Deep Learning*, em inglês). Técnicas de *Deep Learning* foram inicialmente utilizadas em reconhecimento automático de fala, reconhecimento de imagem e processamento de linguagem natural e mais recentemente na detecção de fraudes em transações financeiras como mostrado por Roy et al. (2018).

Um artigo publicado por Breslow et al. (2017) cita que as técnicas que usam *Deep Learning* precisam de grandes quantidades de dados e de modelos bem ajustados para a detecção de fraude. O artigo também destaca que *Deep Learning* é uma técnica utilizada em larga escala por bancos para combater lavagem de dinheiro, fraude e outros crimes financeiros.

1.2 FLUXO DE AUTORIZAÇÃO DE COMPRA

O momento adequado para a detecção de fraude de uma transação com cartão é durante a etapa de autorização da compra. O processo de autorização de cartão de

crédito envolve pelo menos cinco partes: o portador do cartão, o estabelecimento, a adquirente, a bandeira e o emissor (SANTIAGO, 2014), onde cada um é respectivamente responsável por uma etapa da autorização.

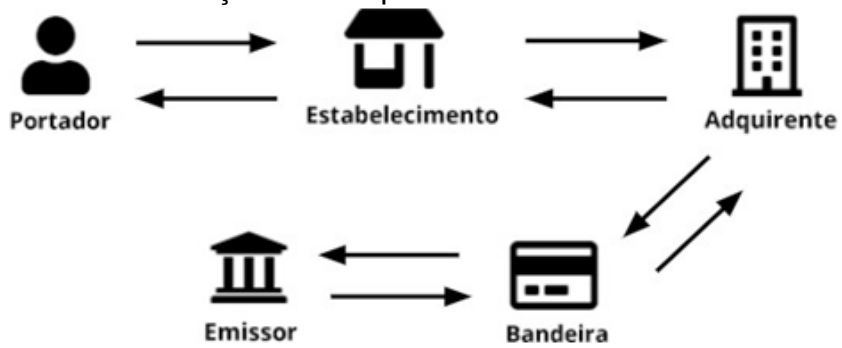
Quando o portador utiliza o seu cartão para fazer uma compra em um estabelecimento, que pode ser físico ou *online*, a transação é primeiramente encaminhada para o aceite da Adquirente, que na maioria dos casos são as maquininhas de cartão de crédito. Após o aceite da Adquirente, a transação é encaminhada para o sistema da Bandeira do cartão de crédito que também dá o aceite da compra. Após este aceite, a Bandeira então encaminha a transação ao Emissor do cartão de crédito, sendo esta a última etapa da autorização.

Por fim, o Emissor, que muitas vezes são bancos ou *fintechs*, valida os dados e dá o aceite final na autorização da compra. Todo o processo de aceite volta até o portador do cartão que tem a sua compra aprovada nesse fluxo.

A Figura 2 ilustra o fluxo de autorização com o relacionamento presente entre as cinco partes envolvidas no processo. O processo total de autorização precisa respeitar o tempo máximo de 15 segundos, ou então a transação é dada como negada. Logo, um sistema de detecção de fraude em tempo real precisa ser ter bom desempenho para não comprometer o processo de compra e gerar frustração ao portador do cartão.

Como o emissor é o responsável pelo aceite final na compra durante a última etapa no fluxo de autorização, não possuir um sistema de detecção de fraude nessa etapa faz com que ele fique reativo apenas a clientes que entram em contato comunicando compras não reconhecidas. Essas compras não reconhecidas podem ser fraudes ou não, necessitando de análise individual feita manualmente. Em outros casos, o emissor deposita sua confiança nos métodos de detecção de fraude da adquirente ou da bandeira, e que esses vão negar transações fraudulentas em etapas anteriores à dele no processo de autorização.

Figura 2 – Fluxo de autorização de compra de cartão.



Fonte: Elaborado pelo autor (2021).

Alguns emissores utilizam sistemas terceirizados para delegar a tarefa de detecção de fraude, porém estes sistemas normalmente tem acesso a poucos dados, pois informações financeiras são extremamente sensíveis e mantidas confidenciais para manter a privacidade dos clientes (MAES et al., 2002). Logo, estes sistemas não conseguem acesso a dados históricos de compra, nem a dados de perfil do cliente, e por isso, não consideram varáveis importantes para a tomada de decisão.

O desenvolvimento de um sistema de detecção de fraude pelo próprio emissor oferece mais confiabilidade ao se adequar às necessidades da empresa e ao perfil de compra de seus clientes, podendo ainda ser treinado com frequência e ser flexível para operar com diferentes sistemas de bandeiras e adquirentes.

1.3 PROPOSTA DO TRABALHO

A hipótese deste trabalho é de ser possível utilizar aprendizado de máquina com um modelo de *Deep Learning* para classificar transações fraudulentas em cartões de crédito no momento da autorização pelo emissor, mesmo após aprovação da adquirente e da bandeira do cartão.

Logo, o objetivo deste trabalho é realizar uma análise de um conjunto de transações de cartões de crédito de uma *fintech* Brasileira utilizando aprendizado de máquina com *Deep Learning* para obter performance e eficácia na detecção de fraudes.

O escopo deste trabalho se limita a utilizar uma base de dados Brasileira de transações de cartões de crédito disponibilizado por uma empresa específica do

ramo de meios de pagamento, e portanto, seu aprendizado está condicionado às características de fraudes de compradores nacionais e que se encaixam em seu público-alvo.

1.4 OBJETIVOS

1.4.1 Objetivo Geral

O objetivo principal deste trabalho é detectar fraudes de transações de cartão de crédito em uma base de dados fornecida por uma *fintech* Brasileira utilizando uma rede neural do tipo *Autoencoder*, criada utilizando algoritmos de *Deep Learning*.

1.4.2 Objetivos Específicos

1. Treinar a rede neural com registros da base rotulados como "transações legítimas" para que o modelo encontre padrões nos dados;

2. Utilizando a base de testes, determinar um limiar de separação das transações em classes investigando a taxa de erro do resultado da rede neural;

3. Classificar os registros de saída da rede neural como "legítimas" ou "fraudes" baseado no limiar definido anteriormente;

4. Gerar uma matriz de confusão com o resultado da classificação para estudar o desempenho do modelo na tarefa de detecção de fraudes.

1.5 ORGANIZAÇÃO DO TRABALHO

Esta dissertação está dividida em cinco capítulos além desta introdução.

O capítulo 2 apresenta um referencial teórico sobre os temas e conceitos que serão abordados ao longo do trabalho. Nesse capítulo também serão apresentados trabalhos correlatos ao tema de detecção de fraudes.

Posteriormente apresenta-se o capítulo 3, detalhando a base de dados, descrevendo seus atributos e explorando seu conteúdo. Esse capítulo também traz detalhes do modelo *Autoencoder* construído e as métricas de avaliação utilizadas para verificar seu desempenho.

No capítulo 4 são apresentados os experimentos realizados com o modelo de rede neural e discutidos os resultados obtidos.

O capítulo 5 traz as reflexões sobre o alcance do objetivo proposto e sugestões para pesquisas futuras.

2 REFERENCIAL TEÓRICO

Neste capítulo serão apresentados alguns conceitos e técnicas utilizadas ao longo do trabalho, além de trabalhos correlatos que estudam o tema de detecção de fraudes utilizando modelos de aprendizado de máquina.

2.1 BASES DESBALANCEADAS

As bases de dados desbalanceadas são aquelas em que o subconjunto de uma das classes aparece com uma frequência maior que os dados das demais classes (FACELI et al., 2011). De acordo com Shakya (2018), a maioria dos desafios do mundo real possui distribuição de classes desbalanceadas, e isso se mostra presente também no contexto da detecção de fraude,s onde o número classes rotuladas como fraude é muito baixo em comparação com o número de classes consideradas legítimas.

Segundo Faceli et al. (2011), vários algoritmos têm o desempenho prejudicado na presença de dados desbalanceados, pois eles tendem a favorecer a classificação de dados da classe majoritária. Por conta disso, é importante definir quais algoritmos de aprendizado de máquina podem ser aplicados na presença de dados desbalanceados bem como quais métricas de avaliação devem ser usadas para calcular o desempenho de um modelo preditivo (SHAKYA, 2018).

Existem diversas técnicas na literatura para lidar com o desbalanceamento de classes de uma base de dados. Uma delas é a geração de um novo conjunto através da re-amostragem de dados para que as classes se tornem igualmente representadas (BHATTACHARYYA et al., 2011). Isso é feito para reduzir os efeitos de favorecimento da classe majoritária durante a fase de treinamento do modelo, e esta re-amostragem pode ser feita através de subamostragem e sobreamostragem (SILVA, 2020).

Na subamostragem, a base de dados original é particionada em subconjuntos que possuem uma distribuição de classes mais equilibrada (CHAN et al., 1999). O trabalho de Bhattacharyya et al. (2011) faz uso de uma estratégia de subamostragem aleatória, onde registros da classe majoritária são aleatoriamente

removidos até que os subconjuntos tenham distribuições mais balanceadas entre as duas classes. Esses subconjuntos são então usados para avaliação do modelo de seu trabalho.

Na sobreamostragem, a base de dados se torna balanceada por meio de um aumento da quantidade de amostras da classe minoritária. Essa técnica consiste em selecionar membros da classe minoritária, criar novas instâncias artificiais desses dados através da duplicação, por exemplo, e os adicionar ao novo conjunto de treinamento (SILVA, 2020), como utilizado no trabalho de Fiore et al. (2017).

Uma desvantagem da subamostragem é que há perda de dados durante o processo, e não há como prever o quanto que esses dados perdidos poderiam ter sido importantes para o treinamento do modelo. Já na sobreamostragem não há perda de dados porém esta técnica adiciona um grande custo computacional ao sistema ao criar novos registros (SILVA, 2020).

Neste trabalho foi utilizada a base de dados em sua forma original sem interferências de amostragem com intuito de representar melhor o desafio de classificação de fraudes, pois as bases reais tem natureza desbalanceada.

2.2 APRENDIZADO DE MÁQUINA

De acordo com Shakya (2018), o Aprendizado de Máquina (ou *Machine Learning*, em inglês) é definido como um campo da inteligência artificial que fornece ao sistema a capacidade de aprender automaticamente a partir de uma experiência, sem a intervenção humana, e tem como objetivo prever os resultados futuros da forma mais precisa possível, utilizando vários modelos algorítmicos. O aprendizado de máquina é muito diferente das abordagens convencionais de computação, em que os sistemas são programados explicitamente para calcular ou resolver um problema (SHAKYA, 2018).

Para isso, esse sistemas empregam um princípio de inferência chamado indução, no qual se obtêm conclusões genéricas a partir de um conjunto particular de exemplos. Assim, algoritmos de aprendizado de máquina aprendem a induzir uma função ou

hipótese capaz de resolver um problema a partir de dados que representam instâncias do problema a ser resolvido (FACELI et al., 2011).

Sistemas que utilizam aprendizado de máquina são usados de forma bem-sucedida em diversas áreas para a solução de problemas reais, como reconhecimento de fala, detecção de mensagens de *e-mail* indesejadas, condução de automóveis autônomos e diagnóstico de doenças (GRUS, 2019) (FACELI et al., 2011).

As técnicas de aprendizado de máquina podem ser categorizadas em dois tipos: sistemas de aprendizado supervisionado e não supervisionado.

2.2.1 Aprendizado supervisionado

No aprendizado supervisionado, os rótulos de entrada e saída são fornecidos ao modelo a ser treinado, que usa esses dados na fase de treinamento e extrai os padrões dos dados de entrada (SHAKYA, 2018).

No contexto de detecção de fraudes, Santiago (2014) descreve que os métodos supervisionados examinam transações previamente classificadas a fim de determinar futuras transações fraudulentas. Logo, os métodos supervisionados são treinados apenas para discriminar entre transações legítimas e fraudes com base em transações já conhecidas anteriormente, não sendo possível detectar novos padrões ou técnicas diferentes das que já estejam no conjunto de dados (KOU et al., 2004).

Existem algumas dificuldades no treinamento de modelos de aprendizado supervisionado no que diz respeito à obtenção de dados pré-rotulados. Uma delas é que a obtenção dos dados pode ser muito custosa, pois normalmente dados de transações bancárias e de crédito são protegidos por serem considerados dados sensíveis. Uma outra dificuldade é de que a rotulação dos dados pode estar incorreta, e por conta disso, levar o modelo a obter resultados errados (SANTIAGO, 2014).

2.2.2 Aprendizado não supervisionado

O aprendizado é categorizado como não supervisionado quando os modelos são treinados e avaliados com dados sem rótulos de saída pré-definidos e, portanto, o classificador aprende diferentes padrões e estruturas para gerar o resultado (SHAKYA, 2018). O aprendizado não supervisionado também é conhecido como uma tarefa de aprendizado descritiva, onde a meta é explorar ou descrever um conjunto de dados já que os modelos de aprendizado de máquina nesta categoria não fazem uso do atributo de saída (FACELI et al., 2011).

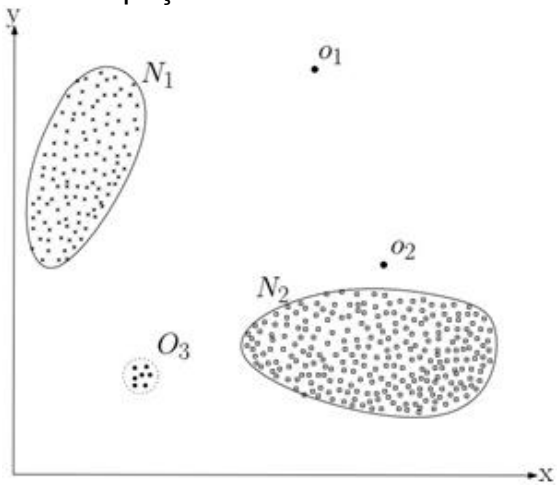
Para detecção de fraudes, o uso de métodos não supervisionados permite que tipos de fraudes anteriormente não descobertos possam ser detectados. A vantagem neste caso é que os modelos não precisam do conhecimento dos rótulos de classe para serem treinados, e a decisão sobre a identificação de uma transação como fraude é tomada com base nos padrões e características dela. O modelo é treinado com as transações legítimas de modo que ele possa diferenciar entre legítima e fraude para as próximas transações (REZAPOUR, 2019).

2.3 DETECÇÃO DE ANOMALIAS

Detecção de anomalias é a prática de identificar itens ou eventos que não estão de acordo com um comportamento esperado ou não se correlacionam com outros itens em um conjunto de dados (ZHANG; GARDNER; VUKOTIC, 2019). Chandola, Banerjee e Kumar (2009) definem anomalia como padrões de dados que não estão em conformidade com uma definição de comportamento normal. A Figura 3 ilustra anomalias detectadas em um espaço de duas dimensões. O conjunto de dados tem duas regiões normais N_1 e N_2 que concentram a maioria das observações. Os pontos que são distantes dessas regiões como o_1, o_2 e a região O_3 são considerados anomalias.

A detecção de anomalias é uma técnica de aprendizado e não supervisionado, utilizado de forma extensiva em uma ampla variedade de aplicações, como detecção de fraudes em seguro e planos de saúde, detecção de intrusão para segurança cibernética, detecção de falhas em sistemas críticos de segurança e vigilância militar para atividades inimigas (CHANDOLA; BANERJEE; KUMAR, 2009).

Figura 3 – Anomalias em um espaço de duas dimensões.



Fonte: Chandola, Banerjee e Kumar (2009)

Patcha e Park (2007) citam que uma das vantagens de um sistema de detecção de anomalias é de que o atacante ou fraudador não tem conhecimento das regras, e portanto não sabe quais atividades ele pode realizar sem gerar suspeitas ou alarmar o sistema. Uma outra vantagem desse tipo de sistema é de ser capaz de detectar anomalias previamente desconhecidas, sem a necessidade de ser treinado novamente.

De acordo com Chandola, Banerjee e Kumar (2009), técnicas de detecção de anomalias são recomendadas para analisar bases desbalanceadas de maneira não supervisionada, pois a premissa nesse caso é de que as instâncias de classes classificadas como normais são bem mais frequentes do que que as anomalias.

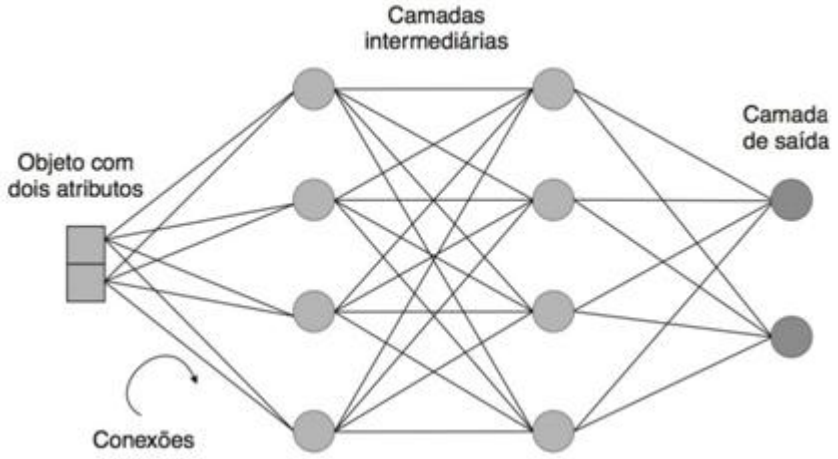
A análise de transações de cartões de crédito pode ser classificada como um problema de detecção de anomalias onde um dado semelhante aos dados vistos anteriormente é considerado como uma transação legítima. Dados que não são se assemelham ou que não podem ser categorizados com a maioria dos dados já vistos são, portanto, classificados como fraudes.

2.4 REDES NEURAIS ARTIFICIAIS

Redes Neurais Artificiais são modelos preditivos que tem como inspiração a estrutura e funcionamento da operação de um cérebro humano (GRUS, 2019). Uma característica desses modelos é conter unidades chamadas de neurônios, que são

responsáveis por computar funções matemáticas. Os neurônios são dispostos em uma ou mais camadas e interligados por um grande número de conexões geralmente unidirecionais. As conexões possuem pesos que ponderam a entrada recebida por cada neurônio da rede e podem assumir valores positivos ou negativos (FACELI et al., 2011). A Figura 4 ilustra um exemplo de Rede Neural. As redes neurais são responsáveis por resolver uma grande variedade de problemas, como reconhecimento de escrita à mão e detecção de rosto (GRUS, 2019).

Figura 4 – Rede neural.



Fonte: Faceli et al. (2011)

2.4.1 Redes Neurais Profundas

Algoritmos de Redes Neurais Profundas (*Deep Learning*, em inglês) são uma classe de algoritmos de aprendizado de máquina que usam várias camadas de processamento não lineares para extração e transformação de características. As características descobertas em uma camada formam a base para o processamento da camada seguinte (ROY et al., 2018).

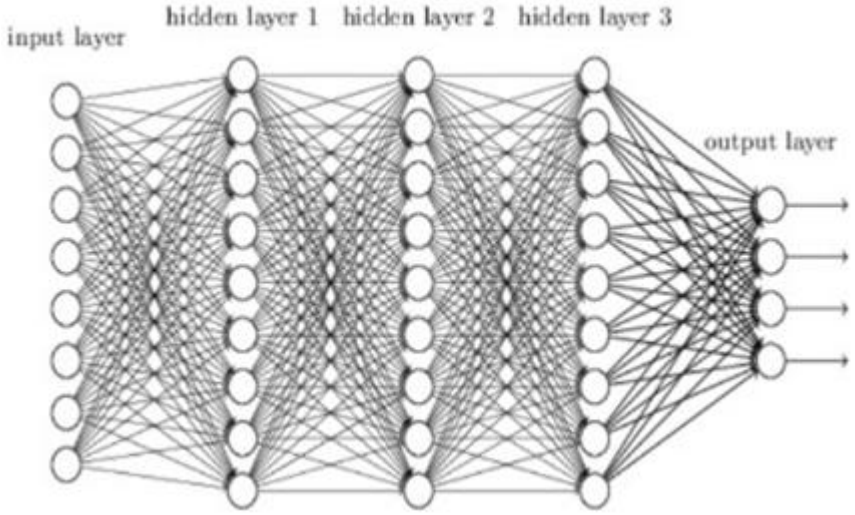
As instâncias da base de dados de treinamento desse tipo de rede neural passam por uma estrutura hierárquica que define pesos para determinadas características conforme são analisados. Se o modelo prevê uma observação incorreta, essa observação volta pela estrutura e os pesos são recalculados. Ao final deste processo, o modelo otimiza a função e escolhe as melhores características do conjunto de treinamento (RUSHIN et al., 2017).

Assim, esses algoritmos aprendem múltiplas camadas de características ou de representações dos dados. As características de nível superior são derivadas de características de nível inferior para formar uma representação hierárquica. (PANDEY, 2017).

Algoritmos de *Deep Learning* podem ser tanto de aprendizado supervisionado quanto de aprendizado não supervisionado, e suas aplicações incluem análise de padrões (não supervisionados) e classificação (supervisionados) (PANDEY, 2017). Pumsirirat e Yan (2018) citam que ao selecionar um algoritmo de aprendizagem profunda para resolver um determinado problema, o pesquisador deve saber o problema real que está sendo estudado e como cada algoritmo funciona.

A Figura 5 apresenta uma rede neural profunda contendo uma camada de entrada (*input layer*), três camadas intermediárias (*hidden layers*) e uma camada de saída (*ouput layer*).

Figura 5 – Rede neural profunda.



Fonte: Pumsirirat e Yan (2018).

2.4.2 Autoencoders

Um *Autoencoder* é uma rede neural profunda que recebe vários atributos de dados como entrada e em seguida codifica estes atributos através de um processo de redução de dimensionalidade. Depois de concluído, os dados compilados são

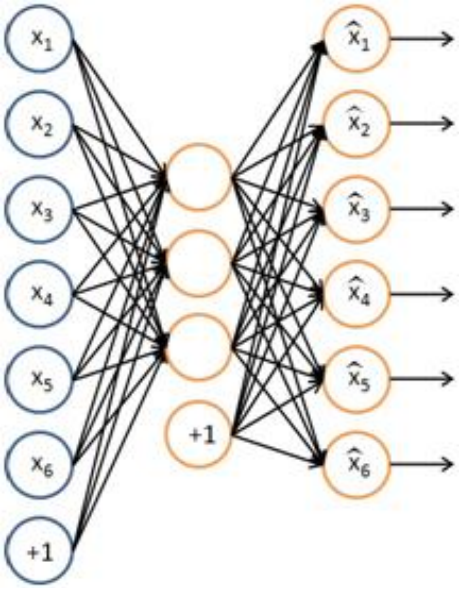
reconstruídos visando obter a entrada original (RUSHIN et al., 2017). Assim, o modelo *Autoencoder* não tem como objetivo fazer uma classificação como resultado, mas sim, alcançar novamente uma reconstrução fiel dos dados que foram fornecidos como entrada.

A Figura 6 ilustra um *Autoencoder* com uma camada de entrada contendo dados que vão de X_1 a X_6, uma camada intermediária que busca extrair as características principais dos dados e uma camada de saída com os dados reconstruídos.

O procedimento de converter dados de entrada em um espaço de baixa dimensão é denominado codificação e a operação reversa que reconstrói os dados originais é chamado de decodificação (KAZEMI; ZARRABI, 2017). Através da codificação e decodificação dos dados, o modelo consegue mapear as principais características da base de dados.

Desta forma, após o treinamento da rede neural, qualquer dado de entrada que passe pelo processo de codificação e não tenha características semelhante com as já conhecidas pelo *Autoencoder* é então tratado como uma anomalia. Uma transação fraudulenta ao ser submetida a um *Autoencoder*, por exemplo, apresentaria vários erros de reconstrução durante a fase de decodificação por ter características diferentes de uma transação legítima.

Figura 6 – Representação de um *Autoencoder*.



Fonte: Kazemi e Zarrabi (2017).

2.5 TRABALHOS CORRELATOS

Os trabalhos correlatos descritos abaixo usam modelos de *Autoencoder* para a classificação de transações fraudulentas de cartão de crédito.

No trabalho de Pumsirirat e Yan (2018) foram utilizadas três bases diferentes de transações de cartões de crédito (uma Alemã, uma Australiana e outra Europeia) para validar dois métodos de *Deep Learning*: *Autoencoders* e *Restricted Boltzmann Machines* (RBMs). Ambos os métodos são de aprendizado não supervisionado para detecção de anomalias. Cada modelo foi avaliado individualmente com cada uma das bases de dados. As métricas utilizadas por eles foram o Erro quadrático médio (*Mean Squared Error* ou MSE, em inglês), a Raiz quadrada do erro-médio (*Root-Mean-Squared Error* ou RMSE, em inglês) e a Curva Característica de Operação do Receptor (*Receiver Operating Characteristic Curve* ou *ROC curve*, em inglês).

O autor concluiu que o modelo *Autoencoder* tem uma melhor performance do que o *Restricted Boltzmann Machine*, principalmente em bases de dados maiores, como é o caso da base Europeia. As bases Alemã e Australiana não atingiram bons resultados por serem bases muito pequenas e portanto não apropriadas para modelos de *Deep Learning*.

O objetivo do artigo de Rezapour (2019) foi estudar o comportamento da aplicação de três métodos de aprendizado não supervisionado para detectar fraudes de cartão de crédito. Os métodos aplicados foram *One-Class SVM*, *Autoencoder* e *Multivariate Outlier Detection*, este utilizando a distância de *Mahalanobis* como medida de classificação. O autor indica que a base de dados é desbalanceada, pois as fraudes correspondem a 0,17% do total de registros e, por conta disso, uma subamostragem aleatória foi aplicada para balancear a base de dados.

O estudo conclui que o modelo *Autoencoder* foi o que obteve o maior sucesso porque apresentou o menor número de fraudes classificadas incorretamente como falsos positivos e falsos negativos. O estudo evitou avaliar os métodos com outras métricas de desempenho, pois ele afirma que cada um dos três modelos foi treinado de forma diferente e, portanto, não podem ser comparados entre si.

O objetivo de Al-Shabi (2019) foi construir um modelo *Autoencoder* para detectar fraudes de cartão de crédito em uma base de dados Europeia, contendo 284.807 registros com 492 transações fraudulentas. A base é utilizada em sua forma original sem aplicação de nenhuma técnica de amostragem, sendo dividido em 80% para treinamento e 20% para testes do modelo. No final do experimento, os autores definem um limiar para a taxa de erro de reconstrução que definirá uma transação como fraudulenta.

O *Autoencoder* é avaliado com quatro valores diferentes de limiar, e os autores concluem que o valor ideal deve ser um equilíbrio entre a detecção de mais fraudes verdadeiras (maior exatidão) e, ao mesmo tempo, manter um valor aceitável de casos de falsos positivos. Os melhores resultados são para o limiar 5: Exatidão (0,98), Precisão (0,011), Sensibilidade (0,64) e *F1-Score* (0,19).

O estudo de Misra et al. (2020) apresenta uma forma de detecção de fraude em dois estágios usando um modelo não supervisionado de *Autoencoder* combinado com um modelo de classificação supervisionado. A base de dados é a mesma do trabalho de Al-Shabi (2019), contendo 0,17% de transações rotuladas como fraudes. O *Autoencoder* é aplicado como um primeiro estágio para detectar e extrair as principais características do conjunto de dados, gerando uma nova base de dados com menos características, porém mais significativas.

Essa base de dados reduzida é então submetida à segunda etapa, que é o processo de classificação com três diferentes modelos previamente treinados: *Multilayered Perceptron* (MLP), *K-nearest Neighbor* (KNN) e Regressão Logística (LR). As métricas de desempenho utilizadas no estudo são Exatidão, Precisão, Sensibilidade e *F1-Score*, sendo esta última a mais importante de acordo com os autores. A conclusão do estudo é que a aplicação do *Autoencoder* combinado com o classificador MLP obteve os melhores valores de *F1-Score* (0,8265), Exatidão (0,9994) e Precisão (0,8534) quando comparados aos demais modelos.

2.6 CONSIDERAÇÕES SOBRE TRABALHOS CORRELATOS

A metodologia do trabalho correlato de Al-Shabi (2019) é bastante semelhante à apresentada neste trabalho, por fazer uma análise de limiar de taxa de erro para

separação de classes. Porém a base de dados do trabalho correlato de Al-Shabi (2019) contém transações de cartão de crédito coletadas por apenas dois dias de portadores de cartão da Europa. Além disso, os registros dessa base passaram por um processo prévio de transformação e foram anonimizados, sendo assim, suas características e formatos são desconhecidos. Também não há informações do processo de rotulação prévio de classes, se foi feito de forma manual ou automática.

Já este trabalho apresenta um modelo de rede neural para classificação de fraudes em uma base de dados de cartão de crédito brasileira. A base de dados proprietária deste trabalho possui um tamanho 139 vezes maior do que a base de dados Europeia, além de ter mais de 10 vezes o número de registros de fraudes. Por conta disso, essa base brasileira é mais desbalanceada do que a Europeia e por isso possui maior complexidade no processo de validação de performance do modelo.

Também é de conhecimento de que os registros de fraude dessa base foram rotulados manualmente, conforme será explicado no capítulo seguinte. Ao contrário da base Europeia, as características dessa base brasileira são conhecidas e permitem uma análise exploratória dos dados e o que cada um deles representa. Por conta dessa base possuir tamanho, características e perfis de compra diferentes, ela é mais adaptada aos desafios do país.

3 MATERIAIS E MÉTODOS

Neste capítulo serão apresentados e descritos os materiais e métodos utilizados neste trabalho. A base de dados utilizada no trabalho também será apresentada, bem como as informações de sua origem, a descrição de seus atributos e a exploração dos seus dados. Este capítulo também descreve o modelo desenvolvido para a análise de fraude.

3.1 BASE DE DADOS

Este trabalho utiliza uma base de dados própria com transações de cartões de crédito disponibilizada por uma *fintech* Brasileira do ramo de meios de pagamento. Essa base contém transações de cartões de crédito reais feitas por clientes da empresa. As transações foram coletadas durante todo o ano de 2019, do dia 01 de Janeiro ao dia 31 de Dezembro do mesmo ano. A base contém 39.571.558 transações de cartões de crédito no total, sendo que 39.550.254 são transações legítimas e 21.304 são consideradas fraudes

A quantidade de fraudes representa apenas 0,054% do total de transações conforme apresentado pela Tabela 1. Como o total de transações de fraude representa menos de 1% da base, trata-se de uma base desbalanceada. Esta situação de desbalanceamento é a mesma encontrada na base de dados Europeia utilizada nos estudos dos trabalhos correlatos de Pumsirirat e Yan (2018), Al-Shabi (2019) e Misra et al. (2020).

Tabela 1 – Distribuição de classes da Base de Dados

Classe	**Quantidade**	**Porcentagem**
Legítimas	39.550.254	99,946%
Fraudes	21.304	0,054%
TOTAL	39.571.558	100%

Fonte: Elaborado pelo Autor (2021).

3.1.1 Rotulação dos Dados

As transações desta base de dados foram rotuladas como fraudes por meio do contato dos clientes relatando compras não reconhecidas por eles. A transação só é então confirmada e rotulada como fraude após tal contato e após uma análise manual pela área responsável da *fintech*. A área da *fintech* analisa cada caso individualmente para confirmar se houve mesmo uma fraude na compra. Para isso, os analistas de fraude utilizam diversos dados da conta do cliente e de seu comportamento em transações anteriores, mas não é um processo determinístico por conta de cada transação e cada cliente terem características diferentes. Portanto, um desafio deste trabalho é de que podem haver transações legítimas que são fraudes mas não foram rotuladas como tal, porque não foram relatadas corretamente pelos clientes.

As transações também são analisadas previamente por meio de sistemas terceirizados de detecção de fraudes existentes na adquirente e na bandeira do cartão de crédito, porém o resultado dessa classificação não chega até o emissor pois são etapas de autorização anteriores a dele no fluxo de autorização de compra. Por conta disso, as fraudes detectadas por estes sistemas da adquirente ou da bandeira ficam restritas ao domínio deles e não aparecem como registros da base de dados utilizada neste trabalho.

O sistema existente na adquirente analisa regras simples, como verificação do horário da transação (se feitas pela madrugada, por exemplo, são mais suspeitas de serem fraudes), se são feitas presencialmente ou *online*, o país onde ocorreu a compra e se a digitação da senha foi necessária. Por motivo de privacidade, os dados do cliente não são analisados pela adquirente, bem como os dados de transações passadas. As regras do sistema anti-fraude da adquirente são fixas e aplicadas a qualquer compra. Esse sistema depende de uma pessoa responsável que observa o conjunto de regras e as ajusta de tempos em tempos.

Já o sistema utilizado pela bandeira do cartão de crédito, caso exista, é uma caixa preta e suas regras não são divulgadas por motivos de segurança. Dessa forma, o sistema de detecção de fraudes desenvolvido neste trabalho lida apenas com os casos de fraudes mais complexas que chegaram até a etapa de autorização do

emissor, ou seja, que não foram detectadas previamente pelos sistemas de fraude da adquirente e da bandeira.

3.1.2 Dicionário de Dados

Os atributos coletados para formar a base de dados estão apresentados na Tabela 2. A base é composta apenas por atributos numéricos, 12 deles no total. O valor da compra está em reais, ou foi previamente convertido em reais pelo próprio processo de autorização da compra pela *fintech* caso seja de moeda estrangeira. O *Ticket* Médio representa a média de gastos de compras anteriores do cliente no momento da transação, ou seja, a soma do valor total de todas as compras anteriores do cliente dividido pela quantidade de compras.

O atributo Hora é uma representação numérica com valores entre 0-23 do horário em que a transação foi realizada. O dia da semana é um valor numérico entre 0-6 onde 0 e 6 representam Domingo e Sábado, respectivamente. A moeda da compra está representada na base de dados na forma de um código numérico de 3 dígitos definido conforme o padrão internacional ISO 4217.

Dois marcadores (também chamados de *flags*, em inglês) são utilizados com valores de 0 ou 1 para saber de forma binária se: a transação foi feita de forma presencial ou *online* e se é a primeira compra do cliente ou não. No caso dessas duas *flags*, o valor 0 significa "falso" e o valor 1 significa "verdadeiro". A base também possui um atributo para a quantidade de transações do mesmo cliente realizadas nos últimos 10 minutos. O MCC é uma sigla para *Merchant category code*, em inglês, e é um código numérico de quatro dígitos estabelecido pelas bandeiras de cartão de crédito de acordo com padrão ISO 18245. Este código indica a natureza do estabelecimento onde a compra foi realizada. Por fim, o atributo Classe indica se a transação é legítima (0) ou fraude (1).

Tabela 2 – Atributos da base de dados

Atributo	Descrição
Valor	Valor da compra convertido em reais
Ticket médio	Ticket médio do cliente no momento da compra, em reais
Hora	Horário da compra
Dia da Semana	Dia da Semana da compra
Moeda	Código ISO 4217 da moeda da compra
Online	*Flag* indicando se a compra foi presencial ou não
Primeira compra	*Flag* indicando se a é a primeira compra do cliente
Parcelas	Quantidade de parcelas da compra
Limite	Limite do cartão do cliente no momento da compra
Quantidade	Quantidade de transações do mesmo cliente nos últimos 10 minutos
MCC	*Merchant category code* - Código do estabelecimento
Classe	Indicador de transação legítima ou fraude

Fonte: Elaborado pelo Autor (2021).

3.1.3 Exploração dos Dados

O atributo do Valor da compra da base de dados possui a maior variedade de valores, e por isso é um bom ponto de partida para a análise exploratória dos dados. A Tabela 3 apresenta os dados estatísticos deste atributo. A partir deles é possível observar que a média do valor de compra de transações de fraude é maior, mesmo que o valor máximo de uma compra fraudulenta ainda tenha sido menor do que o valor máximo de uma compra legítima. A tabela também mostra que o valor mínimo de compra de uma transação legítima foi um valor negativo, e o motivo de existirem transações com valor menor ou igual a zero na base de dados será explicado na seção seguinte.

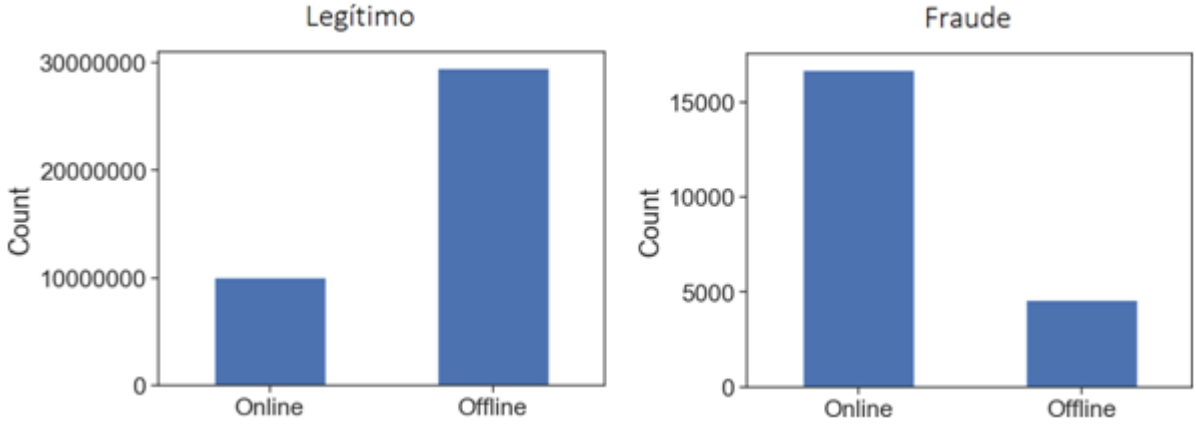
Tabela 3 – Estatísticas do Valor de Compra

	Legítimas	Fraudes
Quantidade	39.550.254	21.304
Média	75,97	150,24
Valor mín.	-16.029,54	0,01
Valor máx.	113.457,46	20.557,71
25%	13,00	22,37
50%	28,57	50,98
75%	68,80	179,44

Fonte: Elaborado pelo Autor (2021).

O comportamento do atributo *Online*, que indica se a transação foi feita de modo presencial ou não, é diferente nos conjuntos de transações legítimas e fraudes. Apesar do número de transações feitas pela *internet* ter crescido conforme descrito primeiro capítulo deste trabalho, essa modalidade de compra ainda não superou as transações presenciais, conhecidas como *Offline*, na observação do conjunto de transações legítimas (que são 99,94% do total da base). Já no conjunto de transações fraudulentas o comportamento é exatamente o oposto, e as compras *Online* são a maioria dos casos. A quantidade de transações *Online* e *Offline* de cada um desses conjuntos está ilustrado pela Figura 7. Essa diferença no comportamento mostra que este atributo será relevante para a detecção de fraudes pelo modelo.

Figura 7 – Quantidade de transações *Online* e *Offline*.



Fonte: Elaborado pelo autor (2021).

Outro atributo que indica o comportamento das transações é conhecido como *Merchant category code* ou MCC. Este é um código de quatro dígitos padronizado pelas bandeiras de cartão de crédito e indicam a natureza do estabelecimento onde a compra foi realizada, conforme explicado na seção anterior. Este trabalho analisou quais foram as cinco categorias com a maior quantidade de compras nos conjuntos de transações legítimas e de fraudes, como mostram as Tabelas 4 e 5 respectivamente. A descrição das categorias estão em inglês pois foram extraídas conforme o padrão ISO 18245 referente a cada código MCC.

Tabela 4 – Categorias com mais compras Legítimas

Categoria	**Quantidade**
Grocery Stores, Supermarkets	6.077.180
Taxicabs and Limousines	2.917.884
Service Stations	2.693.859
Fast Food Restaurants	2.321.704
Eating places and Restaurants	1.969.572

Fonte: Elaborado pelo Autor (2021).

Tabela 5 – Categorias com mais compras Fraudulentas

Categoria	**Quantidade**
Fast Food Restaurants	2.980
Direct Marketing	2.086
Taxicabs and Limousines	1.781
Cable and other pay television	1.400
Telecommunication Services	1.116

Fonte: Elaborado pelo Autor (2021).

No conjunto de transações legítimas, a categoria com mais compras é "*Grocery Stores, Supermarkets*", que são compras feitas em mercearias e supermercados como diz a tradução direta do inlgês. A segunda categoria com mais compras legítimas é "*Taxicabs and Limousines*", que diz respeito a transações feitas em aplicativos de transporte individual. A terceira categoria é "*Service Stations*", que são compras feitas em postos de gasolina no Brasil. A quarta categoria, "*Fast Food Restaurants*" são restaurantes de redes de *fast food* e compras em aplicativos de pedidos de comida por entrega. A quinta e última categoria com mais compras legítimas é "*Eating places and Restaurants*" que são transações feitas em restaurantes conforme a tradução direta do inglês.

Através destes dados é possível observar que as transações legítimas seguem dois hábitos de consumo em específico: alimentação e transporte. Outra observação interessante é que o primeiro item da lista, que são as compras em supermercados e mercearias, representam mais que o dobro do segundo colocado.

Já no conjunto de transações fraudulentas, a categoria "*Fast Food Restaurants*" aparece com o maior número de compras, com um número bem próximo da segunda categoria que é "*Direct Marketing*". Esta segunda categoria possui uma descrição genérica, mas essas transações têm origem em maquininhas de cartão de crédito individuais, muito populares no Brasil para pessoas que trabalham informalmente como autônomos ou pequenos negócios. Como terceiro item da lista aparecem as transações feitas em aplicativos de transporte individual categorizadas como "*Taxicabs and Limousines*". Na quarta categoria estão as compras categorizadas em "*Cable and other pay television*" que são transações feitas para serviços de TV a cabo ou assinaturas recorrentes de aplicativos de *streaming* de filmes e séries. O último item da lista com mais compras fraudulentas é a categoria "*Telecommunication Services*", que são transações de compra de crédito para celulares pré-pagos.

Logo, a análise de categorias com mais compras mostra que as transações fraudulentas tem um comportamento diferente das legítimas. Enquanto as transações legítimas tem foco em hábitos básicos de consumo, as de fraude tem foco em compras de ganho rápido e que na maioria das vezes não são presenciais e não exigem a senha do cartão, como aplicativos de transporte e de comida. Além disso, na lista das categorias de fraudes ainda aparecem itens de serviços de telefonia pré-paga e de lazer como assinatura de serviços de entretenimento.

3.2 PRÉ-PROCESSAMENTO DOS DADOS

3.2.1 Eliminação Manual de Atributos

Durante a exploração da base foram encontrados 106.551 registros de transações que possuem o valor de compra menor ou igual a zero. Transações com valor igual a zero existem devido à um mecanismo que verifica se os dados de um cartão de crédito são válidos antes de fazer a cobrança verdadeira, que é registrada posteriormente como outra transação. Este tipo de transação de verificação de dados é conhecida como *Zero Dollar Authorization*, em inglês, e não gera nenhuma cobrança ao portador do cartão. Já transações com valores negativos são considerados como créditos na fatura dos clientes. A natureza desses créditos é

variada, podendo ocorrer por cobranças indevidas, estorno de compras ou adiantamento de valores da próxima fatura.

Essas transações são geradas de forma automática e não representam, de fato, um registro de compra feito por um cliente portador do cartão e portanto não devem haver fraudes neste conjunto. Foi verificado que nenhum dos 106.551 registros foram classificados como fraude (ou seja, nenhum deles possui Classe igual a 1) como era esperado, e portanto eles foram removidos da base por não serem transações verdadeiras.

Depois da limpeza desses registros a base de dados passou a ter 39.465.007 transações de cartão de crédito, com 39.443.703 transações legítimas e 21.304 fraudes.

3.2.2 Normalização dos dados

Antes de serem submetidos ao modelo, os dados numéricos precisam passar por uma função de redimensionamento pois muitos atributos não são comparáveis entre si por possuírem escalas e unidades diferentes como, por exemplo, número de parcelas e valor da transação.

Este processo é conhecido como normalização dos dados. Nele, os dados são redimensionados para uma distribuição normal para que cada dimensão tenha média 0 e desvio padrão 1 conforme proposto por Grus (2019) e aplicado no trabalho de Shakya (2018).

O processo de normalização de dados evita que um modelo de rede neural privilegie, de forma errada, um determinado atributo com escala maior em detrimento de outros de menor escala. Se essa padronização não for realizada, o modelo terá seu desempenho afetado (SHAKYA, 2018).

A função de normalização é aplicada individualmente para cada atributo numérico da base de dados. A sua equação é dada por:

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

onde *x* é o valor do atributo escolhido para ser normalizado, μ é o valor da média do atributo e σ é o valor do desvio padrão do mesmo atributo. Os valores de μ e σ são calculados com as seguintes fórmulas:

$$\mu = \frac{1}{N}\sum_{i=1}^{N}(x_i) \tag{2}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{3}$$

onde *N* é a quantidade total de registros do atributo *x*.

3.3 MODELO AUTOENCODER

Este trabalho foi desenvolvido utilizando a linguagem de programação *Python* 3 na plataforma *Jupyter Notebook*. Para o desenvolvimento do modelo foram utilizadas classes disponibilizadas pela biblioteca *Keras* com *TensorFlow*.

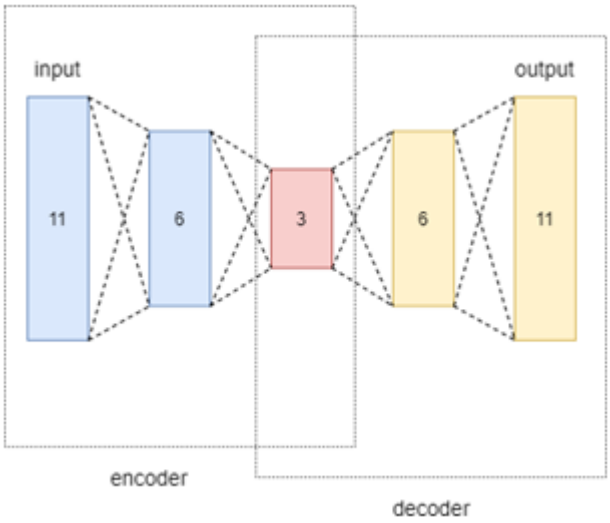
A dimensão das camadas de entrada e saída do *Autoencoder* é a mesma quantidade de características da base de dados. No caso deste trabalho são 11 neurônios. As camadas intermediárias entre a entrada e saída, também chamadas de camadas ocultas, possuem um número de neurônios menor do que a quantidade de características. Essa redução possibilita que o *Autoencoder* extraia as principais características do conjunto de dados e então as reconstrua novamente para atingir a mesma quantidade de características na saída. As camadas de redução são chamadas de *encoder* e as de reconstrução de *decoder*.

O modelo é formado por duas camadas ocultas no *encoder* com 6 e 3 neurônios respectivamente, e então o *decoder* com o mesmo número de neurônios de forma espelhada. Isso vai garantir que o registro de saída possua o mesmo número de características do registro de entrada.

A arquitetura de camadas do *Autoencoder* está representada pela Figura 8.

Para executar o *Autoencoder* foi preciso definir mais três parâmetros: métrica geral, função de otimização e métrica de perda. A métrica geral é o que o modelo utiliza para avaliar seus próprios resultados durante a etapa de treinamento e para isso foi escolhida a exatidão (*accuracy*). A função de otimização escolhida foi a *Adam*, ideal para modelos que lidam com grandes bases de dados. Já para a métrica de perda, que também é um valor avaliado durante o treinamento, foi escolhido o Erro Quadrático Médio conforme descrito na seção seguinte.

Figura 8 – Arquitetura do *Autoencoder*.



Fonte: Elaborado pelo autor (2021).

3.3.1 Métrica da Taxa de Perda do Autoencoder

Dado que o objetivo do *Autoencoder* é reconstruir na saída os mesmos atributos informados na entrada, é preciso medir o quanto de perda há no processo de reconstrução. O cálculo desta perda será importante para definir o limiar que vai distinguir um registro de ser uma anomalia ou não. Os registros que possuem uma alta taxa de perda na reconstrução são considerados anomalias por não serem iguais aos já conhecidos pelo modelo e, portanto, esses serão classificados como fraude nos termos deste estudo.

A métrica utilizada para medir a taxa de perda na reconstrução é o Erro Quadrático Médio (*mean squared error* ou MSE, em inglês), a mesma utilizada pelo trabalho de Pumsirirat e Yan (2018). Isso permite avaliar o modelo de forma a verificar se ele

teve alta acurácia ao reconstruir os valores, sem estar necessariamente enviesado nos resultados. Quanto menor o valor do erro quadrático médio, melhor é o resultado do modelo. O valor ideal do MSE é um número próximo de 0 que significará uma baixa perda na reconstrução do registro pelo *Autoencoder*.

A equação do MSE é dada por:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2 \tag{4}$$

onde *N* é a quantidade total de registros, *Y* é a representação do registro original e *Yˆ* a representação do registro depois de ter sido reconstruído pelo *Autoencoder*.

3.3.2 Amostragem

A divisão para gerar os subconjuntos de treinamento e teste foi feita usando o atributo Mês. As transações de Janeiro a Outubro (valores 1 a 10) foram separadas para treinamento e os meses de Novembro e Dezembro (valores 11 e 12, respectivamente) foram separadas para teste do modelo. Como esse atributo foi usado para dividir os registros, ele não foi usado na avaliação do modelo. Essa estratégia foi escolhida porque simula melhor o mundo real. Os meses de Novembro e Dezembro, nessa estratégia, simulam as novas transações e fraudes ocorridas ao longo do tempo, utilizando apenas os registros anteriores como dados de treinamento.

Os modelos preditivos para detecção de anomalias devem ser treinados somente com dados da classe majoritária, que no caso deste trabalho são as transações legítimas. Isso permite que o modelo detecte anomalias (as fraudes) ao encontrar transações que não estão relacionadas com o padrão aprendido. Por conta disso todos os registros de fraude (onde o atributo Classe é igual a 1) foram removidos da base de treinamento. Ao final as bases de treinamento e teste possuem 30.449.205 e 8.999.221 registros respectivamente.

3.3.3 Métricas de Avaliação

Com base na classe verdadeira dos dados e na classe prevista pelo *Autoencoder*, os resultados do modelo preditivo foram categorizados como: verdadeiro positivo (*True Positive*, TP em inglês), falso negativo (*False Negative*, FN em inglês), falso positivo (*False Positive*, FP em inglês) e verdadeiro negativo (*True Negative*, TN em inglês). Essas categorias serão colocadas em uma Matriz de Confusão conforme ilustrado pela Figura 9.

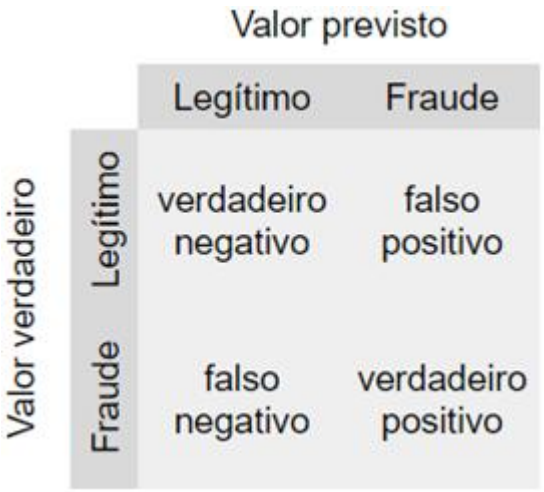
As seguintes métricas de avaliação para o modelo foram calculadas com a matriz de confusão: *Accuracy* (exatidão), *Recall* (sensibilidade) e *Specificity* (especificidade). O cálculo destas métricas é dado conforme as fórmulas 5, 6 e 7 a seguir. O resultado dessas métricas são números na escala de 0 a 1, onde 1 representa a excelência na predição.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{5}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{6}$$

$$Specificity = \frac{TN}{(TN + FP)} \tag{7}$$

Figura 9 – Matriz de confusão.



Fonte: Elaborado pelo autor (2021).

Como este trabalho utiliza uma base desbalanceada, as métricas descritas acima não são suficientes para refletir a performance do modelo. Como a quantidade de transações legítimas é mais representada na base (99,94% do total), um modelo que, por exemplo, erre ao classificar todas as transações de fraude vai possuir uma exatidão muito alta. Isso remete a uma falsa impressão de que o modelo é eficiente, quando na verdade não está detectando fraude alguma.

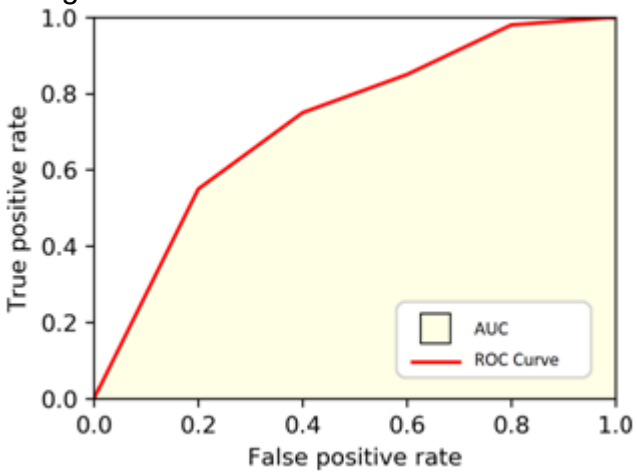
Por conta disso, outras duas métricas conhecidas como *Matthews Correlation Coefficient* (MCC) e *Area Under The Curve* (AUC) foram utilizadas para análise da performance por não serem afetadas pelo desbalanceamento das classes. Essas métricas foram utilizadas para avaliar os classificadores nos trabalhos de Awoyemi, Adetunmbi e Oluwadare (2017) e Bhattacharyya et al. (2011) dado sua relevância na avaliação de problemas de classificação binária em bases desbalanceadas, como é o caso da base de fraudes de cartão de crédito. O cálculo do MCC é dado pela fórmula a seguir:

$$MCC = \frac{TP.TN - FP.FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (8)$$

O resultado do MCC varia entre -1 e +1, onde um valor de +1 representa uma classificação excelente e o valor de -1 representa distinção total entre classificação e predição.

O AUC é calculado pela área bidimensional abaixo da curva ROC (*Receiver Operating Characteristic curve*). O gráfico da curva ROC representa a taxa de falsos positivos (*False Positive Rate*) no eixo X e a taxa de verdadeiros positivos (*True Positive Rate*) no eixo Y em diferentes limiares de classificação conforme ilustrado pela Figura 10. O ponto ideal da curva ROC é o canto superior esquerdo do gráfico, pois isso maximiza a área abaixo da curva.

Figura 10 – Exemplo de gráfico da Curva ROC.



Fonte: Elaborado pelo autor (2021).

O resultado da AUC é um valor que varia entre 0 e 1. Um modelo excelente possui AUC próximo de 1, o que significa que ele tem uma boa medida de separabilidade entre as classes. Um modelo ruim tem AUC próximo do 0, que significa que a classificação está sendo feita de forma inversa ao esperado. Quando o AUC é 0.5 significa que o modelo não possui capacidade alguma de separação de classes.

4 EXPERIMENTOS, RESULTADOS E DISCUSSÃO

Neste capítulo serão apresentados os experimentos realizados com o modelo *Autoencoder*

bem como os resultados obtidos na detecção de fraudes com a base de transações Brasileira.

4.1 TREINAMENTO E TESTE DO MODELO

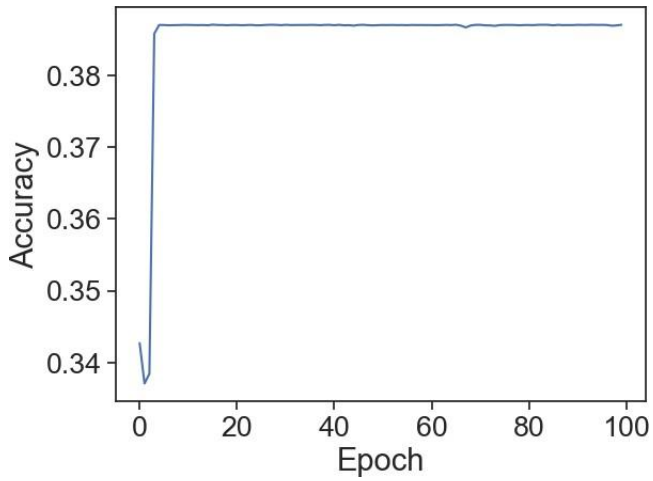
O modelo *Autoencoder* criado no capítulo anterior foi submetido a uma etapa de treinamento com 100 épocas e com o tamanho de lote (*batch size*, em inglês) de 128 unidades. Essas configurações permitem que o modelo seja treinado de forma paralela utilizando todos os recursos disponíveis ganhando velocidade em bases com grande volume de dados.

Durante este processo o modelo se auto-avalia a cada época buscando otimizar-se baseado nos parâmetros que foram definidos em sua inicialização. A base de treinamento contém 30.449.205 registros, onde todas transações são legítimas porque o modelo deve ser treinado apenas com dados da classe majoritária conforme explicado no capítulo anterior.

Como o objetivo do *Autoencoder* é reconstruir em sua saída os mesmos registros fornecidos como entrada, a base de treinamento é fornecida duas vezes para treinar o modelo. A base é fornecida tanto como parâmetro de entrada quanto como parâmetro de resultado da rede para que ela possa se autoavaliar e calcular o quanto de perda está havendo durante o processo de reconstrução.

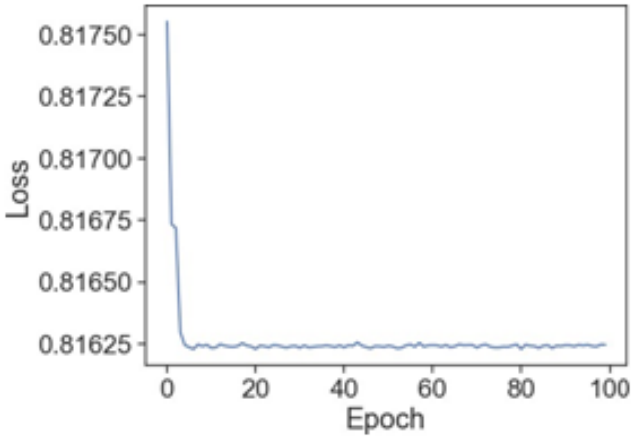
O processo de treinamento do modelo durou cerca de 4 horas e 40 minutos. A curva de acurácia e a taxa de perda (representado pelo *mean squared error*, MSE) de cada uma das épocas do treinamento do modelo estão ilustrados pelas Figuras 11 e 12 respectivamente. É possível notar que o modelo aumenta a acurácia e reduz o MSE a cada época do treinamento através da autoavaliação e ajustes dos pesos entre as camadas intermediárias para encontrar as características mais importantes do conjunto de dados.

Figura 11 – Acurácia do *Autoencoder*.



Fonte: Elaborado pelo autor (2021).

Figura 12 – Taxa de perda do *Autoencoder*.



Fonte: Elaborado pelo autor (2021).

Após ser treinado, o *Autoencoder* é submetido à etapa de predição utilizando a base de testes. A base de testes contém 8.999.221 transações de cartão de crédito, com 8.994.498 registros classificados como legítimos e 4.723 classificados como fraudes. O resultado da etapa de predição do modelo é uma matriz no mesmo formato da base de entrada, contendo 8.999.221 registros que foram reconstruídos contendo 11 características em cada um.

A etapa de predição levou cerca de 1 minuto e 5 segundos para ser concluída, ou seja, este foi o tempo total que o *Autoencoder* levou para reconstruir todos os registros da base de testes. Nesse tempo de processamento, cada registro não

levou meio segundo para ser reconstruído. Em seguida, o *Mean Squared Error* é calculado para cada um destes registros para saber quanto de perda houve no processo de reconstrução.

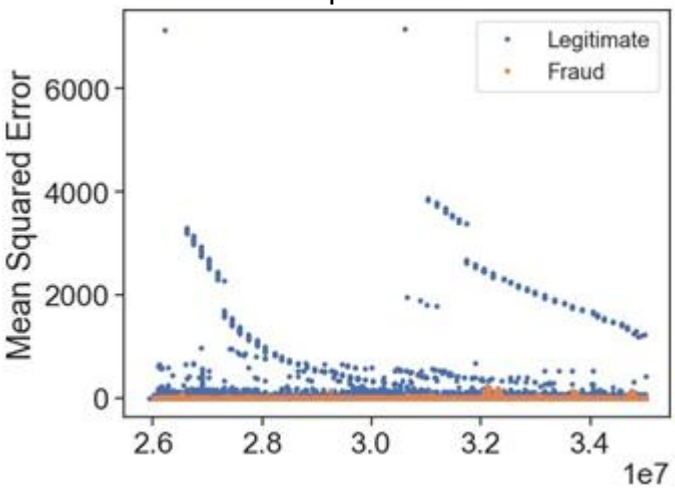
4.2 DEFINIÇÃO DO LIMIAR DO *AUTOENCODER*

A classificação dos registros reconstruídos como anomalias e, portanto, fraudes é feito através da criação de um limiar. De posse das taxas de erro de cada uma das transações de saída do *Autoencoder*, é preciso encontrar um limiar para definir com qual quantidade de erro um registro será classificado como fraude. A Figura 13 ilustra as diferentes taxas de erros (*mean squared error*) dos registros com o rótulo original das classes da base de testes (legítimas e fraudes).

É possível observar que as classes não são linearmente separáveis pois muitas transações de fraude possuem erros de reconstrução semelhantes a transações legítimas. Além disso, existem transações legítimas com uma alta taxa de erro, e algumas delas podem ser fraudes que ainda não foram relatadas pelos clientes titulares dos cartões.

A Figura 14 ilustra os valores das métricas *Specificity* e *Recall* para diferentes limiares de erro de reconstrução. O gráfico mostra que existe um balanceamento entre as duas métricas, onde enquanto uma métrica aumenta, a outra diminui. Após a análise destes dados, foi escolhido o valor 3 como o limiar da taxa de erro para separação dos registros.

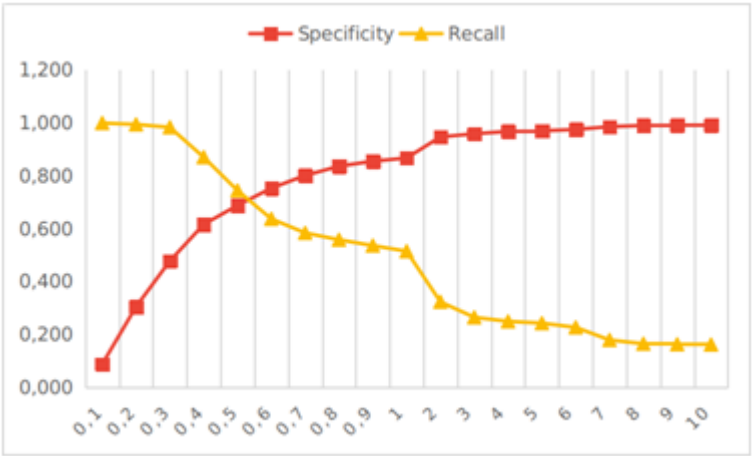
Figura 13 – Taxa de erro do *Autoencoder* por Classe.



Fonte: Elaborado pelo autor (2021).

Tal valor maximiza a métrica de *Specificity*, não havendo aumento significativo após este ponto.

Figura 14 – *Specificity* e *Recall* para diferentes limiares.



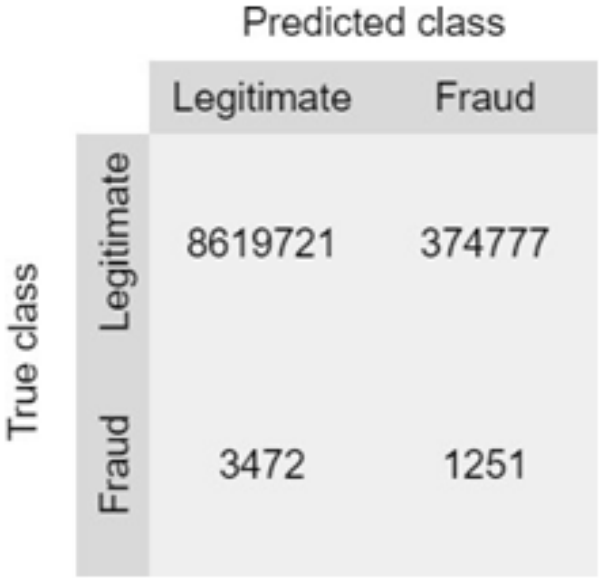
Fonte: Elaborado pelo autor (2021).

4.3 RESULTADOS DA CLASSIFICAÇÃO

Com a definição do limiar da taxa de erro de reconstrução, as transações resultantes da fase de predição do *Autoencoder* foram classificadas como legítimas ou fraudes. Após essa classificação, as métricas de avaliação do modelo foram geradas e são apresentadas a seguir.

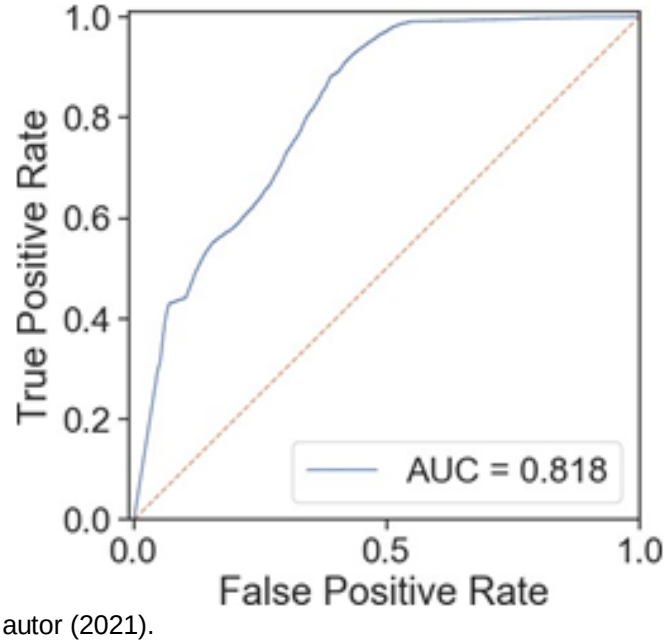
A Matriz de confusão do *Autoencoder* com o resultado da classificação após a definição do limiar está ilustrado pela Figura 15. O Gráfico da Curva ROC com a taxa de verdadeiros positivos e falso positivos, para cálculo da AUC, está ilustrada pela Figura 16. As métricas de avaliação extraídas a partir da matriz de confusão e do gráfico da curva ROC estão consolidadas e apresentadas na Tabela 6.

Figura 15 – Matriz de Confusão do *Autoencoder*.



Fonte: Elaborado pelo autor (2021).

Figura 16 – Gráfico da Curva ROC do *Autoencoder*.



Fonte: Elaborado pelo autor (2021).

Tabela 6 – Métricas do *Autoencoder*

Métrica	Valor
Accuracy	0.95
Specificity	0.95
Recall	0.26
MCC	0.02
AUC	0.81

Fonte: Elaborado pelo autor (2021).

A partir destes dados é possível observar que os valores de *Accuracy* e *Specificity* chegaram a 0,95 (próximos de 1, o valor de uma classificação ideal), como era esperado por conta do desbalanceamento de classes da base. Apesar de existirem classificações erradas elas não parecem ter grande impacto nestes valores. Portanto, ambas as métricas podem levar a uma falsa impressão de bom desempenho. Essas duas métricas usam verdadeiros negativos (transações legítimas, a classe mais representada) como um numerador em sua fórmula.

A métrica *Recall* foi de 0,26, um valor menor do que as duas métricas anteriores. Isso ocorre porque o modelo previu mais falsos negativos do que verdadeiros positivos, conforme observado pela matriz de confusão, ou seja, ele classifica mais transações fraudulentas como sendo legítimas do que pode detectar fraudes reais.

Como as fraudes reais pertencem à classe menos representada na base de dados, essa métrica também pode levar a uma falsa impressão de baixo desempenho. É possível aumentar o número de classificações de verdadeiros positivos e, portanto, melhorar o *Recall*, mas isso também pode aumentar o número de classificações de falsos positivos.

As métricas AUC e MCC são importantes para a continuidade da análise de desempenho do *Autoencoder*, visto que as métricas anteriores são sensíveis ao desbalanceamento de classes. Um valor MCC positivo e um valor AUC de 0,81 (superior a 0,5 e mais próximo de 1) indicam que o modelo possui uma boa capacidade geral de distinção de classes e não está classificando de forma inversa ao que era esperado.

Esses resultados foram descritos, apresentados e publicados no Simpósio Brasileiro de Automação Inteligente (SBAI), que aconteceu no dia 17 de Outubro de 2021.

4.4 COMPORTAMENTO SOB O PONTO DE VISTA DO EMISSOR

O modelo *Autoencoder* é capaz de classificar corretamente 95% das transações, conforme mostrado pela métrica de exatidão (*Accuracy*). No entanto, para cada classificação errada, há um custo diferente do ponto de vista do emissor do cartão de crédito. Este modelo também aprovaria 3.472 transações fraudulentas como sendo legítimas (falsos negativos). Essas transações gerariam perdas financeiras de qualquer forma se não houvesse um sistema de detecção de fraude.

Da perspectiva de um emissor de cartão de crédito que decide implementar um sistema de fraude com *Autoencoder*, seu modelo classificaria 1.251 transações como fraudes corretamente e 374.777 incorretamente. Se o emissor decidir bloquear automaticamente todas as transações classificadas como fraude, não haverá perda financeira de 1.251 transações que seriam fraudulentas.

Mas esse bloqueio também iria abranger as outras 374.777 transações legítimas classificadas incorretamente como fraude (os falsos positivos). Isso pode causar frustração ao cliente, porque essas seriam compras legítimas que não estão sendo aprovadas pelo modelo.

Se o emissor do cartão de crédito preferir mudar o comportamento e melhorar a classificação para detectar mais transações fraudulentas reais, a análise do Gráfico da curva ROC mostra que há um valor de limiar onde é possível aumentar a taxa de classificação de verdadeiros positivos (ou seja, aumentar a identificação de fraudes reais), mas ao custo de aumentar onúmero de falsos positivos e, como resultado, isso pode aumentar as frustrações dos clientes. Este *trade-off* deve ser cuidadosamente analisado pelo emissor, pois afeta diretamente a experiência do cliente.

5 CONSIDERAÇÕES FINAIS

Casos de fraudes em cartões de crédito se tornaram um assunto recorrente nos últimos anos devido ao crescimento da *internet* e da facilidade de compras *online* através de *e-commerces*. Isso faz com que bancos e instituições financeiras desempenhem um papel importante no combate a este tipo de fraude, desenvolvendo sistemas automatizados para reduzir custos de análise manual, reduzir prejuízos financeiros e aumentar sua credibilidade.

Modelos de aprendizado não supervisionado e de detecção de anomalias são eficazes para análises de bases grandes e desbalanceadas, e conseguem detectar novos padrões de comportamento sem precisarem de um novo treinamento. Por isso, neste trabalho foi desenvolvido uma rede neural do tipo *Autoencoder* utilizando algoritmos de *Deep Learning* para a detecção de fraudes em transações de cartões de crédito. Este tipo de modelo de rede neural reconstrói em sua saída os mesmos registros fornecidos na entrada, de forma a extrair as características mais importantes dos registros da base. Para cada registro reconstruído na saída, o modelo também fornece a taxa de erro resultante do processo de reconstrução.

O modelo *Autoencoder* criado neste trabalho foi treinado e validado usando uma base de dados própria fornecida por uma *fintech* Brasileira do ramo de meios de pagamento. Essa base contém 39.571.558 registros de transações de cartão de crédito feitos no ano de 2019, sendo que destes apenas 21.304 são rotulados como fraude tratando-se, portanto, de uma base desbalanceada.

Após analisar a taxa de erro de reconstrução dos registros do conjunto de teste, foi definido um limiar de separação para que cada registro pudesse ser classificado como uma transação legítima ou fraudulenta. Os experimentos apresentaram resultados satisfatórios mostrando que o modelo *Autoencoder* mantém a separabilidade de classes na base de dados brasileira com dados reais de transações, diferente dos trabalhos que utilizam a base de dados Europeia. Além disso, o *trade-off* de performance do ajuste do limiar da taxa de erro se comporta da mesma maneira do que nos trabalhos correlatos.

Um valor de MCC positivo (0,02) e um AUC de 0,81, que não são afetados pelo desbalanceamento da base de dados, também mostraram que o *Autoencoder* tem

uma boa capacidade de separação de classes. Vale ressaltar que a base de dados contém apenas as fraudes mais desafiadoras, que não foram detectadas anteriormente pelos sistemas de anti-fraudes da adquirente e da bandeira do cartão de crédito. Além disso, o resultado é promissor considerando que a base de dados pode conter transações legítimas que na verdade são fraudes, já que a rotulação de cada transação é feita através do contato de clientes que relatam compras não reconhecidas e após uma análise manual pela área responsável da *fintech*.

Apesar do resultado do experimento ter conseguido classificar com sucesso as transações que eram verdadeiras fraudes da base de teste, ainda houve transações legítimas classificadas erroneamente como fraude (os falsos positivos) pelo *Autoencoder*. As transações que caem nesta categoria podem gerar frustração na experiência dos clientes, pois estes teriam suas compras negadas caso o emissor do cartão de crédito decidisse implementar um sistema de bloqueio automático de todas as transações classificadas como fraude.

O emissor do cartão deve avaliar o custo de introduzir um sistema de detecção de fraude que consiga evitar perdas financeiras mas que pode, de alguma forma, impactar na experiência do cliente. Reduzir o valor do limiar do *Autoencoder* para detectar mais transações de fraudes verdadeiras (os verdadeiros positivos) é um *trade-off* em relação à experiência, pois também aumenta o número de casos de falsos positivos.

Misra et al. (2020) reconhece esse impacto ao dizer que qualquer tipo de sistema de detecção de fraude estaria sujeito a erros, como identificar de forma errada uma transação legítima como fraude ou vice-versa. É necessário encontrar um equilíbrio para minimizar essas situações. Um grande número de fraudes perdidas pode gerar prejuízos para pessoas e empresas. Por outro lado, muitos casos de transações legítimas declaradas como fraude faria com que as pessoas deixassem de confiar na organização com tal sistema. Portanto, o problema se torna bastante desafiador.

A análise visual das taxas de erro de reconstrução do *Autoencoder* mostra que existem muitas similaridades entre transações legítimas e fraudes, não sendo possível separá-las linearmente. Isto pode acontecer pois um fraudador tenta ao máximo fazer uma transação parecer legítima para evitar o bloqueio da mesma.

5.1 TRABALHOS FUTUROS

A publicação da base de dados está planejada de modo que ela possa ser usada e posteriormente comparada com outros trabalhos relacionados. O foco do trabalho foi encontrar um modelo de rede neural com uma análise de limiar adequado do *Autoencoder* que forneça uma boa classificação de transações fraudulentas. A base de dados será publicada de forma anônima para sigilo dos dados, uma vez que a empresa não permite a divulgação de seu nome.

Uma melhoria para trabalhos futuros seria combinar diferentes algoritmos de *Deep Learning* através da Aprendizagem em Conjunto (*Ensemble learning*, em inglês) (SOHONY; PRATAP; NAMBIAR, 2018) buscando reduzir a taxa de falsos positivos para melhorar a experiência do cliente, sem penalizar a detecção de verdadeiros positivos (as fraudes verdadeiras).

Outra sugestão de melhoria seria implementar níveis de confiança nas taxas de erro de reconstrução do *Autoencoder* para que a classificação não seja somente binária com limiar fixo. Desta forma, cada nível de confiança pode levar a ações diferentes como, por exemplo, exigir uma verificação do cliente portador do cartão enviando uma mensagem de texto para que ele confirme ou não a compra, antecipando uma possível fraude.

REFERÊNCIAS

AL-SHABI, MA. Credit card fraud detection using autoencoder model in unbalanced datasets. **Journal of Advances in Mathematics and Computer Science**, p. 1–16, 2019.

AWOYEMI, John O.; ADETUNMBI, Adebayo O.; OLUWADARE, Samuel A. Credit card fraud detection using machine learning techniques: A comparative analysis. In: IEEE. INTERNATIONAL CONFERENCE ON COMPUTING NETWORKING AND INFORMATICS (ICCNI). 2017. **Proceedings...** [S.l.], 2017. p. 1–9.

BANCO CENTRAL DO BRASIL. **Estatísticas de Pagamentos de Varejo e de Cartões no Brasil.** 2021. Disponível em: https://www.bcb.gov.br/estatisticas/spbadendos. Acesso em: 1 out. 2021.

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

BRESLOW, Stuart et al. The new frontier in anti–money laundering. **McKinsey & Company**, New York, nov. 2017.

CHAN, Philip K et al. Distributed data mining in credit card fraud detection. I**EEE Intelligent systems**, n. 6, p. 67–74, 1999.

CHANDOLA, Varun; BANERJEE, Arindam; KUMAR, Vipin. Anomaly detection: a survey. **ACM computing surveys (CSUR)**, v. 41, n. 3, p. 15, 2009.

FACELI, Katti et al. **Inteligência artificial**: uma abordagem de aprendizado de máquina. [S.l.: s.n.], 2011.

FIORE, Ugo et al. Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. **Information Sciences**, Elsevier, 2017.

GRUS, Joel. **Data science from scratch: first principles with python**. [S.l.]: O'Reilly Media, 2019.

KAZEMI, Zahra; ZARRABI, Houman. Using deep networks for fraud detection in the credit card transactions. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE-BASED ENGINEERING AND INNOVATION (KBEI). 4., 2017. **Proceedings…** [S.l.], 2017. p. 0630–0633.

KNIEFF, Ben. **Global consumer card fraud**: where card fraud is coming from. [S.l.: s.n.], 2016.

KOU, Yufeng et al. Survey of fraud detection techniques. In: IEEE INTERNATIONAL CONFERENCE ON NETWORKING, SENSING AND CONTROL. 2004. **Proceedings…** [S.l.], 2004. v. 2, p. 749–754.

MAES, Sam et al. Credit card fraud detection using bayesian and neural networks. In: INTERNATIONAL NAISO CONGRESS ON NEURO FUZZY TECHNOLOGIES. 1., 2002. **Proceedings…** [S.l.: s.n.], 2002. p. 261–270.

MISRA, Sumit et al. An autoencoder based model for detecting fraudulent credit card transaction. **Procedia Computer Science**, Elsevier, v. 167, p. 254–262, 2020.

OXFORD. Fraud. In: **Oxford Learner's Dictionary**. [s.n.], 2021. Disponível em: https://www.oxfordlearnersdictionaries.com/us/definition/english/fraud.

PANDEY, Yamini. Credit card fraud detection using deep learning. **International Journal of Advanced Research in Computer Science**, v. 8, n. 5, 2017.

PATCHA, Animesh; PARK, Jung-Min. An overview of anomaly detection techniques: Existing solutions and latest technological trends. **Computer networks**, Elsevier, v. 51, n. 12, p. 3448–3470, 2007.

PUMSIRIRAT, Apapan; YAN, Liu. Credit card fraud detection using deep learning based on auto-encoder and restricted boltzmann machine. **International Journal of advanced computer science and applications,** v. 9, n. 1, p. 18–25, 2018.

REZAPOUR, Mahdi. Anomaly detection using unsupervised methods: credit card fraud case study. **Int J Adv Comput Sci Appl**, v. 10, n. 11, 2019.

ROY, Abhimanyu et al. Deep learning detecting fraud in credit card transactions. In: SYSTEMS AND INFORMATION ENGINEERING DESIGN SYMPOSIUM (SIEDS). 2018. **Proceedings…** [S.l.], 2018. p. 129–134.

RUSHIN, Gabriel et al. Horse race analysis in credit card fraud-deep learning, logistic regression, and gradient boosted tree. In: SYSTEMS AND INFORMATION ENGINEERING DESIGN SYMPOSIUM (SIEDS). 2017. **Proceedings...** [S.l.], 2017. p. 117–121.

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) - Universidade de São Paulo, 2014.

SHAKYA, Ronish. **Application of machine learning techniques in credit card fraud detection**. [S.l.: s.n.], 2018.

SILVA, Bruno Riccelli dos Santos. **Uma análise comparativa de técnicas de subamostragem para projetos de sistemas de detecção de intrusão em redes de computadores** 2020. 85 f. Dissertação (Mestrado em Engenharia de Teleinformática) - Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza, 2020.

SOHONY, Ishan; PRATAP, Rameshwar; NAMBIAR, Ullas. Ensemble learning for credit card fraud detection. In: ACM INDIA JOINT INTERNATIONAL CONFERENCE ON DATA SCIENCE AND MANAGEMENT OF DATA. 2018. **Proceedings…** [S.l.: s.n.], 2018. p. 289–294.

ZHANG, James; GARDNER, Robert; VUKOTIC, Ilija. Anomaly detection in wide area network meshes using two machine learning algorithms. **Future Generation Computer Systems**, Elsevier, v. 93, p. 418–426, 2019.