COIMBRA BUSINESS SCHOOL

iscac Politécnico de Coimbra

100 ANOS

Análise comparativa de algoritmos de Inteligência Artificial na deteção da Fraude Transacional em contexto bancário: uma revisão sistemática de literatura

Pedro Alexandre T.F. da Mata

ISCAC | 2023

COIMBRA BUSINESS SCHOOL

iscac Politécnico de Coimbra

100 ANOS

Pedro Alexandre Teixeira Fangueiro da Mata

Análise comparativa de algoritmos de Inteligência Artificial na deteção da Fraude Transacional em contexto bancário: uma revisão sistemática de literatura

Coimbra, agosto de 2023

Pedro Alexandre Teixeira Fangueiro da Mata

Análise comparativa de algoritmos de Inteligência Artificial na deteção da Fraude Transacional em contexto bancário: uma revisão sistemática de literatura

Dissertação ao Instituto Superior de Contabilidade e Administração de Coimbra para cumprimento dos requisitos necessários à obtenção do grau de **Mestre em Análise de Dados e Sistemas de Apoio à Decisão**, realizado sob a orientação do Professor Doutor António Trigo.

Coimbra, agosto de 2023

TERMO DE RESPONSABILIDADE

Declaro ser o autor desta dissertação, que constitui um trabalho original e inédito, que nunca foi submetido a outra Instituição de ensino superior para obtenção de um grau académico ou outra habilitação. Atesto ainda que todas as citações estão devidamente identificadas e que tenho consciência de que o plágio constitui uma grave falta de ética, que poderá resultar na anulação da presente dissertação.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: presidencia@iscac.pt; www.iscac.pt

AGRADECIMENTOS

Do Porto a Coimbra. De Direito aos dados. Uma viagem, uma aventura, repleta de medos e receios, de estudo e dedicação, de amizades e memórias.

Vários são os meus agradecimentos.

Primeiro aos meus colegas e amigos que conheci ao longo deste mestrado, que sem eles, muitas mais dificuldades, ânsias e desespero enfrentaria. A eles, pela ajuda, calma e apoio ao longo destes dois anos.

À minha família e amigos de longa data, por estarem sempre presentes, por ouvirem constantemente as lamurias e dificuldades daquilo que agora parece tão pouco e fugaz.

À minha namorada, Cátia, que nas longas noites de após aulas, durante as longas chamadas telefónicas em Coimbra, ouvia o debitar de matéria, as histórias do dia-a-dia, as minhas incertezas, e o constante conflito interno se teria capacidade em continuar.

Ao meu orientador, Professor Doutor António Trigo, por todo o apoio, pelas várias conversas, opiniões, aconselhamentos e acima de tudo, pela compreensão tida.

A todos, um eterno Obrigado.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: presidencia@iscac.pt; www.iscac.pt

RESUMO

Reconhecer a dificuldade da mitigação e deteção de fraude não significa que ficamos simplesmente reduzidos à esperança de que tudo irá correr para o melhor. Nos dias que correm, a fraude, é cada vez mais frequente e flexível, constantemente atualizando-se, sendo necessária uma luta diária constante, em que a passividade social não é solução. As instituições financeiras estão constantemente na linha da frente nesta batalha, principalmente a banca, setor altamente escrutinado, e incessantemente desejável pelos vários agentes fraudulentos, alvo das suas mais variadas técnicas e tentativas criminosas. Numa análise à perspetiva da fraude transacional, e suscitando o apoio da tecnologia, mais concretamente da Inteligência Artificial, levanta-se a questão: "Qual ou quais os melhores algoritmos de deteção de fraude transacional no contexto bancário?"

Ora, através de uma Revisão Sistemática da Literatura, e reduzindo a 8 referências bibliográficas credíveis e comparáveis, analisamos, segundo a opinião dos seus autores, o melhor ou os melhores algoritmos de deteção de fraude transacional no setor bancário/financeiro. Após esta análise, agrupamos os resultados obtidos pelas 8 referências e observamos aqueles que foram selecionados mais vezes atingindo a um pódio de 2 algoritmos – Regressão Logística e *Random Forest*.

Apesar das conclusões, apenas 1 dos 8 algoritmos teve em consideração a perspetiva bancária/financeira, as suas particularidades e desafios, seja na ótica operacional e/ou de tempo de processamento transacional. Por outro lado, a escolha do pódio ser bastante dúbia, já que os dois algoritmos apenas foram escolhidos mais 1 vez que os restantes das referências. Não é de todo consensual esta preferência. Estamos perante uma área de intensa investigação, que apesar de atualmente haver uma necessidade de uma abordagem mais aprofundada por parte da futura bibliografia, seguramente diferente resultados, análises e conclusões virão.

Palavras-chave: Fraude; Fraude bancária; Fraude transacional; Deteção de fraude; Inteligência artificial; Algoritmos de aprendizagem automática; Algoritmos de aprendizagem profunda.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: presidencia@iscac.pt; www.iscac.pt

ABSTRACT

Recognizing the difficulty of mitigating and detecting fraud does not mean that we are simply reduced to hoping that everything will turn out for the best. Nowadays, fraud is increasingly frequent and flexible, constantly updating itself, requiring a constant daily struggle, in which social passivity is not a solution. Financial institutions are constantly at the forefront of this battle, especially banks, a highly scrutinized sector, and incessantly desirable by various fraudulent agents, being the target of their most varied techniques and criminal attempts. In a transactional fraud analysis perspective, and calling for the support of technology, more specifically Artificial Intelligence, the question arises: "Which are the best algorithms for detecting transactional fraud in the banking context?"

Now, through a Systematic Literature Review, and reducing to 8 credible and comparable bibliographical references, we analyze, according to the authors' opinion, the best transactional fraud detection algorithms in the banking/financial sector. After this analysis, we grouped the results obtained by the 8 references and observed those that were selected more times reaching a podium of 2 algorithms – Logistic Regression and Random Forest.

Despite the conclusions, only 1 of the 8 algorithms considered the banking/financial perspective, its particularities, and challenges, whether from an operational perspective and/or transactional processing time. On the other hand, the choice of the podium is rather dubious, since the two algorithms were only chosen 1 more time than the rest of the bibliography. This preference is not at all consensual. This is an area of intense research, and although there is currently a need for a more in-depth approach in future literature, there will certainly be different results, analyses, and conclusions to come.

Keywords: Fraud; Bank fraud; Transactional fraud; Fraud detection; Artificial intelligence; Machine learning algorithms; Deep learning algorithms.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: presidencia@iscac.pt; www.iscac.pt

ÍNDICE GERAL

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: presidencia@iscac.pt; www.iscac.pt

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: presidencia@iscac.pt; www.iscac.pt

ÍNDICE DE FIGURAS

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: presidencia@iscac.pt; www.iscac.pt

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: presidencia@iscac.pt; www.iscac.pt

ÍNDICE DE TABELAS

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: presidencia@iscac.pt; www.iscac.pt

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: presidencia@iscac.pt; www.iscac.pt

LISTA DE ABREVIATURAS, ACRÓNIMOS E SIGLAS

- *Accuracy (AC)*
- *Adaptive Boosting (AdaBoost)*
- *Adaptive Synthetic (ADASYN)*
- *Antes de cristo (A.C)*
- *Area Under the Precision-Recall (AUPR)*
- *Area Under the Receiver Operating Characteristic Curve (AUC-ROC)*
- *Artificial Neural Network (ANN);*
- *Charles Sturt University (CSU)*
- *Conditional Random Fields (CRF);*
- *Decision Tree/Árvore de Decisão (DT/AD)*
- *Edited Nearest Neighbour (ENN)*
- *Extra Tree (ET)*
- *Extreme Gradient Boosting (XGBoost)*
- *False Positive (FP)*
- *Gated Recurrent Unit (GRU);*
- *Hidden Markov Model (HMM);*
- *Information Gain (IG)*
- *Inteligência Artificial (IA)*
- *Internet Protocol (IP)*
- *K-Nearest Neighbors (KNN)*
- *LightGBM (LGBM)*
- *Long Short-Term Memory – Conditional Random Fields (LSTM-CRF)*
- *Long Short-Term Memory (LSTM)*
- *Machine Learning (ML)*
- *Majority Voting Ensemble Learning (VOT)*
- *Mathhews Correlation Coefficient (MCC)*
- *Maximum Entropy Markov (MEM);*
- *Naive Bayes (NB)*
- *Oversampling (OS)*

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: presidencia@iscac.pt; www.iscac.pt

- *Precision (PR)*
- *Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)*
- *Principal Component Analysis (PCA)*
- *Random Forest (RF)*
- *Random Under Sampling (RUS)*
- *Recall (RC)*
- *Recurrent Neural-Work (RNN)*
- *Registre de Commerce et des Sociétes (RCS)*
- Regressão Logística (RL)
- Revisão Sistemática da Literatura (RSL)
- *Sequence-Aware Undersampling (Seq-US)*
- *Stimulus, Capability, Opportunity, Rationalization e Ego (SCORE)*
- *Stochastic Gradient Descent (SGD);*
- *Support Vector Machine (SVM)*
- *Synthetic Minority Oversampling Technique (SMOTE)*
- *True Positive (TP)*
- *True Positives (TP)*
- *Undersampling,(US)*
- *XGBoost (XGB)*

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: presidencia@iscac.pt; www.iscac.pt

INTRODUÇÃO

O mundo atual está em constante transformação. As mudanças súbitas e por vezes radicais são cada vez mais frequentes. Ora, o crime como o conhecemos não ficou estagnado face a estas alterações. Ele evoluiu, tornou-se mais moldável, complexo e obriga diariamente a um combate mais exigente e permanente dos governos, organizações e até de cada indivíduo.

Nas organizações a história complica-se, especialmente quando falamos de fraude. Num panorama cada vez mais digital, as organizações geram e retêm cada vez mais informação em formato eletrónico. Contudo, mesmo com uma análise mais aprofundada, com mais informação disponível, procedimentos, e trabalhadores especializados, o crime da fraude persiste.

À medida que a economia mundial se move para a era digital, dentro das organizações, as instituições financeiras são daquelas que possivelmente, enfrentam um dos maiores desafios.

A indústria dos serviços financeiros providencia serviços para empresas e indivíduos, sendo a expressão "serviços financeiros" um termo lato, que varia desde os serviços bancários, investimentos, seguros, entre outros, o que faz com que a fraude possua assim, um vasto campo de aplicabilidade. Há vários pontos de acesso para esta prática criminosa, mas a Internet e a digitalização das transações tornaram este "trabalho" ainda mais fácil. Para além da fraude através do métodos tradicionais, como as caixas de multibanco, agora, a preocupação foca-se nas transações por cartões de crédito, débito, apps bancárias e *web banking*.

Num relatório publicado pela consultadora Deloitte (2021), intitulado de "*Fraud Survey Portugal 2021: Clear and focused attention*", constatamos que o setor de serviços financeiros surge como o mais propenso a situações de fraude (47%), sendo que 58% das empresas inquiridas considera que o impacto do Covid-19 promoveu o aumento de situações de fraude. Apesar da maioria das empresas considerar que tem havido um esforço e preocupação na resolução deste problema, 61% dos inquiridos considera que no

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

caso das transações fraudulentas, as empresas não possuem as ferramentas tecnológicas que permitam identificar, avaliar e mitigar os resultados dessas mesmas transações (Deloitte, 2021, p. 9 e p.70).

A escalada da tecnologia, apesar de uma mais-valia, constitui "uma faca de dois gumes". Por um lado, o crescente desenvolvimento da tecnologia permitiu às empresas novas fontes de negócios, clientes e internacionalização, por outro, os meios digitais obrigaram a um controlo mais exigente, com um maior nível de risco, e a um investimento em meios de computação capazes de detetar e mitigar práticas fraudulentas.

Apesar do contributo humano na análise, avaliação e monitorização da fraude, baseada numa enorme quantidade laboral e intervenção humana. É necessária uma abordagem em mecanismos automatizados para apoiar o sistema de deteção e mitigação de fraude. Uma menor intervenção humana, poderá levar a um sistema mais eficiente e eficaz para detetar estas práticas.

O foco da nossa dissertação será nesta preocupação emergente. Perante a necessidade de combater a fraude digital nas instituições financeiras, temos como objetivo, através de uma Revisão Sistemática da Literatura (RSL), encontrar e comparar os algoritmos de Inteligência Artificial (IA) mais usados no processo de fraude transacional, expondo criticamente, qual ou quais, melhor se adequam no contexto bancário.

Por último, o presente trabalho está estruturado em quatro capítulos: 1) Introdução e o respetivo enquadramento teórico que retrata a razão da fraude, a sua história, evolução e a necessidade de combate e mitigação, assim como uma particularidade dos *datasets* de fraude; 2) a metodologia utilizada, mais concretamente, a formulação do problema e a estratégia de pesquisa; 3) apresentação dos registos obtidos através da pesquisa, e a sua correspondente análise; 4) análise comparativa para identificação dos melhores algoritmos identificados nos registos da revisão de literatura; 5) por último, as devidas conclusões aplicadas ao contexto bancário e referentes críticas.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

1 ENQUADRAMENTO TEÓRICO

Uma boa investigação ou discussão necessita sempre de ser clara e precisa nas suas definições e tópicos a serem abordados. Começamos por definir fraude e a salientar as suas características essenciais e únicas, história, motivações humanas e sociais, e a sua permanência no decurso do tempo. De seguida, realçamos a necessidade de uma mudança de paradigma, com o foco na abordagem automatizada, cada vez mais necessária face a escalada da tecnologia. Abordando por último, o desequilíbrio dos *datasets* de fraude, e a necessidade de colmatar este problema.

1.1 A fraude e a razão humana

Segundo o dicionário de Oxford Learner's podemos definir fraude como "crime de trair alguém com intuito de obter dinheiro ou bens ilegalmente", que apesar desta definição capturar a essência da fraude, não precisa, nem descreve a natureza e as características da fraude.

A fraude não é um fenómeno recente. Inerente ao próprio ser humano, movida por vontade e desejo, a fraude como a conhecemos não parece ter um fim próximo. O consequente aumento do número de casos de fraude, a necessidade de os governos implementarem nos seus ordenamentos nacionais legislação cada vez mais restritiva e punitiva, parece não abrandar a constante escalada desta prática ilícita (Albrecht et.al., 2019, p.2-4)

Desde os tempos da Grécia antiga, relatamos casos de fraude, um dos primeiros ocorreu a 300 anos A.C, quando Hegestratos, comerciante grego, celebrou um contrato de seguro a favor do seu navio e mercadoria, que possuía duas condições: 1) reembolsar a seguradora com juros, quando a mercadoria fosse devidamente entregue; 2) em caso de perda total ou parcial do navio, Hegestratos seria restituído pelo valor do empréstimo. Este contrato era comumente conhecido como "bottomry" (Subramanian, 2014, p.1). Aproveitando as clausulas deste contrato, Hegestratos engendrou um plano com o intuito de defraudar a companhia de seguros. Desde logo, não colocaria qualquer mercadoria no navio. De seguida, tentaria afundar o navio. Neste sentido, reembolsado pela seguradora,

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

tanto quanto ao valor do navio, assim como da respetiva carga. Ainda não sendo suficiente, venderia posteriormente, a mercadoria a um terceiro comerciante (Subramanian, 2014, p.1). Infelizmente para Hegestratos, quando se preparava para afundar o navio, foi apanhado em flagrante pela tripulação do mesmo, tendo acabado por falecer afogado quando tentava fugir (Subramanian, 2014, p.1) Apesar deste caso ter ocorrido há mais de 2300 anos, a história parece repetir-se. De uma forma mais articulada, engenhosa e transversal a todas as linhas de negócio, e não apenas a seguradoras, esta prática acompanha de mãos dadas, o passado, presente e possivelmente, o futuro.

Será necessidade? Desejo? Ou algo mais? Questão esta, levantada pelo criminologista Donald Cressey, que nos anos 50, realizou um estudo com 250 reclusos ao longo de um período de 5 meses, estudando e verificando o comportamento dos mesmos (Kassem & Higson, 2019, p.191). Estipulou duas regras no seu estudo: 1) a pessoa em causa ter aceitado uma posição de confiança de boa vontade; 2) teria de ter violado essa mesma confiança. Após os 5 meses, concluiu que três fatores são necessários para haver essa quebra de confiança, desde logo: 1) um problema financeiro que não é transferível; 2) oportunidade para cometer a violação de confiança; 3) racionalização pelo infrator dessa mesma confiança. Com o passar dos anos, esta hipótese começou a ser conhecida como a Teoria da Fraude Triangular, que representa a pressão ou motivos que levam ao ato ilícito, oportunidade e racionalização (Figura 1.1). Sendo inerentes vários fatores segundo autor, seja por dívidas, fracasso pessoal, isolamento físico e mental, necessidade de status social, entre outras situações (Kassem & Higson, 2019, p. 191-192).

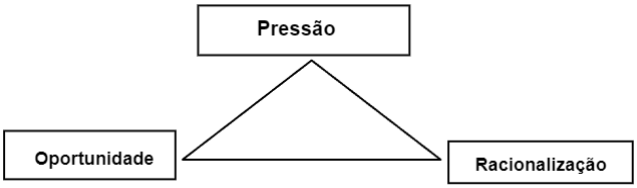

Figura 1.1 - Teoria da Fraude Triangular por Cressey´s

Fonte: Adaptado de Kassem & Higson, 2019, pp.192

Com a evolução das mentalidades e paradigmas, a Teoria Triangular, sofre uma ligeira modificação, passando a conotar-se de Teoria da Fraude de Diamante. Ora, esta teoria,

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

contrariamente à anterior, não é tripartida, adicionando um quarto lado ou um modelo há teoria já existente. Segundo *Wolfe* e *Hermanson*, fraude por si só, só pode ocorrer se a pessoa que comete tal ilícito tiver capacidade para o cometer, isto é: 1) temos de estar perante uma pessoa de autoridade; 2) temos de conhecer os sistemas contabilísticos da empresa; 3) confiança que não será apanhada; e 4) capacidade para lidar com o stress do ato praticado (Figura 1.2) (Kassem & Higson, 2019, p. 193).

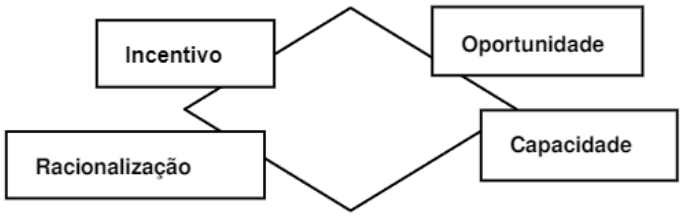

Figura 1.2 - Teoria da Fraude de Diamante por Wolfe e Hermanson

Fonte: Adaptado de Kassem & Higson, 2019, p.194

Ora, este pensamento de que não ser apanhado ou julgado pelo seus atos, para *Duffield* e *Grabosky*, representa um sentimento de ego, quase transversal a todos os tipos de fraude. Constitui um papel fundamental no porquê que as pessoas decidem cometer fraude, assim como, a forma que a cometem (Vousinas, 2019, p.376) . Com esta nova característica, o autor *Georgios L. Vousinas*, com o intuito de tentar explicar a existência de fraude e o seu crescente aumento de casos, criou o modelo *S.C.O.R.E*, que consiste no acrónimo das palavras: estímulo, capacidade, oportunidade, racionalização e ego (Figura 1.3).

Figura 1.3 - Modelo S.C.O.R.E

Fonte: Adaptado de Vousinas, 2019, p. 377

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Este modelo segundo o autor, tem uma maior aplicabilidade quando estamos perante atos praticados por um único indivíduo. Contudo, maior parte dos crimes de fraude organizacional, ocorre através de um coluio, com mais duas ou três pessoas. Este ato fraudulento é muito mais difícil de prever ou detetar, ainda para mais quando envolve empresas/organizações terceiras, que impossibilita a eficaz fiscalização deste atos. O crime organizado, constitui um dos maiores problemas na deteção, prevenção e previsão da fraude como a conhecemos (Vousinas, 2019, p.377-378). Perante esta nova realidade, o autor, afirma que no modelo surge mais uma alteração, contendo a nova característica do conluio (a sua aplicação apenas ocorre numa atuação coletiva), agora em formato de hexágono (Figura 1.4) (Vousinas, 2019, p.379).

Figura 1.4 - O alargado modelo S.C.O.R.E em formato de hexágono

Fonte: Adaptado de Vousinas, 2019, p. 379

Neste sentido, a necessidade de compreender a atuação da fraude, seja numa perspetiva individual ou coletiva, realça tópicos-chave, ou características do comportamento humano, que poderão servir de apoio teórico para possíveis modelos práticos de previsão e deteção de fraude, essenciais face à escalada deste tipo de crime.

1.2 A fraude no setor bancário na era digital

Os bancos são especialmente vulneráveis à fraude. Mas porque será? Segundo uma alegada conversa com o famoso criminoso, *Willie Sutton*, quando questionado por um repórter, "Porque roubas bancos constantemente?", ao que *Sutton* responde, "Porque é

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

onde está o dinheiro". Esta declaração praticamente sumariza o porquê de os bancos serem tão populares quando falamos de fraude (Saporta & Maraney, 2022, p. 185).

À medida que a economia converge para a era digital, os serviços financeiros enfrentam inúmeros desafios. Os serviços financeiros providenciam serviços a empresas e a indivíduos, seja na área da banca, investimento, seguros, entre outros. A aceleração do uso da tecnologia digital, exige um esforço na prestação e segurança destes serviços.

O pagamento com dinheiro, a relação pessoal entre banca e cliente, até o próprio uso de caixas de multibanco, são práticas que tendem a desaparecer. Com o rápido movimento à era digital, consumidores conseguem gerir as suas contas e finanças a partir de casa, através de aplicações ou do computador, o que aumentou com o surgimento do Covid-19 (Saporta & Maraney, 2022, p. xvii).

Cada vez mais, estas instituições apostam na tecnologia e na expansão dos seus serviços remotamente. Aliado a este fator, a evolução tecnológica, para além de permitir um crescimento substancial na maioria dos negócios, através de uma maior capacidade de internacionalização dos seus produtos e serviços, surge a fraude digital, onde o dinheiro e a oportunidade andam de mãos-dadas.

Ficaram para trás os tempos de um ato físico na fraude. A necessidade dos criminosos de roubar informações de cartões de crédito, seja num pagamento numa caixa de multibanco, ou em qualquer terminal de pagamento de uma loja, para posteriormente utilizarem em compras online, já se encontra ultrapassado. Segundo um artigo da empresa portuguesa, *Feedzai*, apesar desta prática física de fraude de cartões de crédito não ter totalmente desaparecido, o principal foco das equipas de prevenção e de combate à fraude, reside nas transações onde a componente física do cartão é inexistente. Com o aumento das interações e comércio online, mais de 83% de todos os ataques de fraude ocorrem, envolvendo cartões de crédito, débito e pré-pagos, única e simplesmente de forma online (Saporta & Maraney, 2022, p. xvi).

Atualmente existem várias formas de pagamento por cartão, que estão largamente disponíveis e constituem uma das principais formas de pagamento em vários países. Os

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

meios tecnológicos e digitais, têm ultrapassado os métodos tradicionais de pagamento, assim como a forma como nós usamos e circulamos o dinheiro. Houve assim, uma revolução nos nossos meios de pagamento, de um estado físico para uma transformação digital. A paisagem da política monetária que outrora existiu, sofreu uma drástica alteração (Seera *et al.*,2021, p.1-2).

Porém, nem tudo é mau. Devido ao aumento de utilizadores dos meios tecnológicos e a digitalização das transações, cada passo dado pelo cliente é um ponto de informação, ou seja, traduz-se em dados. Estes pontos de acesso, permitem ao banco obter, analisar e providenciar um melhor serviço ao cliente, e o mesmo ocorre na segurança e combate da fraude. A informação é chave nesta luta diária, que nos dias que correm, não apenas a fraude se tornou muito mais frequente, mas mais volátil, inconstante, e consequentemente, de difícil deteção.

1.3 A necessidade de uma nova estratégia

Nós apenas conseguimos observar indicadores, sintomas ou avisos de fraude. Uma vez detetados, eles devem ser investigados para determinar se estamos perante uma verdadeira fraude ou não. No caso de uma grande quantidade de informação como é o caso de um banco, essa investigação pode tornar-se complicada e morosa. É necessário estabelecer a melhor forma de mitigar e realçar os verdadeiros avisos de fraude para ser possível dar a atenção devida pelas equipas de combate à fraude.

Ora, a banca costuma ter acesso às mais e melhores ferramentas e programas informáticos de mitigação, o que permite uma melhor proteção das instituições e dos seus clientes, contra possíveis vulnerabilidades. Mas só isto não basta. É necessária uma ligação entre os vários departamentos internos do banco e as fontes externas de informação (por exemplo: *DowJones*; *Registre de Commerce et des Sociétes* (RCS)). Esta massificação da informação é o que possibilita a esquematização de padrões de comportamento dos praticantes da fraude, prevalecendo o combate à sua deteção e mitigação. A disponibilidade da informação a partir de várias fontes, seja na rede local bancária (departamentos internos), transações online, redes sociais, *Internet Protocol* (IP),

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

telecomunicações, *etc.*, contribui para a eficácia no combate à fraude (Saporta & Maraney, 2022, p. 183-184).

Mas como podemos processar esta quantidade massiva de dados? E qual será a forma mais eficiente? Esta última questão, constitui o cerne deste trabalho. Várias instituições financeiras utilizam diferentes métodos de deteção e mitigação de fraude, uns como iremos ver, mais utilizados que outros.

Com tendência a desaparecer, uma das técnicas mais utilizadas consiste num modelo de deteção de fraude baseada no conhecimento, avaliação e monitorização humana. Esta abordagem clássica, da utilização do conhecimento humano para determinar o que é fraude ou não, apesar de ultrapassada, é uma das principais ferramentas ainda utilizadas e representa um bom ponto de começo para as instituições financeiras, incorporarem meios e sistemas de deteção e mitigação de fraude. Contudo, o nosso foco será na perspetiva futura. Mais concretamente na tendência crescente do uso da IA neste combate.

Uma abordagem automatizada para manter um sistema de deteção de fraude é sempre mais favorável, nem que seja, por necessitar menos esforço humano, o que por si só poderá levar a uma maior eficiência e eficácia. Neste sentido, podemos resumir em três as razões o interesse e vantagens desta nova metodologia automatizada, afastada da intervenção humana (Baesens, *et al.*, 2015, p. 17-19):

- **Precisão:** Modelos estatísticos de deteção de fraude, oferecem uma maior capacidade de deteção que a metodologia clássica. Desde logo, através do processamento de grandes quantidades de dados de informação, padrões de fraude que antes não eram visíveis ao olho humano, tornam-se aparentes. Por outro lado, o foco é diminuir a interação humana, ou seja, utilizar as equipas de prevenção de fraude apenas para os "verdadeiros" casos de fraude, e não sobrecarregar os limitados recursos humanos com casos falsos positivos.
- **Eficácia operacional:** Outro ponto de vista consiste na rapidez da mitigação e libertação transacional. Ora, quando avaliamos uma transação de um cartão de crédito, uma decisão quase imediata é necessária, seja para bloquear ou libertar a respetiva transação. A possibilidade de bloquear determinadas transações como

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

fraudulentas, quando não o são, comporta um risco concorrencial e uma possível quebra na carteira de clientes, incomportável num mercado cada vez mais competitivo.

- **Custos:** Manter e desenvolver um sistema de deteção de fraude com mero conhecimento e experiência humana, torna-se bastante exigente. A volatilidade da fraude, faz com que haja a necessidade de alterar constantemente o sistema, porque o que era até aquela altura conhecido, deixou de o ser. Logo, um sistema automatizado, é neste ponto de vista, uma mais-valia.

Apesar das abordagens de fraude terem evoluído e avançado ao longo dos últimos anos, a fraude continua a ser difícil de detetar. A complexidade deste problema baseia-se na sua volatilidade. Não apenas os mecanismos de deteção de fraude ficam mais eficazes, mas os próprios criminosos adaptam formas mais complexas e inventivas de ultrapassar estes mesmos mecanismos. Como tal, é difícil de determinar qual a melhor solução para detetar a fraude. Será meramente automatizada a melhor forma? Humana? Híbrida? Depende da análise em concreto do negócio e atividade em causa. Teremos de tratar igual o que é igual e diferente o que é diferente.

1.4 As soluções de prevenção de fraude e a questão controversa

Esta é uma questão complexa. Depende de cada modelo de negócio e sua correspondente finalidade. É difícil determinar qual a "melhor" forma de prevenir a fraude ou ferramentas para o efeito. Todas as soluções possuem vantagens e desvantagens.

Independentemente da solução, lidar com o problema da fraude é sempre bastante dispendioso. Seja por meios meramente humanos, tecnológicos ou híbridos, há sempre custos e riscos associados.

Há sempre a necessidade de realizar uma avaliação de risco e determinar se os benefícios são superiores aos eventuais custos. Alguns riscos podem ser desde logo assumidos, com a implementação de ferramentas de controlo mais eficazes, sabendo desde logo que os custos inerentes a esta implementação serão superiores às perdas previstas ou presumidas. Por exemplo, um banco ao emitir cartões de crédito, pode reduzir o número de transações

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

fraudulentas se forem implementadas ferramentas de segurança avançadas, mas terá um custo monetário avultado associado ou, até criar constrangimentos aos clientes que poderão levar à saída dos mesmos, ultrapassando os custos tidos com as transações fraudulentas até então (Gee, 2015, p.4). Outro exemplo, é o bloqueio por parte do banco de transações legitimas que aparentam ser suspeitas para reduzir o número de perdas por fraude, que também irá criar constrangimentos aos clientes (Isson, 2018, p. 161-162).

Os bancos possuem um papel fundamental na economia mundial, seja na atribuição de crédito, a oferta de vários produtos financeiros, e até o escrutínio e estabilidade de um sistema de pagamentos funcional. Contudo a exigência regulamentar, também surge da relação de interesses que possui com os respetivos governos. Por um lado, a benesse dos depósitos bancários, contrapõe com o seu papel de mutuante em caso de crises governamentais. Há assim a necessidade de uma pressão regulamentar, que apesar de ser intrusiva e intensa nos vários modelos negociais da banca, aparenta ser necessário para salvaguardar interesses de bem maior (Dill, 2020, p. 16-17).

Neste sentido, com a pressão dos clientes, reguladores e a necessidade de redução de custos, o processo de deteção e mitigação da fraude tem de ser eficaz e célere e não colocar em causa o normal funcionamento do sistema de pagamentos e operações. Como já referido anteriormente, maior parte das instituições, incluindo a banca, desconstrói este processo em tarefas automatizadas e humanas (híbridas), reduzindo em quatro principais focos de ação (Figura 1.5): 1) alertar as operações/transações que constituem um risco de fraude; 2) a investigação desses alertas; 3) ação em relação aos suspeitos de fraude; 4) monitorização de fraude *à posteriori* (Nesvijevskaia, 2021, p. 3-4).

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

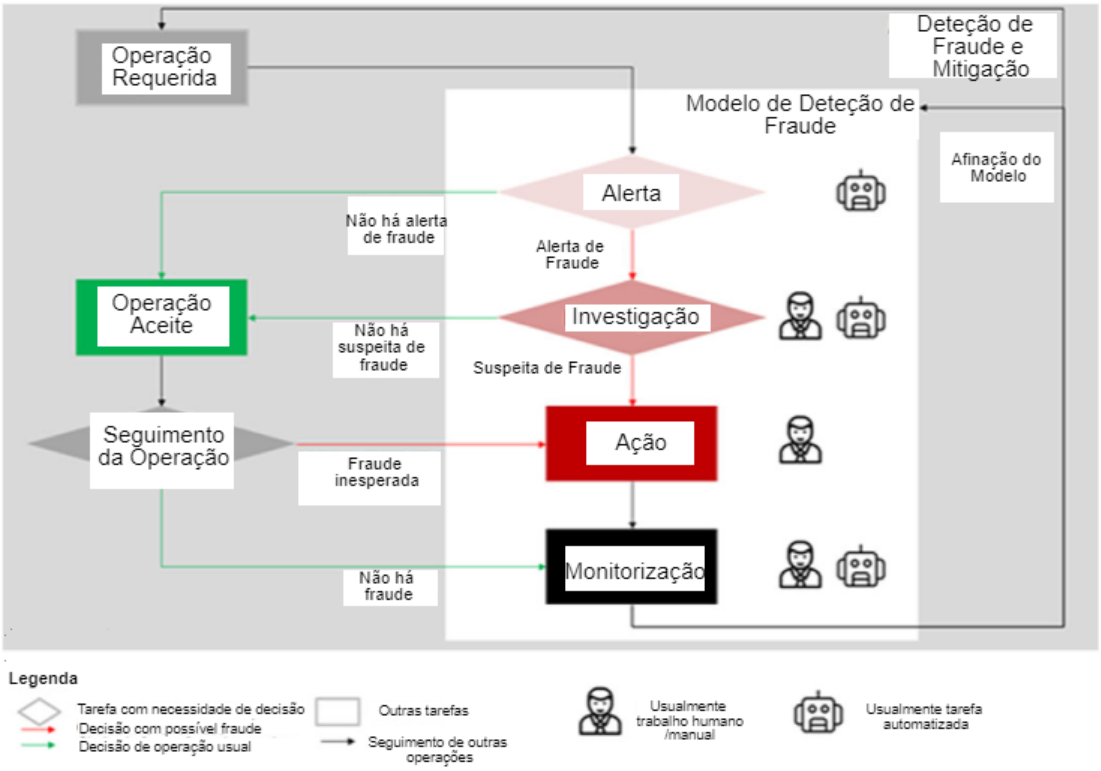

Figura 1.5 - Processo simplificado de gestão de fraude

Fonte: Adaptado de Nesvijevskaia, 2021, p. 4

Independentemente do caso particular da banca, existem soluções de deteção de fraude de cariz unicamente tecnológico, humano, e até uma junção das duas vertentes, adotando para melhor entendimento, a estrutura apresentada pelos autores Saporta & Maraney (2022, p. 56-58):

- **Sistema de Regras:** Consiste num sistema já bastante utilizado na área financeira, em que são utilizadas regras para avaliar as transações. Por exemplo, podemos determinar que as transações de valor superior a 250 euros necessitam de ser revistas manualmente por parte da equipa de prevenção de fraude. O mesmo poderá ser aplicado não quanto ao valor, mas sim à área geográfica da devida transação, ou da atividade proveniente da mesma, *etc*. São várias as regras que podemos estipular, assim como as características adjacentes a cada uma delas. O aspeto negativo deste tipo de sistema, é a generalização da regra. Por exemplo o caso de um país como a Rússia, será que queremos bloquear todas as transações

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

provenientes deste país? Mesmo estipulando uma outra regra, seria na mesma um número avultado de transações. Não obstante, este tipo de sistema é bastante fácil de trabalhar e gerir.

- ***Machine Learning (ML):*** Especialmente nos últimos 5 anos tem sido um tema bastante debatido, e o conceito é bastante simples: máquinas que consigam ser treinadas para reconhecer transações ou determinada atividade, como fraudulenta ou legitima, tendo como base, exemplos passados. Uma das principais vantagens deste sistema consiste na sua capacidade de adaptarem à volatilidade da fraude, ou às novas invenções e ideias criadas por estes agentes fraudulentos. Ao contrário do sistema de regras, o ML consegue mais facilmente detetar padrões e avaliar de uma mais forma aprofundada o que deve ser considerado como fraude ou não. Por outro lado, estes sistemas podem conter aquilo que é chamado do efeito "*black box*", em que algumas vezes é difícil perceber o porquê de determinadas decisões ou quais os fatores que tiveram em consideração quando tomaram essa mesma decisão. Isto faz com que seja difícil alterar o sistema, quando este mesmo comete erros.
- **Modelo Híbrido:** Os modelos híbridos combinam sistemas de ML com o sistema de regras. Pode começar com um sistema de regras que posteriormente foram adicionando ML para um propósito específico, ou até o contrário. Segundo os autores, este é um dos modelos mais utilizados atualmente pelas várias instituições financeiras, já que combina as vantagens de ambos sistemas de regras e ML.

1.5 Desequilíbrio dos dados nos *datasets* de fraude

Aquando da escolha de um modelo pela instituição financeira, especialmente quando esta opta por modelos híbridos ou de ML, há uma particularidade necessária a ter em conta: o desequilíbrio dos *datasets* de fraude. Diz-se que um *dataset* é desequilibrado quando há uma discrepância significativa entre o número de exemplos das diferentes classes, neste caso a classe positiva (transações legítimas) e a classe negativa (transações fraudulentas).

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Ora, o principal problema desta discrepância de classes consiste na falta de representação da classe minoritária ou negativa, sendo esta usualmente a classe alvo de estudo ou escrutínio. Segundo um exemplo realçado por Alberto Fernández *et al.* (2022): "imaginem que estamos a lidar com uma aplicação médica, em que temos de diferenciar entre os tumores malignos e benignos de um tipo específico de cancro, que foram estudados após uma biopsia. Neste caso em concreto, seria muito mais importante estudar as consequências dos tumores malignos que os benignos, já que os malignos podem ser fatais (…) A verdade é que queremos 100% de *Accuracy* em ambas as classes (…) porém a verdade é que a os classificadores tendem a obter uma excelente *Accuracy* na classe maioritária e resultados péssimos na classe minoritária." (Fernández *et al.*, 2022, p. 20-21).

Numa aplicação ao mundo da fraude, esta discrepância pode influenciar os resultados e consequente a performance dos algoritmos de ML. Neste sentido, existe uma necessidade de suprir e equilibrar estas classes, ou encontrar algoritmos que trabalhem bem com *datasets* desequilibrados, já que o nosso objetivo é detetar as transações fraudulentas.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

2 METODOLOGIA

Para a realização do presente trabalho, utilizou-se a metodologia da Revisão Sistemática de Literatura (RSL), que segundo o guia da Charles Sturt University (CSU, 2023) "a revisão sistemática de literatura identifica, seleciona e avalia criticamente literatura, com intuito de responder a uma pergunta claramente formulada. A revisão sistemática deve seguir um protocolo ou plano claramente definido, onde os critérios são claramente declarados antes da realização da revisão". Logo, consiste em determinar uma estratégia de pesquisa ponderada, que tenha como um foco específico ou uma pergunta definida.

Para documentar as evidências encontradas ao longo da pesquisa de forma clara e transparente, guiamo-nos pela diretriz para a revisão sistemática de literatura designada por: *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) (Moher, D., 2009).

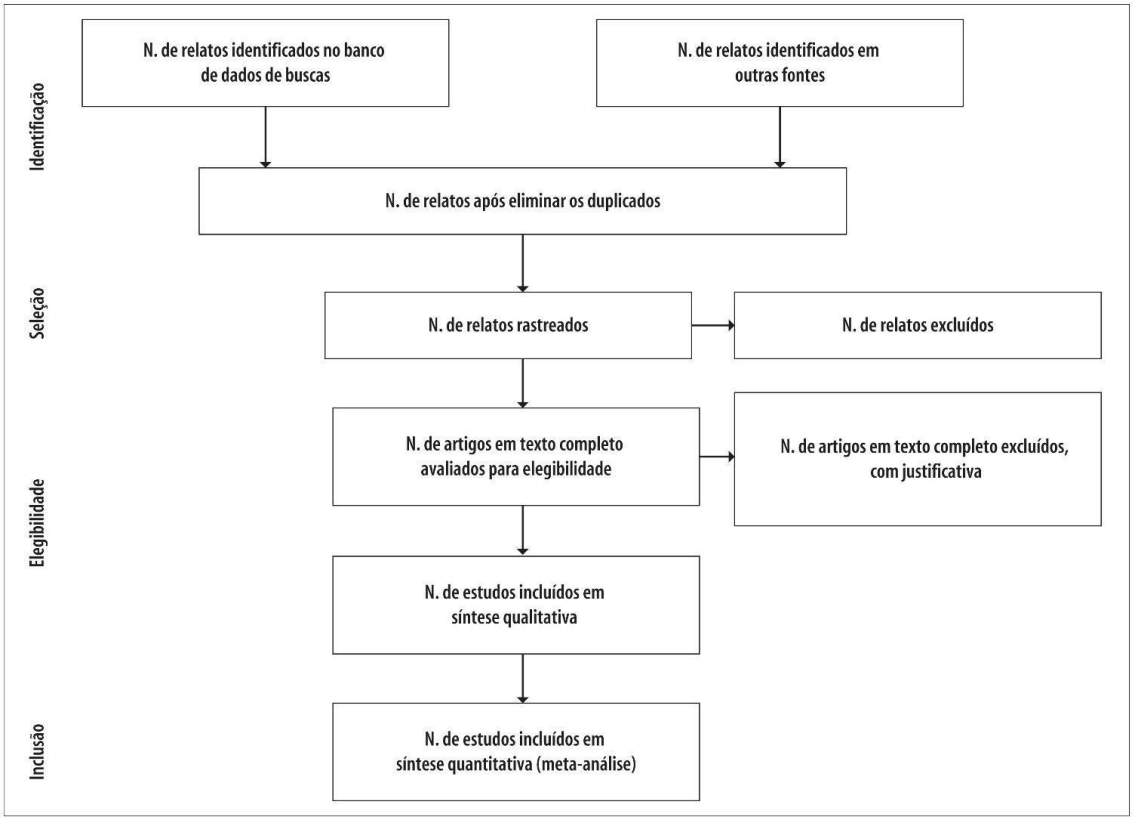

Figura 2.1 - Fluxo da informação com as diferentes fases de uma revisão sistemática de literatura

Fonte: Adaptado de Moher, D. et al., 2009

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Esta diretriz fornece um diagrama de fluxo (Figura 2.1), assim como as fases mais importantes da RSL, seleciona e identifica qual a literatura mais importante, fazendo posteriormente uma triagem ou inclusão na presente dissertação.

Possibilita o mapeamento das publicações/literatura elegível e excluída em cada fase do processo de elaboração da revisão sistemática, demonstrando visualmente, todo este encadeamento (Kuhn, 2022, p. 22).

2.1 Formulação do problema

Como já referido anteriormente, toda a nossa pesquisa/investigação, irá incidir na resposta ou pelo menos na tentativa de tal, de qual ou quais os melhores algoritmos de deteção de fraude transacional, com uma posterior aplicação ao contexto bancário. Neste sentido, iremos investigar e debruçar sobre os artigos que, no nosso entender, elucidam esta questão controversa.

2.2 Estratégia de pesquisa

A estratégia de pesquisa foi baseada no uso de algumas palavras-chave em inglês, mais concretamente: "*fraud"*, "*artificial intelligence"* e "*transactions"*, para artigos. Relativamente à palavra "*artificial intelligence*", podíamos ter de alterar esta palavra-chave de pesquisa para termos com um significado semelhante, como: *Machine Learning* e/ou *algorithms.* O mesmo processo foi aplicado à palavra-chave "*transactions*", que podíamos ter de alterar para "*banking*" e/ou "*financial institutions*"

Realizamos para o efeito duas pesquisas, aplicando os operadores lógicos OR e AND do inglês. Numa primeira pesquisa, em todo o corpo do texto, em que fixamos como hiato temporal de 2017 a 2022, e as palavras-chave anteriormente definidas (Figura 2.2).

```
YEAR >= 2017 =< 2022 AND FULL DATA (FRAUD AND (ALGORITHIMS OR MACHINE LEARNING OR
ARTIFICIAL INTELLIGENCE) AND (TRANSACTIONS OR BANKING OR FINANCIAL INSTITUTIONS))
```

Figura 2.2 - Fixação de termos de pesquisa em todo o corpo de texto

Fonte: Elaboração própria

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Numa segunda pesquisa, em apenas no campo do título e do *abstract* , em que fixamos um hiato temporal de 2017 a 2022, e as palavras-chave anteriormente definidas (Figura 2.3).

YEAR >= 2017 =< 2022 AND TITLE AND ABSTRACT (FRAUD AND (ALGORITHIMS OR MACHINE LEARNING OR ARTIFICIAL INTELLIGENCE) AND (TRANSACTIONS OR BANKING OR FINANCIAL INSTITUTIONS))

Figura 2.3 - Fixação de termos de pesquisa no título e abstract

Fonte: Elaboração própria

2.3 Fontes de pesquisa e sua aplicação

A pesquisa foi realizada entre o início do mês de outubro de 2022 e meados do mês de fevereiro de 2023, através de duas plataformas distintas: *Dimensions.ai* e o *B-on*.

O *Dimensions.ai* consiste numa plataforma que permite pesquisar e analisar várias fontes de informação numa única plataforma. Permite encontrar publicações de ensaios clínicos, patentes, documentação legal, artigos, etc, visualizar e analisar esta quantidade massiva de dados e traçar relações entre estes mesmos dados (Dimensions.ai, 2023).

No mesmo sentido, o *B-on* consiste numa plataforma de pesquisa de informação científica, que tem como intuito disponibilizar "o acesso ilimitado e permanente às instituições de investigação e do ensino superior aos textos integrais de milhares periódicos científicos e *ebooks* online de alguns dos mais importantes fornecedores de conteúdos, através de assinaturas negociadas a nível nacional (…) dando acesso a milhares de publicações cientificas e é hoje uma referência no acesso à informação cientifica internacional" (B-on, 2023).

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Ora, aplicando as diferentes palavras-chave em cada uma das plataformas, obtemos os seguintes resultados[1]:

1) **No caso da plataforma *Dimensions*:**

Quando procuramos única e exclusivamente pelas palavras-chave e critérios mencionados anteriormente, verificamos uma quantidade avultada de resultados (44,667 mil resultados) (Figura 2.4). Para reduzir o número de resultados, para além de filtrar pela data, aplicamos um outro critério, que consiste em filtrar a pesquisa das palavras-chave apenas no título e *abstract* dos artigos e informação em causa (116 resultados) (Figura 2.5).

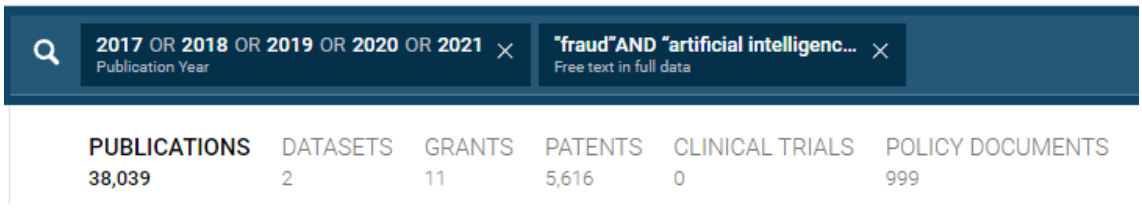

Figura 2.4 - Pesquisa das palavras-chave na plataforma Dimensions com filtro apenas na data

Fonte: Dimensions.ai (2023)

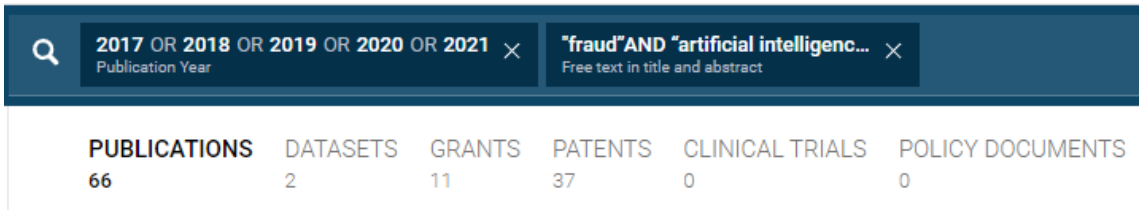

Figura 2.5 - Pesquisa das palavras-chave na plataforma Dimensions com filtro do título e abstract

Fonte: Dimensions.ai (2023)

[1] Pesquisa realizada a 15/01/2023

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

2) No caso da plataforma *B-on*:

O mesmo problema ocorre nesta plataforma (59,060 mil resultados), quando utilizamos as referidas palavras-chave e filtramos pela data (Figura 2.6). Será necessário reduzir o número de resultados, retirando o filtro pré-selecionado de "aplicar a palavras relacionadas" e "pesquisar também no texto integral dos artigos", concentrando e convergindo melhor a pesquisa (379 resultados) (Figura 2.7).

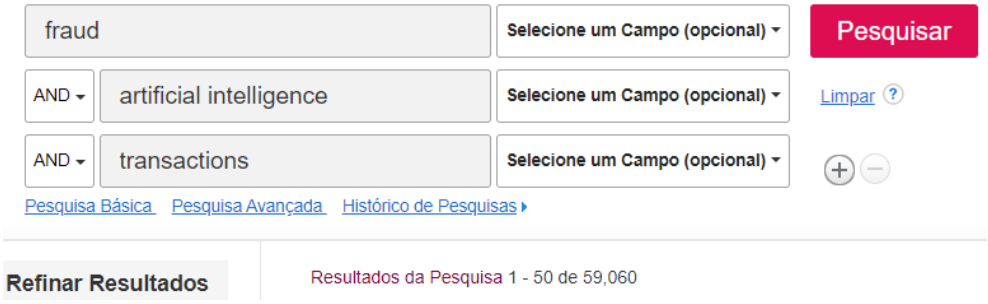

Figura 2.6 - Pesquisa das palavras-chave na plataforma B-on com filtro apenas na data

Fonte: B-on (2023)

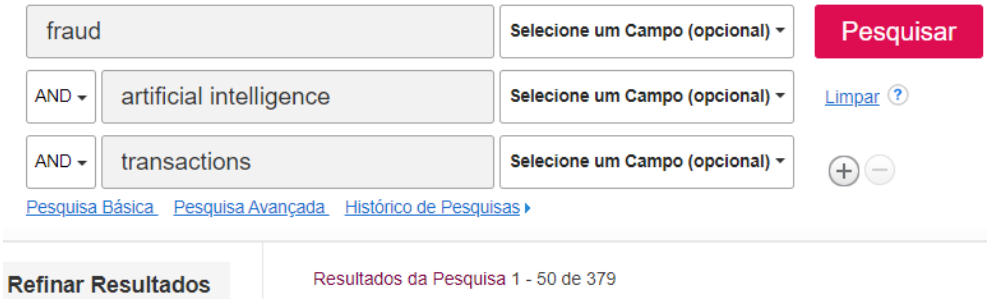

Figura 2.7 - Pesquisa das palavras-chave na plataforma B-on com diferentes filtros

Fonte: B-on (2023)

Necessário relembrar, que os valores dos resultados obtidos em ambas as plataformas podem sofrer alterações, seja devido à data em que ocorra a pesquisa, seja pela utilização de termos semelhantes às palavras-chave referidas.

2.4 Critérios de elegibilidade

Para esta revisão, foram escolhidos os artigos que descrevem os seus estudos e resultados obtidos, com abordagens técnicas, que tenham como objetivo comparar estes mesmos

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

resultados e diferenciar qual ou quais o(s) melhores algoritmos de deteção de fraude transacional.

2.4.1 Critérios de inclusão

Dada a importância deste assunto, e o consequente surgimento de artigos sobre esta matéria, há uma necessidade de diferenciar e selecionar aqueles que provém de artigos de revistas científicas, ou quando não o são, identifica-se o contributo científico dos mesmos, seja por uma escrita bem estruturada, sistematizada, e que compara vários algoritmos, justificando devidamente os seus resultados, ou, empresas e instituições desta área de atuação que publicam artigos e newsletters relevantes (por exemplo: *Feedzai*).

A pesquisa compreendeu apenas artigos em idioma inglês, uma vez que é o idioma predominante das publicações em análise.

2.4.2 Critérios de exclusão

Foram excluídos títulos que não estavam disponíveis nas bases de dados da *Dimensions* e *B-on*, e aqueles que, dentro destas mesmas plataformas, eram referentes a pesquisas secundárias, que não se focavam no tópico em questão e/ou sem uma verdadeira contribuição de relevo.

Durante a fase de triagem dos títulos relevantes, foi estipulada uma regra para restringir o número de títulos que seriam alvo de análise e comparação. Ora, ditamos que apenas aqueles que fizessem uma comparação entre 2 ou mais algoritmos de deteção de fraude e que utilizassem métricas de classificação como por exemplo o *F1-Score*, é que poderiam fazer parte desta seleção (Tabela 2.1).

Neste sentido, conseguimos reduzir drasticamente o número de títulos de ambas as plataformas, assim como facilitar uma comparação dos resultados entre os diferentes títulos.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Tabela 2.1 - Exemplo do Output da Regra em causa

Model Name	Accuracy	Precision	Recall	F1_score	Time in Sec
Decision Tree Classifier	99.958%	99.945%	99.971%	99.958%	67
MLP Regressor	99.389%	99.390%	99.389%	99.389%	1244
Random Forest Classifier	99.950%	99.950%	99.950%	99.950%	1550
Complement NB	78.753%	70.222%	99.799%	82.438%	10

Fonte: Megdad et al., 2022, p. 37

2.5 Diagrama de Fluxo

Os detalhes de cada fase estão apresentados na Figura 2.8, seguindo o anteriormente estipulado no subtópico 2.2, quanto às bases de dados de pesquisa, palavras-chave e regras.

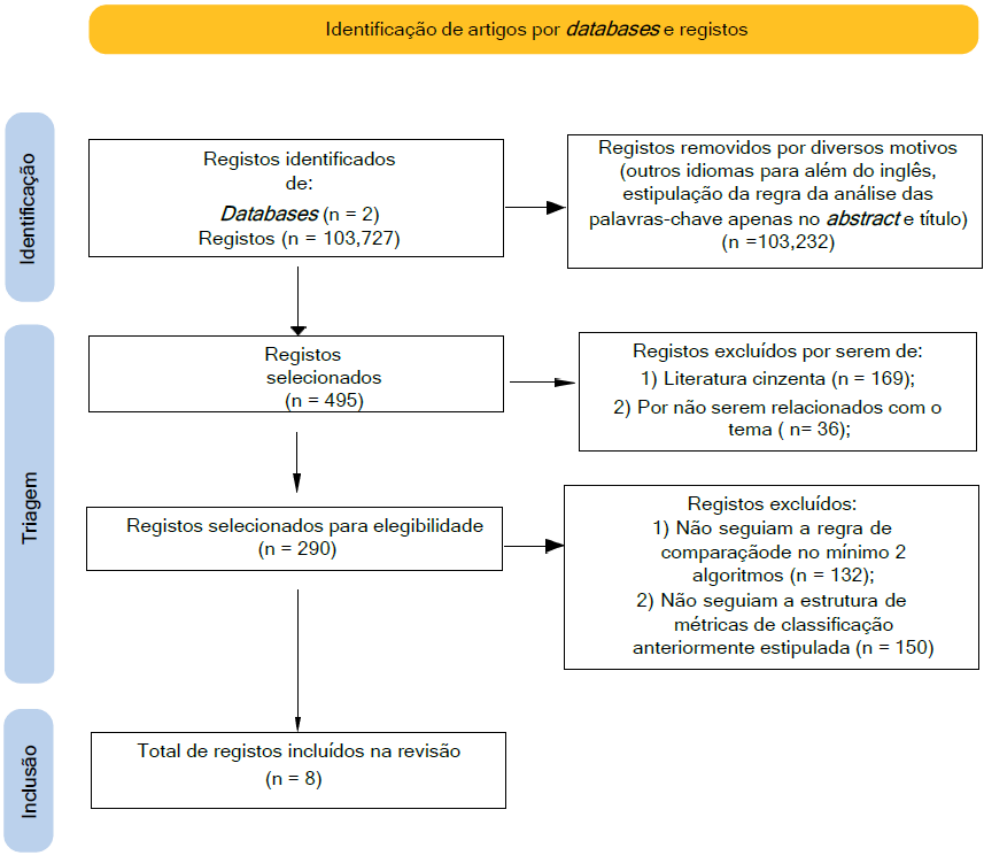

Figura 2.8 - Diagrama de fluxo PRISMA

Fonte: Elaboração própria

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

A fase de identificação resultou em 103,727 registos provenientes de 2 fontes de informação distintas (*B-on* e *Dimensions.ai*), dentro dos quais, 103,232 registos foram excluídos por possuírem idiomas diferentes do inglês e/ou pela restrição da pesquisa das palavras-chaves apenas no título e *abstract* (o que reduziu consideravelmente os resultados).

Na fase de triagem, observamos que do processo de exclusão anterior, obtivemos um total de 495 registos, sendo que destes, foram excluídos 169 registos por serem de literatura cinzenta e/ou de fontes dúbias, assim como 36 registos por não estarem integralmente relacionados com o tema, apenas fazendo uma "ligeira" abordagem aos tópicos suscitados. Ora, num total de 198 registos excluídos anteriormente, perfazendo assim 290 registos que possuíram características de elegibilidade, foram também excluídos deste número, aqueles que, não seguiam as regras cumulativas anteriormente estipuladas, como: 1) a necessidade de 2 ou mais algoritmos de deteção de fraude, para possibilitar uma comparação entre os mesmos (n = 132); 2) a necessidade de uma estrutura de métricas de classificação como por exemplo (*F1-score*, *Accuracy*, *Recall*, *etc*) (n = 150).

Face ao exposto, totalizamos assim, 8 registos finais de revisão na fase de inclusão, que irão ser alvo de análise e devida interpretação ao longo da presente dissertação.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

3 RESULTADOS

Com o intuito de estruturar e esquematizar os 8 artigos finais que iremos analisar ao longo desta dissertação, decidimos dividir por subtópicos cada um destes registos, verificando a questão controversa, dados a serem analisados, algoritmos invocados, comparações e conclusões, em cada um destes mesmos, e por fim sumarizar os resultados destes 8 registos num sumário de resultados, com intuito de responder à nossa questão de investigação identificada no capítulo anterior: *Qual ou quais os melhores algoritmos de deteção de fraude transacional?*

Para tornar mais fácil a leitura do documento foram colocadas em anexo as descrições dos algoritmos identificados (Apêndice 1), bem como as principais métricas (Apêndice 2) e técnicas de equilíbrio de dados (Apêndice 3).

3.1 *Exploratory analysis and implementation of machine learning techniques for predictive assessment of fraud in banking systems*

O presente artigo consiste num trabalho realizado por Lellis, *et al.* (2022) com intuito de atender à problemática da fraude financeira nos sistemas bancários, através da análise de algoritmos de ML.

Os autores decidiram estruturar o presente artigo em: 1) uma análise exploratória com intuito de clarificar as variáveis que estão a influenciar este processo de avaliação dos algoritmos de deteção de fraude; 2) retratar a questão dos *datasets* não serem equilibrados utilizando técnicas como *Random Under Sampling* (RUS), *Synthetic Minority Oversampling Technique* (SMOTE) e *Adaptive Synthetic* (ADASYN); 3) posteriormente classificar e distinguir os casos de fraude, através de algoritmos como a Regressão Logística (RL); *Naive Bayes* (NB); *K-Nearest Neighbors* (KNN) e técnicas *Percepton*; 4) por último apresentar os devidos resultados e considerações, tendo em conta cada um dos cenários possíveis.

Com o objetivo de clarificar e possivelmente servir de modelo de comparação ao sistema bancário, no caso em estudo, os autores, utilizaram um *dataset* com características de *Big*

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Data, possuindo mais de 6 milhões de transações de um banco internacional (não foi indicado pelos autores a instituição financeira em questão).

O *dataset* possui 11 variáveis, desde a variável do montante até à variável que determina em código binário se estamos perante uma situação de fraude ou não (Tabela 3.1).

Tabela 3.1 - Análise descritiva do dataset e respetivas variáveis

	Step	Amount	OldbalanceOrg	newbalanceOrig	oldbalanceDest	newbalanceDest	isFraud	isFlaggedFraud
mean	243.4	179861.9	833883.1	855113.7	1100701.7	1224996.4	0	0
Std	142.3	603858.2	2888242.7	2924048.5	3399180.1	3674128.9	0	0
Min	1.0	0	0	0	0	0	0	0
25%	156.0	13389.6	0	0	0	0	0	0
50%	239.0	74871.9	14208.0	0	132705.7	214661.4	0	0
75%	335.0	208721.5	107315.2	144258.4	943036.7	1111909.2	0	0
max	743.0	92445516.6	59585040.4	49585040.4	356015889.4	356179278.9	1	1

Fonte: Lellis et al., 2022, p. 120

Não obstante, os autores afirmam que de 6.362.620 transações, 8213 transações consistem em casos de fraude, representando um total de 0.13%. Ora apesar da percentagem parecer ínfima, é de salientar que na Tabela 3.1, a média do valor das transações ronda os 179 mil dólares, e o valor máximo perto de um trilião de dólares. Como tal, 0.13% em número de transações, ou 1,05% quando comparando com o valor monetário, constitui uma perda em fraude de 12 biliões de dólares. Além disso, a distribuição monetária das transações que são alvo de fraude, quando comparada com as transações regulares ou normais, é bastante distinta (Figura 3.1), com transações regulares a variar entre os 0 e 250 mil dólares, e as transações fraudulentas a variar entre os 150 mil e 1.5 milhões de dólares.

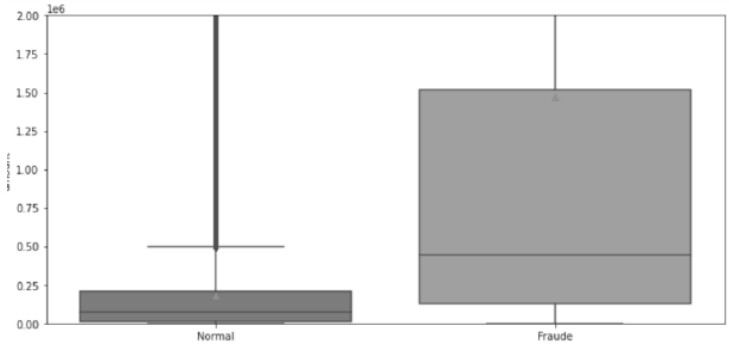

Figura 3.1 - Concentração do valores relativos a transações regulares e transações fraudulentas

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Fonte: Lellis et al., 2022, p. 120

Devido a esta variação e não equilíbrio no registo de transações fraudulentas e não fraudulentas, foram utilizadas várias técnicas de equilíbrio dos dados, tendo em consideração 70% do *dataset* original para conjunto treino, e 30% para conjunto teste (no apêndice 3 é possível ver um resumo das técnicas de equilíbrio de *datasets* encontradas nos artigos analisados):

- *Random Under Sampling* (RUS): esta técnica descarta aleatoriamente conjuntos da classe maioritária (neste caso dos casos de não fraude), até atingir uma distribuição equilibrada no conjunto treino.
- *Synthetic Minority Oversampling Technique* (SMOTE)*:* consiste numa técnica que gera novos dados sintéticos, que não sobrepõe aos dados já existentes, mas aproxima-se dos pontos minoritários no espaço dimensional, com intuito de equilibrar a distribuição no conjunto treino.
- *Adaptive Synthetic* (ADASYN): esta técnica gere dados sintéticos, em que não sobrepõe aos dados já existentes, contudo quando se aproxima dos pontos minoritários no espaço dimensional, aproxima-se daqueles que são difíceis de "aprender", afastando-se dos mais fáceis para este modelo.

Após o processo de equilíbrio, foram implementados quatro algoritmos de ML, com intuito de testar a efetividade para detetar e mitigar as transações fraudulentas. Os algoritmos utilizados foram:

- Regressão Logística (RL):
- *K-Nearest Neighbours* (KNN)
- *Naive Bayes* (NB)
- *Perceptron*

Com a clarificação destes algoritmos, e com o intuito de entender qual o modelo mais favorável, os autores utilizaram a AUC (área debaixo da curva ROC ou *Area Under the Receiver Operating Characteristic Curve*), sendo uma das métricas mais importantes para classificar a performance de um determinado modelo. O AUC-ROC, demonstra o quanto

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

o modelo é capaz de distinguir entre classes. Quanto maior for o AUC, melhor será o modelo a prever classes de valores binários [0,1]. Não obstante, os autores decidiram também adicionar a *Accuracy* como métrica de classificação, sendo que esta consiste no número de verdadeiros positivos mais os verdadeiros negativos, a dividir pelo total de amostras do *dataset*. A Tabela 3.2 apresenta os resultados obtidos pelos autores com as diferentes configurações.

Tabela 3.2 - Análise descritiva dos algoritmos e classificações

ML Model	Dataset Balancing	Accuracy	AUC	Number of False Positives	Number of False Negatives
Logistic Regression	Unbalanced	0.998	0.713	1877	1413
	RUS	0.908	0.908	174634	227
	SMOTE	0.912	0.910	168445	226
	ADASYN	0.836	0.913	312104	25
Naive Bayes	Unbalanced	0.992	0.584	13008	2031
	RUS	0.964	0.719	67465	1298
	SMOTE	0.963	0.721	68390	1288
	ADASYN	0.132	0.423	165922	701
KNN	Unbalanced	0.999	0.840	253	789
	RUS	0.941	0.952	12797	93
	SMOTE	0.995	0.945	8595	261
	ADASYN	0.995	0.945	8978	266
Perceptron	Unbalanced	0.992	0.883	15355	559
	RUS	0.687	0.838	597473	27
	SMOTE	0.650	0.819	668006	29
	ADASYN	0.866	0.919	256103	69

Fonte: Lellis et al., 2022, p. 122

1) Após implementar os modelos de ML e obtiver as respetivas métricas de performance, os autores conseguiram chegar às seguintes conclusões (Tabela 3.2): Com a avaliação das métricas de classificação, AC e AUC, o algoritmo da RL e o algoritmo KNN, apresentam a melhor performance, especialmente quando estamos perante dados equilibrados.

2) O algoritmo *NB* foi o menos eficaz quando testada a sua performance, seja em dados equilibrados ou não.

Apesar das conclusões, o estudo realizado por estes autores teve como intuito não uma análise exaustiva, mas sim uma aproximação da integração de algoritmos de ML em sistemas bancários, assim como, as mais-valias que representam na mitigação e deteção de fraude transacional.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

3.2 *Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost*

O presente artigo dos autores Ileberi, *et al.* (2021), surge com a preocupação do avanço galopante das tecnologias e aplicações nas áreas do e-commerce e financeira, aumentando consequentemente o número de transações *online*, assim como a fraude de cartões de crédito, sendo o principal alvo, empresas, comerciantes e bancos. Neste sentido, os autores, através deste estudo tentam implementar algoritmos de ML para a deteção da fraude em cartões de crédito, usando transações reais provenientes da Europa, num *dataset* não equilibrado.

Relativamente ao *dataset* em causa, os autores adotaram uma base de dados com 284807 mil transações reais, em que 99.828% destas são legítimas e 0.172% são fraudulentas. Adicionalmente, este *dataset* contem 30 variáveis, em que 28 destas encontram-se anonimizadas (V1, …, V28), e duas não (*Time* e *Amount*), assim como uma coluna (*Class*), que determina em código binário, em que 0 representa uma transação legítima, e 1 uma transação fraudulenta.

Para resolver a particularidade dos dados não equilibrados, os autores adotaram a técnica SMOTE.

Em termos de algoritmos de ML, que foram alvo de estudo e comparação, os autores adotaram os seguintes:

- Etapa 1:
  - *Support Vector Machine* (SVM)
  - Regressão Logística (RL)
  - *Decision Tree/Árvore de Decisão* (DT/AD)
  - *Random Forest* (RF)
  - *Extreme Gradient Boosting* (XGBoost)
  - *Extra Tree* (ET)
- Etapa 2:
  - *AdaBoost*

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Para além dos algoritmos da Etapa 1, foi também adotado um outro algoritmo de ML, agora na Etapa 2, mais conhecido por *AdaBoost*. Este último, tem como finalidade criar modelos altamente precisos, através da combinação de outros modelos mais simples ou menos eficazes/precisos. O intuito desta divisão de etapas, e consequentemente de algoritmos, foi de demonstrar as possíveis vantagens de utilizar ou não o algoritmo AdaBoost juntamente com outros algoritmos. Ao dividir as tabelas e resultados, permitiria uma maior perceção das diferenças de resultados entre as duas etapas. Para todos os efeitos, os autores adotaram a seguinte estrutura no seu artigo, desde o *dataset*, ao equilíbrio, a divisão por etapas, testagem e resultados (Figura 3.2).

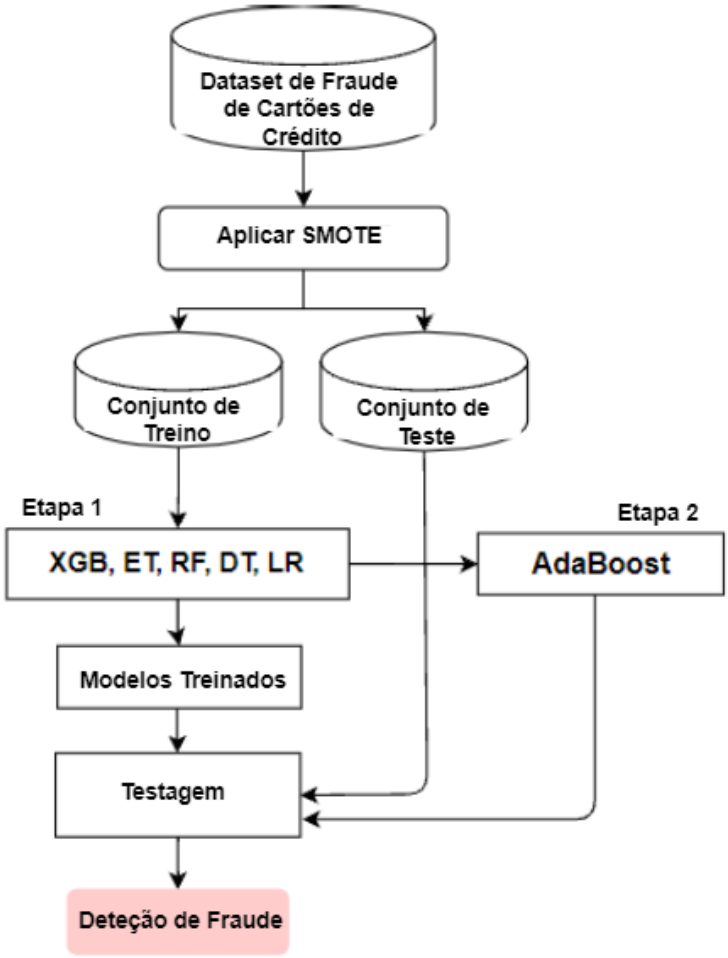

Figura 3.2 - Estrutura de deteção para deteção de fraude adotada

Fonte: Adaptado de Lellis et al., 2022, p. 165288

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Quanto aos resultados, para determinar a capacidade de eficácia e precisão dos respetivos algoritmos, os autores adotaram métricas de classificação como a *Accuracy (AC)*, o *Recall* (RC), a *Precision* (PR), e o *Mathhews Correlation Coefficient* (MCC).

Face ao exposto, os autores obtiveram os seguintes resultados:

Tabela 3.3 - Resultados sem o método AdaBoost

Model	AC	RC	PR	MCC
DT	99.91%	75.57%	79.83%	0.78
RF	99.95%	79.38%	97.19%	0.88
ET	99.95%	78.19%	96.29%	0.86
XGB	99.90%	59.39%	84.04%	0.71
LR	99.90%	56.55%	85.18%	0.59

Fonte: Lellis et al., 2022, p. 165290

Tabela 3.4 - Resultados com o método AdaBoost

Model	AC	RC	PR	MCC
DT	99.67%	99.00%	98.79%	0.98
RF	99.95%	99.77%	99.91%	0.99
ET	99.98%	99.96%	99.93%	0.99
XGB	99.98%	99.97%	99.92%	0.99
LR	98.75%	93.83%	97.56 %	0.94

Fonte: Lellis et al., 2022, p. 165290

Sem o método de *AdaBoost* (Tabela 3.3), observamos que em termos de qualidade da classificação (MCC), o RF foi o que possuiu melhor resultado com 0.88. Em termos de AC, os classificadores que obtiveram a melhor performance foram o RF e o ET com AC de 99.95% e 99.98% consecutivamente.

Com o método *AdaBoost* (Tabela 3.4), observamos que o DT obteve um MCC de 0.98, o que é 0.20 pontos superior que o resultado anteriormente obtido. No mesmo sentido, o algoritmo XGB registou um aumento no MCC de 0.28 pontos, face aos resultados anteriores. Contudo, continuam abaixo em termos de qualidade de classificação, quando

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

comparadas com o algoritmo RF. Relativamente à AC, tanto o XGB como o ET obtiveram um resultado de 99.98%, superior aos 99.95% do RF.

Relativamente ao equilíbrio e não equilíbrio dos dados, o uso do método SMOTE-*AdaBoost* melhorou consideravelmente a performance de todos os modelos quando temos em consideração o PR e o RC. Desde logo, o modelo DT, sem o método SMOTE-*AdaBoost*, atingiu um RC de 75.75%, em contraste aos 99.00% quando aplicado o modelo SMOTE-*AdaBoost*. Quanto à PR, o algoritmo DT atingiu uma precisão de 79.83% sem o modelo SMOTE-*AdaBoost*, mas quando este foi aplicado, aumentou para 98.79%. No mesmo sentido, o MCC, também melhorou de 0.78 para 0.98 pontos. Este padrão foi observado em todos os outros algoritmos que foram considerados neste estudo (Figura 3.3, Figura 3.4 e Figura 3.5).

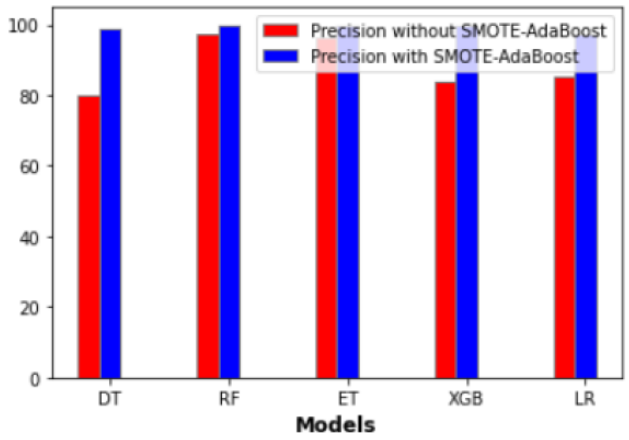

Figura 3.3 - Diferença do método SMOTE-AdaBoost na PR

Fonte: Lellis et al., 2022, p. 165292

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

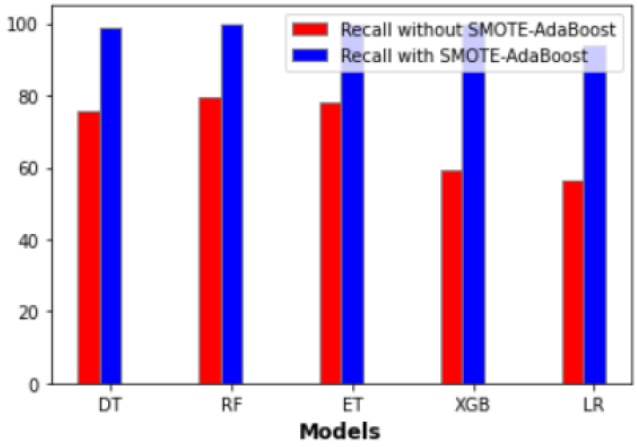

Figura 3.4 - Diferença do método SMOTE-AdaBoost no RC

Fonte: Lellis et al., 2022, p. 165292

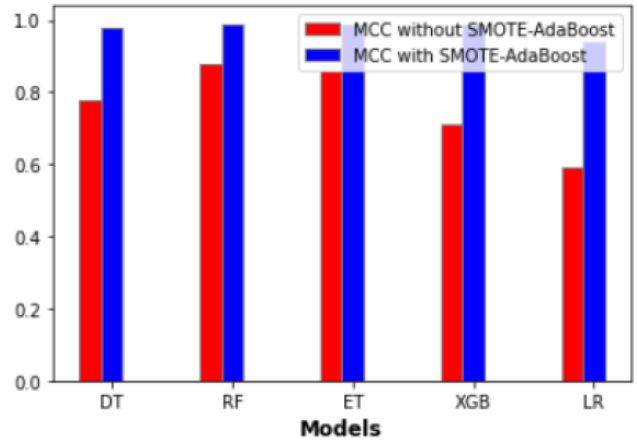

Figura 3.5 - Diferença do método SMOTE-AdaBoost no MCC

Fonte: Lellis et al., 2022, p. 165292

Salientamos que os autores ainda fizeram uma posterior análise utilizando os mesmos algoritmos, adicionando novas métricas de classificação como o AUC, mas com a utilização de dados sintéticos, em que as variáveis desta vez não são anonimizadas (*User, Cart, Year, etc.*). Os dados possuem 24357143 transações legitimas e 29757 transações fraudulentas. Em todos os algoritmos foi utilizado o modelo *AdaBoost*, obtendo os seguintes resultados (Tabela 3.5):

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Tabela 3.5 - Resultados dos algoritmos com o método AdaBoost, num dataset sintético

Model	AC	RC	PR	MCC
DT	99.67%	99.00%	98.79%	0.98
RF	99.95%	99.86%	99.95%	0.99
ET	99.99%	100%	99.93%	0.99
XGB	99.98%	99.99%	99.93%	0.99
LR	100.0%	98.89%	78.82 %	0.15

Fonte: Lellis et al., 2022, p. 165292

O algoritmo que melhor resultado obteve foi o *ET-AdaBoost* com uma AC, RC e PR de 99.99%, e um MCC de 0.99. Um padrão semelhante é possível verificar-se no algoritmo *DT-AdaBoost* e *RF-AdaBoost*.

Assim como, quando confrontando a métrica de classificação AUC, todos os algoritmos obtiveram uma curva ROC de 1, enquanto o algoritmo da RL, obteve o valor 0.66 (Figura 3.6).

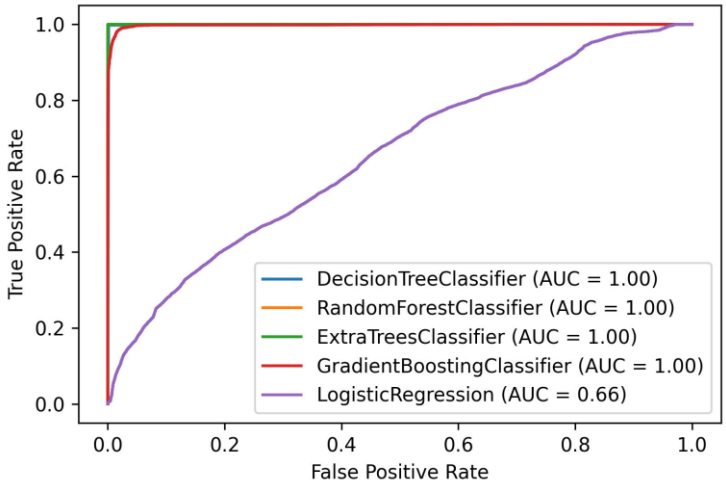

Figura 3.6 - Curva ROC para os diferentes algoritmos

Fonte: Lellis et al., 2022, p. 165293

Por último, os autores concluem que os resultados demonstram que a utilização do método de SMOTE conjuntamente com *AdaBoost* nos modelos de classificação tem um

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

impacto positivo na performance geral, de um sistema de deteção de fraudes transacionais.

3.3 *Efficient Resampling for Fraud Detection During Anonymized Credit Card Transactions with Unbalanced Datasets*

O presente artigo consiste num trabalho elaborado por Mrozek, *et al.* (2020), que tem como objetivo atender à problemática da fraude nos cartões de crédito, especialmente com o rápido crescimento do comércio eletrónico e das compras online.

Os autores decidiram organizar o presente artigo em: 1) estudos anteriores sobre a mitigação e deteção de fraude online; 2) algoritmos de ML que serão alvo de escrutínio; 3) técnicas utilizadas para equilibrar os dados; 4) apresentar a estratégia de avaliação, o *dataset* utilizado e resultados obtidos.

Quanto ao *dataset* em causa, este contém um total de 284,807 mil transações, em que 492 destas constituem fraude, e as restantes são transações legitimas. Estamos perante um ratio de 0.172% entre as duas classes, logo um *dataset* bastante desequilibrado como já seria esperado. Para resolver este problema de desequilíbrio, os autores decidiram utilizar duas técnicas para equilibrar os dados, como o RUS e o SMOTE.

Quanto à divisão do *dataset* em causa, decidiram utilizar a técnica *5-Fold Cross Validation*, em que o *dataset* é dividido em cinco partes diferentes de igual tamanho, onde o número amostras em cada classe (minoritária e maioritária) também é dividido em proporções iguais. Ao longo de todo o processo de validação, uma única parte (ou seja, 20% do *dataset*), é reservada para testar a performance de um determinado algoritmo em causa e, as outras quatro partes (80%) são realocadas ao conjunto de treino. Utilizamos este processo até que as cinco partes sejam utilizadas, e depois faz-se uma média da performance de cada uma das partes, e obtemos posteriormente a performance do algoritmo em causa que estamos a analisar.

De salientar, que os autores para além das técnicas de equilíbrio anteriormente referidas, decidiram testar os vários algoritmos de ML, sem equilibrar o *dataset*, com o intuito de

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

comparar se obteriam melhor resultados se houvesse ou não a aplicação das respetivas técnicas.

Não obstante, em termos de algoritmos de ML, os autores analisaram os seguintes:

- Regressão Logística (RL);
- *Random Forest* (RF);
- *K-Nearest Neighbour* (KNN);
- *Stochastic Gradient Descent* (SGD);

Quanto a métricas de classificação e performance dos respetivos algoritmos de deteção de fraude, utilizaram a AUC, *F1-Score*, *Precision*, *Recall*, *Accuracy*, obtendo os seguintes resultados (Tabela 3.6):

Tabela 3.6 - Comparação da performance dos algoritmos com as variadas técnicas de equilíbrio

Model	Accuracy	Precision	Recall	F1 Score	AUC
LogReg without resampling	99.89%	73.46%	64.62%	68.84%	0.82
LogReg SMOTE resampling	98.14%	7.55%	87.07%	13.90%	0.92
LogReg Under-sampling	96.43%	4.22%	91.15%	0.08%	0.95
RFC without resampling	99.95%	94.16%	76.87%	84.64%	0.88
RFC SMOTE resampling	99.95%	89.70%	82.99%	86.21%	0.91
RFC Under-sampling	**97.43%**	**0.06%**	**100.00%**	**11.82%**	**0.98**
KNN without resampling	99.83%	85.71%	4.08%	7.79%	0.52
KNN SMOTE resampling	93.81%	1.51%	54.42%	2.93%	0.74
KNN Under-sampling	64.52%	0.37%	78.23%	0.75%	0.71
SGD without resampling	99.81%	0%	0%	0%	0.49
SGD with SMOTE resampling	96.18%	3.80%	87.07%	7.28%	0.91
SGD with Under-sampling	0.17%	0.17%	100%	0.34%	0.5

Fonte: Mrozek et al., 2020, p. 432

Numa análise individual a cada algoritmo, a RL sem *resampling* ou equilíbrio, atingiu num panorama geral, resultados mais favoráveis nas mais variadas métricas de classificação. Apesar de o algoritmo de RL com *under-sampling* obter uma capacidade mais elevada de detetar transações, determina incorretamente como fraude um número avultado de transações.

Quanto ao algoritmo de RF, na perspetiva de RUS este detetou todas as transações fraudulentas do *dataset*. Não obstante, os outros métodos também obtiveram resultados bastante positivos, com o RF sem *resampling* a ultrapassar o equilíbrio com SMOTE em algumas classificações (caso da AUC, *Precision* e empatar na *Accuracy*).

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

No algoritmo KNN, os resultados não foram de todo favoráveis, quando comparado com os algoritmos até agora analisados. Apesar do método de RUS e SMOTE melhorarem os resultados do algoritmo, segundo os autores, não é um algoritmo indicado para a deteção de fraude transacional, devido à sua baixa capacidade de deteção e errada classificação de transações legitimas como fraudulentas.

Por último, o algoritmo SGD, também seguindo a lógica anteriormente referida pelos autores, não parece de todo indicado para deteção de fraude transacional. É o primeiro algoritmo dos já referidos até agora, que obteve melhores resultados no *oversampling* (SMOTE), quando comparado com o método de RUS e sem *resampling*. Não obstante, os resultados obtidos continuam a ser de todo favoráveis para uma correta deteção de fraude, devido à sua volatilidade entre as métricas de classificação, com *AC* a 99.81%, mas 0% nas restantes classificações no caso de sem "*resampling*", ou até um Recall de 100%, mas aproximado a 0% nas restantes classificações no caso de RUS.

Numa visão geral, a Tabela 3.6 demonstra que de todas métricas avaliadas, o algoritmo RF com o método de equilíbrio de RUS, que obteve o melhor resultado com 0.98 de AUC, ou seja, há uma probabilidade 98% de corretamente classificar as transações que são fraudulentas, das que não são. Obtendo também um resultado de 100% na métrica de *Recall*, ou seja, detetou todas as transações fraudulentas. De também salientar os ótimos resultados do algoritmo da RL, na perspetiva de RUS e SMOTE, sendo uma das escolhas de eleição, para além do RF, na deteção da fraude transacional.

3.4 *Sequential credit card fraud detection: A joint deep neural network and probabilistic graphical model approach*

Neste presente estudo, os autores Forough, *et al.* (2020), partilham a opinião, que pelas características da natureza sequencial ou constante dos dados transacionais, determinados algoritmos de deteção de fraude, não têm sido alvo da devida atenção, e como tal merecem o seu escrutínio.

Para testar os algoritmos, os autores utilizaram dois *datasets* reais. Um primeiro que retrata transações produzidas por detentores de cartões de crédito europeus, em que das

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

30 variáveis que possui, 28 encontram-se anonimizadas (V1, …, V28), e as restantes duas "*Time*" e "*Amount*" não. Existe também no *dataset*, uma coluna na base de dados, que classifica em código binário se estamos perante fraude (1) ou não (0). O segundo *dataset*, retrata transações produzidas num banco brasileiro, que contem 17 variáveis, e uma coluna que determina através de dois valores (S/N) o que é fraude do que não é. Podemos observar na Tabela 3.7, a estrutura de cada um de os respetivos *datasets*, alvo de análise.

Tabela 3.7 - Estrutura dos dois datasets que são alvo de estudo

Dataset	Normal	Fraudulent	Features	Instances	Fraud/normal ratio (%)
European cards dataset	284,315	492	30	284,807	0.172
The Brazilian dataset	360,792	14,031	17	374,823	3.74

Fonte: Forough, et al., 2020, p. 7

Quanto à problemática do não equilíbrio dos dados, os autores decidiram utilizar um método conotado por *Sequence-Aware Undersampling* (Seq-US), que comparando com os métodos tradicionais de *Undersampling* e *Oversampling*, não mantém a importância dos padrões sequenciais ou constantes dos dados. Este modelo mantém o padrão sequencial antes de uma determinada transação ser reconhecida como fraude, pois essa sequência possui informação essencial que permitiu descodificar determinada transação como fraude (*critical instances*). Após a retenção desta informação, retira do *undersampling* as sequências que não foram essenciais na captura desta transação fraudulenta (*safe instances*).

Os autores dividiram ambos os *datasets*, em 70% conjunto de treino, 10% como validação e 20% como conjunto de teste.

Quanto a métricas de classificação dos algoritmos a serem analisados, foram utilizadas a *Precision*, *Recall*, *F1-Score*, AUROC, e o AUPR (Area Under the Precision-Recall) esta última, consiste numa métrica que tem como principal foco de análise a classe positiva (neste caso as situações de fraude), sendo esta uma curva que combina a *Precision* e o *Recall*).

Relativamente aos algoritmos em análise, estes algoritmos são:

- *Hidden Markov Model* (HMM);

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

- *Maximum Entropy Markov* (MEM);
- *Conditional Random Fields* (CRF);
- *Recurrent Neural-Work* (RNN);
- *Long Short-Term Memory* (LSTM);
- *Gated Recurrent Unit* (GRU);
- *Artificial Neural Network* (ANN);
- *Long Short-Term Memory – Conditional Random Fields* (LSTM-CRF): este é o modelo proposto pelos autores.

Obtendo os seguintes resultados apresentados nas tabelas Tabela 3.8 e Tabela 3.9.

Tabela 3.8 - Comparação dos algoritmos ANN, GRU, LSTM, LSTM-CRF no dataset Europeu

Model	Precision	Recall	F1	AUC-ROC	AUC-PR
ANN	0.8061	0.7416	0.7648	0.8706	0.5955
GRU	0.8626	0.7208	0.7792	0.8602	0.6177
LSTM	0.8575	0.7408	0.7866	0.8702	0.6292
LSTM-CRF	**0.8817***	**0.7569***	**0.8076***	**0.8783**	**0.6623***

Fonte: Forough, et al., 2020, p. 8

Tabela 3.9 - Comparação dos algoritmos ANN, GRU, LSTM, LSTM-CRF no dataset Brasileiro

Model	Precision	Recall	F1	AUC-ROC	AUC-PR
ANN	0.8898	0.7626	0.8187	0.8795	0.6832
GRU	0.8043	0.689	0.7419	0.8413	0.5655
LSTM	0.8776	0.7144	0.7874	0.8847	0.6746
LSTM-CRF	**0.9256***	**0.7955***	**0.8555***	**0.8965**	**0.7439***

Fonte: Forough, et al., 2020, p. 9

Como podemos observar pelos resultados na Tabela 3.8 e Tabela 3.9, o modelo proposto pelos autores LSTM-CRF obteve os melhores resultados em todas as métricas de classificação. Tanto no *dataset* Europeu como no *dataset* Brasileiro, demonstrou segundo os autores como "a melhor técnica atual de deteção de fraude (…) quando comparada com métodos tradicionais de ML, obtendo melhorias significativas em comparação".

Não obstante, os autores também decidiram realizar um outro teste, que foi adicionar mais uma camada ao algoritmo LSTM-CRF, mais concretamente uma camada LSTM (Tabela 3.10 e Tabela 3.11)

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Tabela 3.10 - Resultados do modelo proposto usando duas camadas LSTM no dataset Europeu

Model	Precision	Recall	F1	AUC-ROC	AUC-PR
LSTM-CRF	0.8817	0.7569	0.8076	0.8783	0.6623
LSTM-LSTM-CRF	0.8355	0.7525	0.7874	0.8761	0.6258

Fonte: Forough, et al., 2020, p. 9

Tabela 3.11 - Resultados do modelo proposto usando duas camadas LSTM no dataset Brasileiro

Model	Precision	Recall	F1	AUC-ROC	AUC-PR
LSTM-CRF	0.9256	0.7955	0.8555	0.8965	0.7439
LSTM-LSTM-CRF	0.9425	0.7596	0.8365	0.8789	0.7232

Fonte: Forough, et al., 2020, p. 9

Face aos resultados da Tabela 3.10 e Tabela 3.11, verificamos que adicionando mais uma camada LSTM ao modelo proposto, a performance do mesmo, diminuiu consideravelmente. Neste sentido, não há necessidade de adicionar mais camadas à sequência, sendo que obtemos resultados razoáveis com uma camada única.

Por último, os autores decidiram testar também as várias técnicas de *Oversampling* (OS) e *Undersampling*,(US) nos diferentes *datasets*, tendo como referência o modelo proposto pelos autores LSTM-CRF, obtendo os seguintes resultados (Tabela 3.12 e Tabela 3.13):

Tabela 3.12 - Comparação com as diferentes técnicas de Oversampling e Undersampling tendo como referência o modelo proposto LSTM-CRF, utilizando um ratio de 0.006 no dataset Europeu

Method	Precision	Recall	F1	AUC-ROC	AUC-PR	Tr-time	Instances
No Sampling	0.8817	0.7569	0.8076	0.8783	0.6623	568.86	199,361
Random-US	0.8583	0.7697	0.8031	0.8847	0.6541	179.6	61,686
Seq-US	0.8773	**0.7878**	**0.8204**	**0.8938**	**0.6834**	182.88	61,686
Random-OS	0.8825	0.7621	0.8154	0.8809	0.6711	586.57	200,184
ADASYN-OS	0.891	0.7271	0.7957	0.8635	0.6448	592.58	200,184
SMOTE-OS	**0.8925**	0.7234	0.7936	0.8616	0.6417	590.27	200,184

Fonte: Forough, et al., 2020, p. 10

Tabela 3.13 - Comparação com as diferentes técnicas de Oversampling e Undersampling tendo como referência o modelo proposto LSTM-CRF, utilizando um ratio de 0.005 no dataset Brasileiro

Method	Precision	Recall	F1	AUC-ROC	AUC-PR	Tr-time	Instances
No sampling	**0.9256**	0.7955	0.8555	0.8965	0.7439	745.72	262,372
Random-US	0.8933	0.793	0.84	0.8946	0.7161	588.57	206,481
Seq-US	0.9124	**0.8128**	**0.8596**	**0.9049**	**0.7487**	634.64	206,481
Random-OS	0.8988	0.7995	0.8462	0.898	0.7261	771.12	265,147
ADASYN-OS	0.9179	0.7784	0.8422	0.8878	0.7227	774.77	265,147
SMOTE-OS	0.902	0.7889	0.8416	0.8928	0.7196	773.61	265,147

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Fonte: Forough, et al., 2020, p. 10

Podemos observar que dos resultados obtidos na Tabela 3.12 e Tabela 3.13, o modelo proposto pelos autores de Seq-US atingiu os melhores resultados nas métricas de classificação de Recall, F1-Score, AUC-ROC e AUC-PR.

Em conclusão, os autores são da opinião que o modelo proposto LSTM-CRF atinge melhores resultados quando comparado com o LSTM base, assim como com outros dois algoritmos de DL, como o ANN e o GRU. Realçam também a necessidade de equilibrar os *datasets* de fraude, devido à discrepância entre as classes, aconselhando para isso a utilização da técnica Seq-US.

3.5 *A Neural Network Ensemble with Feature Engineering for Improved Credit Card Fraud Detection*

Neste artigo, os autores Esenogho, *et al.* (2022), tentam demonstrar que os algoritmos de *Deep Learning* (DL) possuem um grande potencial, quando comparados com as fragilidades que os métodos convencionais de ML possuem. Estes últimos, possuem uma particular dificuldade na adaptação às dinâmicas e tendências de compras e bens pelos titulares de cartões de crédito, resultando em classificações erradas destes sistemas convencionais de deteção de fraude. Neste sentido, os autores para abordar esta problemática, decidiram analisar vários algoritmos de ML e comparar os respetivos resultados.

Relativamente ao *dataset* alvo de estudo, este provém de transações de cartão de crédito realizadas num período de 2 dias, por clientes europeus. Estamos perante um *dataset* não equilibrado, com 492 transações fraudulentas e 284807 mil transações não fraudulentas, em que as suas variáveis foram anonimizadas (V1,…,V28), menos duas, mais concretamente o "*Time*" e o "*Amount*", possuindo também uma outra variável conotada de "*Class*" que divide em código binário se estamos perante uma transação fraudulenta (1), ou não (0).

Por estarmos perante um *dataset* que é bastante desequilibrado, foi utilizada a técnica de SMOTE-ENN, que consiste num modelo híbrido de equilibrar a base de dados com

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

oversampling e *undersampling*, na medida em que usa o método SMOTE para aumentar a quantidade de amostras da classe minoritária (*oversampling*), e o ENN (*Edited Nearest Neighbour*) para remover a sobreposição entre instâncias, ou seja, remove algumas das amostras da classe maioritária que distinga dos K "vizinhos" mais próximos.

Quanto às métricas de classificação, foi utilizada a *Sensivity* ou *Recall*, *Specificity* e AUC. Não obstante, para avaliar a performance dos modelos, foi utilizada a técnica Cross-*Validation*, em que divide o *dataset* em duas repartições, a primeira em conjunto treino ou conjunto teste, e a segunda num conjunto de validação. Há vários modelos de Cross-*Validation*, aquele que foi abordado pelos autores, foi o *K-Fold Cross Validation*, que garante que a proporção de amostras de fraude e de não fraude encontradas no *dataset* é preservada em todas divisões ou "*folds*" (no caso em concreto foram 10 divisões).

Quanto aos algoritmos a serem alvo de análise os autores escolheram os seguintes:

- *AdaBoost* ou *Adaptive Boosting;*
- *Long Short-Term Memory Neural Network* (LSTM);
- *Decision Tree* (DT);
- *Support Vector Machine* (SVM);
- *Multilayer Perceptron* (MLP);
- *LSTM Ensemble*;

Relativamente aos resultados, os autores dividiram em duas partes, uma sem a utilização do equilíbrio dos dados (Tabela 3.14) e outra com equilíbrio dos dados utilizando a técnica SMOTE-ENN (Tabela 3.15).

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Tabela 3.14 - Resultados sem o equilíbrio dos dados através do SMOTE-ENN

Algorithm	Sensitivity	Specificity	AUC
SVM	0.583	0.954	0.640
MLP	0.755	0.961	0.810
Decision tree	0.588	0.943	0.690
AdaBoost	0.746	0.975	0.830
LSTM	0.761	0.969	0.780
Proposed LSTM Ensemble	0.839	0.982	0.890

Fonte: Esenogho, et al. (2022), p. 16404

Tabela 3.15 - Resultados com o equilíbrio dos dados através do SMOTE-ENN

Algorithm	Sensitivity	Specificity	AUC
SVM	0.912	0.970	0.940
MLP	0.938	0.982	0.930
Decision tree	0.907	0.951	0.920
AdaBoost	0.968	0.994	0.970
LSTM	0.962	0.978	0.950
Proposed LSTM Ensemble	0.996	0.998	0.990

Fonte: Esenogho, et al. (2022), p. 16404

Quanto aos resultados sem o SMOTE-ENN (Tabela 3.14), ou seja, com um desequilíbrio no *dataset*, conseguimos observar que o modelo/algoritmo LSTM Ensemble proposto pelos autores, obteve um dos melhores resultados quando comparado com os restantes algoritmos, desde logo com uma *Sensivity* de 0.839, uma *Specificity* de 0.982, e um AUC de 0.890. Não obstante, é salientado também a baixa *Sensivity* de uma forma geral, devido ao não equilíbrio dos dados em causa.

Quanto aos resultados com o SMOTE-ENN (Tabela 3.15), ou seja, com um equilíbrio no *dataset*, conseguimos observar que mais uma vez o modelo LSTM Ensemble proposto pelos autores, obteve um dos melhores resultados, desde logo com uma *Sensivity* de 0.996, *Specificity* de 0.998, e um AUC de 0.990. Observamos também, que assim como este modelo proposto, todos os outros algoritmos obtiveram valores de *Sensivity* bastante mais significativos, sendo uma característica essencial na deteção de fraude, já que esta métrica indica a proporção de amostras de fraude foram corretamente detetadas pelo modelo.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Por último, a curva ROC é fulcral para analisarmos a discrepância entre o rácio dos verdadeiros positivos e o rácio dos falsos positivos, que quanto mais próximo do valor 1 (ou quanto superior esquerdo do gráfico), melhor a sua capacidade de prever fraude que os restantes modelos. Mais uma vez, o modelo proposto pelos autores, obteve uma AUC de 0.99, que é superior que os restantes modelos (Figura 3.7).

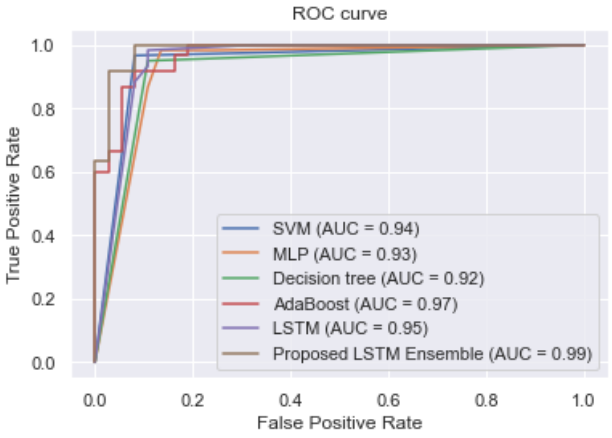

Figura 3.7 - Curva Roc com os vários algoritmos em análise

Fonte: Esenogho, et al. (2022), p. 16404

Não obstante, é importante salientar que neste artigo os autores decidiram comparar tanto os resultados que obtiveram neste *dataset*, com outros modelos utilizados por diferentes autores, assim como testar com diferentes *datasets* o modelo proposto. Decidimos não abordar esta temática, desde logo, porque a comparação das métricas de *Sensivity* e *Specificity* não era amplamente abordada pelos outros autores, e os diferentes *datasets* alvo de posterior análise, possuíam um número de amostras bastante reduzido (30000 mil num caso e 1000 no outro).

3.6 *Fraud Detection in Banking Data by Machine Learning Techniques*

No presente estudo realizado pelos autores Hashemi, *et al.* (2022), o constante avanço da tecnologia e dos respetivos serviços de e-commerce, e o uso cada vez mais recorrente do pagamento por cartões de crédito, resultou num aumento do volume de transações

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

bancárias. Como tal, detetar atividades fraudulentas surge como um tópico fascinante e foco de interesse.

Neste sentido, os autores decidiram analisar um *dataset* composto por 284,807 mil transações de cartão de crédito, realizadas num período de dois dias. Entre estas, 492 são transações fraudulentas (0.17%), sendo o resto legítimas (99.83%), contendo 28 variáveis anonimizadas (V1, …, V28), e duas em que não o são ("*Amount*" e "*Time*"), assim como uma outra variável que através de valores binários, que determina o que é fraude (1) do que não é (0).

Como é possível verificar, existe uma grande discrepância entre estas classes, sendo um *dataset* bastante desequilibrado. Para mitigar este efeito, os autores decidiram utilizar uma técnica de equilíbrio dos dados, chamada de *Bayesian Optimization*. Esta técnica de pré-processamento de dados que tem como objetivo encontrar os melhores conjuntos de parâmetros, que configuram melhor os nossos algoritmos de ML durante o seu treino. Neste caso, os parâmetros utilizados pelos autores, têm como objetivo reduzir o tempo computacional de análise uma transação, e a melhorar a performance dos algoritmos. Ou seja, esta técnica é utilizada para resolver o problema do desequilíbrio entre estas classes, fazendo uma afinação dos pesos dos parâmetros.

Na fase de processamento, os autores utilizaram a técnica 5 *Cross-Fold Validation*, para obter uma comparação da performance no *dataset* desequilibrado. Na medida em que o *dataset* é dividido em cinco partes diferentes de igual tamanho, onde o número amostras em cada classe (minoritária e maioritária) também é dividido em proporções iguais. Ao longo de todo o processo de validação, uma única parte (ou seja, 20% do *dataset*), é reservada para testar a performance de um determinado algoritmo em causa e, as outras quatro partes (80%) são realocadas ao conjunto de treino. Utilizamos este processo até que as cinco partes sejam utilizadas, e depois faz-se uma média da performance de cada uma das partes, e obtemos posteriormente a performance do algoritmo em causa que estamos a analisar.

Quanto às métricas de classificação foram utilizadas a *Accuracy*, *Precision*, *Recall*, o *F1-Score*, a AUC e por último a MCC.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Por último, os autores ainda utilizaram uma técnica para chamada de *Information Gain* (IG), através da escolha de um determinado conjunto de variáveis, que melhoram a performance do classificador na deteção de fraude. Ora, por estarmos perante um *dataset* anonimizado, os autores apenas conhecem a variável "Time" e "Amount", não tendo qualquer outro tipo de informação adicional. Neste sentido, o método IG é usado para selecionar as variáveis mais importantes do *dataset*, com intuito de aumentar o desempenho do modelo em causa. Fá-lo, através da extração das semelhanças entre cada uma das transações e atribuindo um maior peso para aquelas variáveis mais significativas, tendo em consideração a classe das transações legitimas e fraudulentas. Tendo em consideração a Figura 3.8, os autores extraíram as 6 melhores variáveis e foram estas as usadas para avaliar os algoritmos propostos.

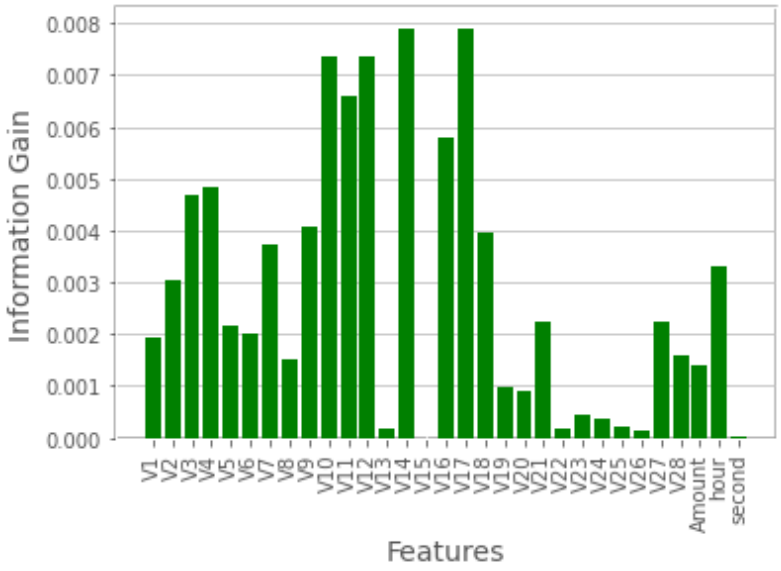

Figura 3.8 - Importância de cada uma das variáveis desconhecidas no dataset, tendo em consideração a classe das transações legitimas e fraudulentas

Fonte: Hashemi, et al. (2022), p. 3038

Quanto aos algoritmos análise, os autores optaram pelos seguintes:

- *Regressão Logística* (RL);
- *LightGBM* (LGBM);
- *XGBoost* (XGB);

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

- *CatBoost*;
- *Majority Voting Ensemble Learning* (VOT);
- *Artificial Neural Network* (ANN): com o apoio da plataforma Keras

Relativamente aos resultados, os autores decidiram abordar esta temática de várias formas. Em primeiro lugar, executaram os algoritmos de ML, salientando a vermelho os melhores resultados que cada algoritmo teve, face aos demais (Tabela 3.16):

Tabela 3.16 - Resultados obtidos utilizando algoritmos de ML

Model	Accuracy	AUC	Recall	Precision	F1-score	MCC
Log_Reg	0.97477	0.9578	0.8730	0.0617	0.1143	0.2248
LGBM	0.99919	0.9472	0.7990	0.7534	0.7699	0.7727
XGB	0.99923	0.9517	0.7949	0.7862	0.7830	0.7864
CatBoost	0.99880	0.9390	0.8096	0.6431	0.7066	0.7158
Vot_Lg, Xg, Ca	0.99924	0.9501	0.8033	0.7720	0.7825	0.7847
Vot_Lg, Xg	0.99927	0.9522	0.8012	0.7901	0.7901	0.7925
Vot_g, Ca	0.99923	0.9492	0.8097	0.7681	0.7823	0.7852
Vot_Lg, Ca	0.99912	0.9459	0.8075	0.7260	0.7581	0.7620

Fonte: Hashemi, et al. (2022), p. 3041

Por outro lado, e como podemos constatar na Tabela 3.17, os autores decidiram dividir os algoritmos de ML de DL. Neste sentido, e recorrendo à plataforma Keras, os autores testaram o algoritmo ANN obtendo os seguintes resultados:

Tabela 3.17 - Resultado obtido utilizando algoritmo de DL - ANN

Model	Accuracy	AUC	Recall	Precision	F1-score	MCC
Keras	0.9994	0.9401	0.8222	0.8043	0.8132	0.8129

Fonte: Hashemi, et al. (2022), p. 3041

Podemos observar na Tabela 3.16, que o algoritmo da RL obteve o valor mais alto de deteção de fraude com 0.9578, contudo contém dos valores mais baixos nos outros critérios.

Ora, para os autores, é necessário ter em consideração uma média da conjuntura de todas as métricas de classificação. Neste sentido os autores na Figura 3.9, demonstram um

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

gráfico de barras que clarifica os algoritmos que tiveram a melhor distribuição numa forma global.

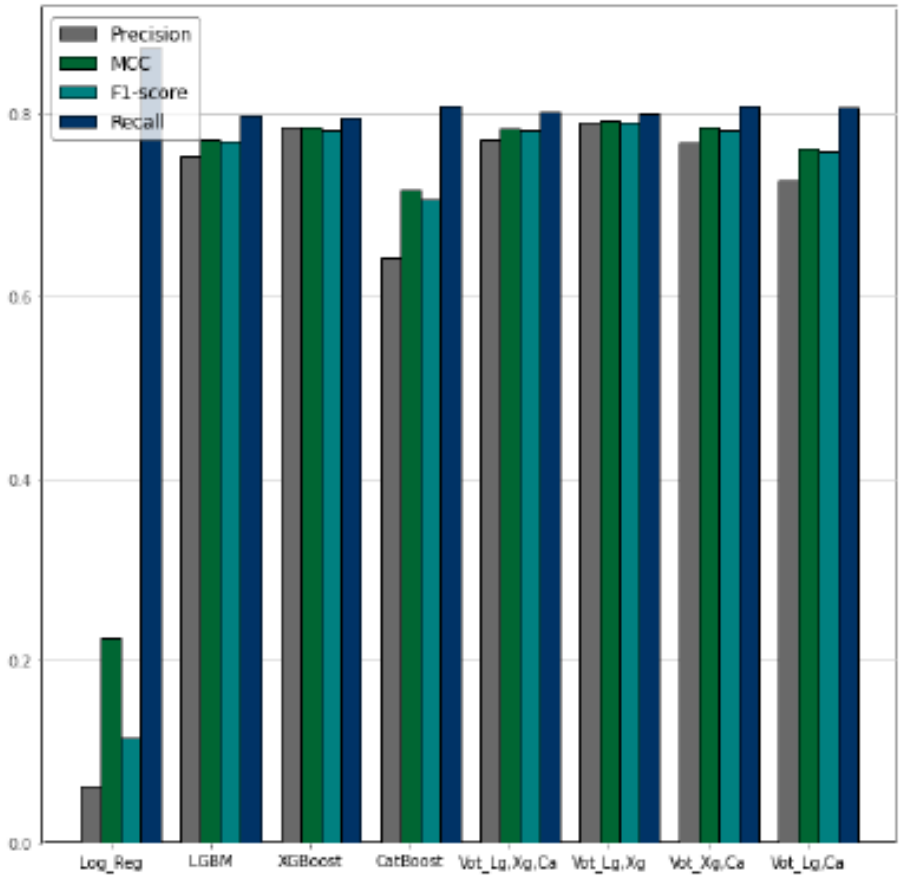

Figura 3.9 - Distribuição dos algoritmos tendo em conta todas as métricas de classificação

Fonte: Hashemi, et al. (2022), p. 3041

Ora, se formos a ter em consideração as métricas de *Precision*, *Recall*, *F1-Score* e MCC, o melhor algoritmo será a combinação entre o LighGBM e o XGBoost (*Majority Voting*), que possuem em média 0.79 nestes critérios (Figura 3.9). No caso dos algoritmos individuais, o XGBoost obteve o melhor resultado.

Por outro lado, se observarmos a Tabela 3.17, o algoritmo de DL obteve uma melhor performance quando comparado com cada um dos algoritmos de ML anteriormente analisados, mesmo aqueles que foram alvo de *Ensemble Majority Voting*, obtendo uma performance em todas as métricas de classificação, menos no AUC.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Por último, os autores decidiram analisar três diferentes algoritmos que foram abordados ao londo do respetivo artigo. Em primeiro lugar, um modelo conotado pela Tabela 3.8 como "*Method presented in [17]*", realizado pelos autores *Altyeb et al* (2020), em que utiliza na mesma o método Bayesian-based com otimização de hiperparametros num algoritmo de LightGBM. Em segundo lugar, conforme a Tabela 3.8, o modelo "Proposed LightGBM", os autores tiveram em consideração o peso das classes para escolher os respetivos hiperparametros, o que segundo os autores, melhorou a capacidade de deteção dos casos de fraude em 50% e o correspondente F1-Score em 20%, quando comparamos com o o "Method presented in [17]". Em último lugar, os autores salientam o modelo criado pelos mesmos, que consiste em dividir o dataset em conjunto de treino e teste, aplicar posteriormente Bayesian Optimization, e de seguida testar a performance dos algoritmos utilizando 5-Fold Cross Validation, aplicando por ultimo as respetivas métricas de classificação.

Como foi possível observar pela Tabela 3.18, o *Proposed Approach*, obteve um melhor resultado global quando comparando as respetivas métricas de classificação.

Tabela 3.18 - Comparação da performance de uma das bibliografias apresentada no artigo [17], do Proposed LightGBM, e Proposed Approach

Model	Accuracy	AUC	Recall	Precision	F1-score
Method presented in [17]	0.984	0.909	0.406	0.973	0.569
Proposed LightGBM	0.9992	0.947	0.799	0.753	0.769
Proposed Approach	0.9993	0.952	0.801	0.79	0.79

Fonte: Hashemi, et al. (2022), p. 3042

Por último, segundo os autores, a sua finalidade em dividir os respetivos resultados em duas tabelas diferentes (Tabela 3.16 e Tabela 3.18), por um lado era demonstrar a melhoria significativa de utilizar Majority Voting (modelos híbridos), não obstante dos resultados obtidos pelo algortimo de DL. Por outro, era realçar a importância dos modelos que utilizam hiperparametros para abordar o tema dos dados não equilibridos, quando comparados com os métodos de "*sampling*", seja por utilizarem menor memória e possuirem um menor tempo de avaliação dos algoritmos, obtendo consequentemente também melhores resultados.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

3.7 *A Closer Look into the Characteristics of Fraudulent Card Transactions*

Segundo os autores Can, *et al.* (2020), o número de transações de cartão de crédito aumenta à medida que a utilização e desenvolvimentos das tecnologias e utilização das plataformas de *e-commerce* evoluem. Desde logo, em 2017 foram realizadas 375 mil milhões de transações, sendo 16.7 milhões destas correspondem a transações fraudulentas. Ora, apesar do ratio ser de 0,006%, pouco significativo perante o panorama geral, a verdade é que causa problemas reputacionais nas várias instituições financeiras. Ora, como maior parte dos bancos, o seu método de deteção de fraude consiste num sistema de regras e, perante as novas tendências de ML e a sua capacidade de detetar fraude mais eficazmente, os autores decidiram abordar esta temática.

Neste sentido, os autores decidiram analisar um *dataset*, contendo informação de 35 bancos da Turquia, com mais de quatro mil milhões de transações de crédito e de débito, entre janeiro e agosto de 2017, em que destas 245 mil transações foram consideradas como fraudulentas. Inicialmente o *dataset* continha 60 variáveis. Não obstante à medida que os autores foram avaliando a viabilidade de cada uma destas variáveis, e selecionando aquelas que eram relevantes para a testagem dos algoritmos (*Feature Selection*), reduziu-se a 30 variáveis alvo de estudo.

Em termos de equilíbrio dos dados, os autores seguiram alguns exemplos da bibliografia referenciada ao longo do artigo, e neste sentido apesar de uma escolha aleatória das transações que iriam fazer parte de cada conjunto de treino e de teste, decidiram agrupar e testar o rácio entre as transações fraudulentas e não fraudulentas. Neste sentido, testaram um ratio de 1:1, 5:1, 10:1, 15:1, 20:1, 25:1, 30:1, 35:1, 40:1, em que posteriormente aplicaram a métrica de classificação F-1 Score para determinar qual seria o melhor ratio para o equilíbrio de dados, optando para um ratio 5:1.

Quanto à divisão do *dataset* aquando da aplicação dos diferentes algoritmos, este foi divido em 70% conjunto de treino e 30% conjunto de teste. Posteriormente o conjunto de treino foi também divido em 70% e 30%, para treino e validação respetivamente. Quanto

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

às métricas de classificação, os autores utilizaram a *Precision*, *Specificity*, *F1-Score*, *Recall* e MCC.

Relativamente aos algoritmos a serem alvo de estudo, os autores decidiram abordar os seguintes:

- Naive Bayes (NB);
- Decision Tree (DT);
- Random Forest (RF);
- Multi-Layer Perceptron (MLP);

Relativamente aos resultados, os autores decidiram apresentar os mesmos tendo em conta dois diferentes cenários:

- Numa primeira análise, testaram os algoritmos tendo em consideração o *dataset* com as 38 variáveis viáveis, dividindo o mesmo entre 70% de conjunto de treino e 30% de conjunto de teste. Obtendo os seguintes resultados (Tabela 3.19):

Tabela 3.19 - Testagem dos algoritmos e as diferentes métricas de classificação

Algorithms	NF	NF	NF	F	F	F	
	Precision	Specificity	F-Measure	Precision	Recall	F-Measure	MCC
Naive Bayes	99.24%	92.90%	95.96%	73.20%	96.44%	83.22%	80.44%
Random Forest	98.54%	98.97%	98.75%	94.77%	92.71%	93.73%	92.50%
Decision Tree	98.66%	98.91%	98.81%	94.52%	93.32%	93.92%	92.65%
Multi-Layer Perceptron	98.84%	98.96%	98.90%	94.83%	94.20%	94.51%	93.41%

Fonte: Can, et al. (2020), p. 166101

- Numa segunda análise, decidiram testar os algoritmos, mas contrariamente à divisão realizada na Tabela 3.19, decidiram usar um mês do ano para treinar o modelo e o mês seguinte para testar a performance do modelo. Decidiram também, realizar a média da performance de cada classificador para os testes que foram realizados (Tabela 3.20).

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Tabela 3.20 - Testagem dos algoritmos dividindo o conjunto de treino e teste por meses, com a correspondente média de cada classificador.

Algorithm	Month	Non-Fraud(%)			Fraud(%)			MCC
		Precision	Specificity	F-Measure	Precision	Recall	F-Measure	
Naive Bayes	Feb.	99.47	92.73	95.98	72.96	97.56	83.49	80.87
	Mar.	99.41	92.85	96.01	73.22	97.25	83.54	80.89
	Apr.	99.37	93.80	96.51	75.87	97.03	85.16	82.67
	May.	99.09	93.34	96.13	74.33	95.75	83.69	80.88
	Jun.	98.73	93.64	96.12	74.89	94.05	83.39	80.35
	Jul.	98.83	92.72	95.68	72.33	94.54	81.96	78.80
	Aug.	99.06	92.60	95.72	72.29	95.65	82.35	79.36
	W.Avg.	**99.21**	**93.16**	**96.08**	**73.92**	**96.30**	**83.63**	**80.86**
Random Forest	Feb.	98.65	98.88	98.76	94.37	93.26	93.81	92.58
	Mar.	98.39	99.17	98.78	95.69	91.93	93.77	92.57
	Apr.	98.64	99.19	98.92	95.87	93.19	94.51	93.44
	May.	98.31	99.15	98.72	95.57	91.51	93.50	92.25
	Jun.	98.07	99.02	98.54	94.89	90.33	92.55	91.13
	Jul.	97.94	98.78	98.36	93.68	89.70	91.65	90.04
	Aug.	97.73	98.77	98.25	93.58	88.65	91.05	89.35
	W.Avg.	**98.35**	**99.04**	**98.70**	**95.07**	**91.76**	**93.38**	**92.10**
Decision Tree	Feb.	98.59	98.85	98.72	94.22	92.96	93.59	92.31
	Mar.	98.65	98.92	98.78	94.54	93.28	93.91	92.69
	Apr.	98.56	99.09	98.83	95.34	92.81	94.06	92.90
	May.	98.09	99.21	98.65	95.85	90.39	93.04	91.75
	Jun.	98.14	98.95	98.54	94.55	90.69	92.58	91.15
	Jul.	97.71	98.55	98.12	92.45	88.50	90.43	88.59
	Aug.	97.45	98.66	98.05	92.91	87.18	89.95	88.07
	W.Avg.	**98.32**	**98.94**	**98.63**	**94.56**	**91.58**	**93.04**	**91.69**
Multi-Layer Perceptron	Feb.	98.75	98.82	98.78	94.09	93.77	93.93	92.71
	Mar.	98.49	99.19	98.84	95.84	92.43	94.10	92.97
	Apr.	98.70	99.14	98.92	95.60	93.52	94.55	93.48
	May.	98.12	99.29	98.70	96.25	90.54	93.31	92.07
	Jun.	98.25	98.93	98.59	94.53	91.25	92.86	91.47
	Jul.	97.98	98.73	98.35	93.44	89.88	91.63	90.00
	Aug.	96.99	98.76	97.87	93.23	84.81	88.82	86.83
	W.Avg.	**98.36**	**99.03**	**98.69**	**95.02**	**91.78**	**93.36**	**92.09**

Fonte: Can, et al. (2020), p. 166101

Ora, observando os resultados obtidos na primeira análise, deparamo-nos que na categoria de percentagem de fraude, nas várias métricas de classificação, o algoritmo MLP foi o que geralmente obteve os melhores resultados. Contudo, o segundo melhor algoritmo, o RF, obteve resultados bastante próximos do MLP, ultrapassando-o até na métrica de classificação *Recall*.

Por outro lado, quando analisamos a testagem dos algoritmos na segunda análise realizada pelos autores, observamos que os papeis inverteram-se. O algoritmo RF, foi o que obteve os melhores resultados no panorama geral. Contudo, observamos que o algoritmo MLP na métrica de classificação *Recall* foi o que obteve o melhor resultado.

Por último, os autores decidiram também analisar os vários algoritmos tendo em conta o tipo de cartão de crédito, montante gasto, características das transações não fraudulentas

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

e fraudulentas, e tempo de processamento de cada algoritmo. Porém, por questões de análise comparativa, e não terem estes tópicos sido abordados com tanta profundidade nos artigos anteriores, decidimos não abordar esta temática.

3.8 *An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine*

Neste estudo elaborado pelos autores Taha, *et al.* (2020), a migração do negócio para a Internet e as transações eletrónicas que cada vez mais substituem o formato físico de pagamento como outrora conhecemos, fizeram com que a segurança transacional e eficácia da deteção de fraude fosse um fator essencial. Perante os desafios que o ML enfrenta na deteção destas transações, os autores para além de compararem vários algoritmos de ML, abordam um modelo próprio (LightGBM com *Bayesian-based Hyperparameter*) que acreditam ser o mais eficaz para detetar e mitigar este problema atual. Face ao exposto analisemos o presente estudo.

Em primeiro lugar, os autores abordaram dois *datasets* diferentes:

- Um primeiro *dataset*, contendo um total de 284,807 mil transações, sendo 284,315 legitimas e 492 fraudulentas, estruturada por 31 variáveis. As transações fraudulentas representam 0.172% do número total de transações. Relativamente às variáveis, 28 destas foram anonimizadas (V1, …, V28), possuindo ainda a coluna "*Time*" e "*Amount*" que não sofreram qualquer anonimização.
- Um segundo *dataset* com 94,683 mil transações no total, sendo 92,589 transações legitimas, e 2.094 fraudulentas, estruturada por 20 variáveis.

Tanto um *dataset* como o outro, possuem uma variável conotada de "*Class*" que classifica em código binário (0 e 1), se estamos perante uma transação fraudulenta ou não.

Relativamente à seleção de variáveis de cada *dataset*, este é um critério fundamental especialmente quando possuímos um grande número de variáveis que podem influenciar os resultados dos algoritmos em análise. No caso do LightGBM, algoritmo proposto pelos autores, este utiliza a técnica de IG, em que seleciona as variáveis mais importantes, com

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

intuito de melhorar o desempenho do modelo. O IG, como já referido anteriormente em outros artigos, tem como finalidade extrair as semelhanças entre as transações e realçar aquelas com um maior peso na classe das transações legitimas e transações fraudulentas.

É importante realçar que no modelo proposto pelos autores, o LightGBM sofreu uma híper parametrização, que tem como finalidade otimizar a performance do algoritmo, determinando por exemplo o parâmetro do número de folhas por árvore, o ratio de aprendizagem, etc, sendo estas características delimitadas pelos autores em prol da otimização.

Salientamos também a necessidade dos autores de dividirem os dados. Desde logo, para não prejudicar a performance dos algoritmos de ML, os autores decidiram abordar a técnica de *K-Fold Cross Validation* , neste caso de 5-Fold, ou seja, dividiram aleatoriamente cada *dataset* em 5 subpartes, em que 20% de cada *dataset* ficará reservada para a validação dos dados, para testar a performance do algoritmo, e os outras restantes 4 partes ou 80% são alocadas ao conjunto de treino. Repetimos este processo 5 vezes, até que cada subparte seja utilizada. Posteriormente calculamos a média das 5 subpartes, e o resultado corresponde à performance do modelo/algoritmo em análise num 5-*Fold Cross Validation*.

Em termos de métricas de classificação, os autores utilizaram a *Precision*, *Recall*, *Accuracy*, AUC e o F1-Score.

Não obstante, os autores decidiram demonstrar a performance da técnica 5-*Fold Cross Validation* , nos dois diferentes *datasets*, utilizando as métricas de classificação supra (Tabela 3.21).

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Tabela 3.21 - Técnica 5-Fold Cross Validation nos diferentes datasets, aplicando as várias métricas de classificação.

Data set	Fold No	AUC	Accuracy	Recall	Precision	F1 score
Data set 1	1	0.9116	0.9832	0.3788	0.9758	0.5458
	2	0.9090	0.9833	0.3824	0.9760	0.5494
	3	0.9066	0.9829	0.3647	0.9770	0.5306
	4	0.9098	0.9829	0.3665	0.9707	0.5316
	5	0.9100	0.9877	0.5373	0.9675	0.6903
	AVG	**0.9094**	**0.9840**	**0.4059**	**0.9734**	**0.5695**
Data set 2	1	0.9428	0.9840	0.2912	0.9531	0.4461
	2	0.9326	0.9838	0.2936	0.9226	0.4453
	3	0.9275	0.9834	0.2792	0.9092	0.4269
	4	0.9141	0.9834	0.2768	0.9061	0.4239
	5	0.9281	0.9833	0.2760	0.8950	0.4217
	AVG	**0.9290**	**0.9835**	**0.2833**	**0.9172**	**0.4327**

Fonte: Taha et al. (2020), p. 25583

Quanto aos algoritmos testados pelos autores, os utilizados foram os seguintes:

- Regressão Logística (RL);
- Support Vector Machine (SVM);
- K-Nearest Neighbour (KNN);
- Decision Tree (DT);
- Random Forest (RF);
- Naives Bayes (NB):
- LightGBM com Bayesian Hyper-parameter: modelo proposto pelos autores;

Obtendo os resultados apresentados na Tabela 3.22.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Tabela 3.22 - Testagem dos algoritmos de ML nos dois diferentes datasets.

Data set	Approach	AUC	Accuracy	Recall	Precision	F1-score
Data set 1	Log Reg	0.7020	0.9685	0.0101	0.0015	0.0030
	SVM Rbf	0.6910	0.9733	0.8695	0.0458	0.0871
	SVM Linear	0.4780	0.9709	0.0106	0.0021	0.0014
	KNN	0.5930	0.9690	0.1498	0.1123	0.1284
	DT	0.6630	0.9560	0.0799	0.4587	0.1375
	RF	0.8690	0.9779	0.6547	0.2522	0.3642
	NB	0.6320	0.8500	0.0111	0.0045	0.0064
	Our Approach	**0.9094**	**0.9840**	**0.4059**	**0.9734**	**0.5695**
Data set 2	Log Reg	0.8570	0.9781	0.7457	0.0839	0.1509
	SVM Rbf	0.7410	0.9781	0.8214	0.0438	0.0833
	SVM Linear	0.7090	0.9778	0.1075	0.0231	0.0833
	KNN	0.7150	0.9756	0.5507	0.2175	0.3119
	DT	0.7130	0.9667	0.3725	0.4312	0.4151
	RF	0.9070	0.9789	0.8168	0.3148	0.4545
	NB	0.8480	0.9592	0.1475	0.1812	0.1626
	Our Approach	**0.9288**	**0.9835**	**0.2833**	**0.9172**	**0.4327**

Fonte: Taha et al. (2020), p. 25585

Como é possível observar na tabela supra, o modelo proposto pelos autores obteve o melhor desempenho que os restantes algoritmos de ML. Contudo, podemos observar que o valor de *Recall* é bastante baixo em ambos os *datasets*, ou seja, a sua capacidade de detetar os verdadeiros positivos, ou aquelas transações que não são realmente fraude. Isto representa um grave problema, especialmente quando deparados com uma instituição financeira. Salientamos também que nenhum dos algoritmos, em ambos os *datasets*, possui valores altos em todas as métricas de classificação, apresentando sempre uma quebra em alguma métrica.

Por último, os autores também decidiram analisar a métrica AUC, comparando com os diferentes algoritmos anteriormente referidos, acrescentado também uma comparação entre o modelo proposto (LightGBM com Bayesian Hyper-parameter), LightGBM sem

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

qualquer parametrização e o algoritmo CatBoost, obtendo os seguintes resultados (Figura 3.10) (Tabela 3.23):

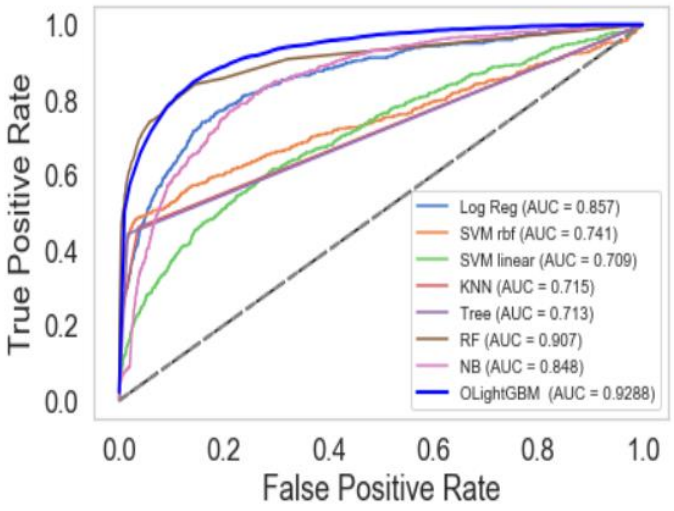

Figura 3.10 - Comparação dos diferentes algoritmos com a métrica de classificação AUC

Fonte: Taha et al. (2020), p. 25585

Tabela 3.23- Comparação do modelo proposto, do LightGBM e Catboost utilizando a métrica de classificação AUC

Approach	AUC
LightGBM	90.62%
Catboost	87.86%
Proposed approach	92.88%

Fonte: Taha et al. (2020), p. 25585

Como podemos observar, o modelo proposto pelos autores obteve 92.88% de AUC, sendo o melhor resultado quando comparado tanto com os algoritmos supra. Segundo os autores, estes resultados demonstram a importância de adotar um método eficiente de otimizar parâmetros, já que melhora consideravelmente os resultados atingidos (2.26% de diferença, quando comparado o modelo proposto com o algoritmo LightGBM sem qualquer parametrização).

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

4 ANÁLISE COMPARATIVA

Neste capítulo apresenta-se a análise comparativa realizada aos estudos identificados no capítulo anterior, com apresentação de tabelas resumo sobre os *datasets*, algoritmos e métricas utilizadas e tempos de execução dos algoritmos. Além das tabelas resumo que permitem sumariar e comparar os resultados obtidos apresenta-se ainda uma discussão sobre os resultados obtidos na deteção de fraude transacional em contexto bancário.

4.1 *Datasets* utilizados

A Tabela 4.1 resume todos os *datasets* encontrados na RSL.

Tabela 4.1 - Datasets utilizados nos registos analisados

Nome dos Registos Bibliográficos	Autores	Ano do registo	Nº Datasets Utilizados	Nº Total de transações nos Datasets	Nº de Variáveis	Transações Legítimas	Transações Fraudulentas	Divisão dos Datasets
A Closer Look into the Characteristics of Fraudulent Card Transactions	Can, et al. (2020)	2020	1	4.000.000.000	60	3.999.755.000	245.000	70% conjunto treino 30% conjunto teste
A Neural Network Ensemble with Feature Engineering for Improved Credit Card Fraud Detection	Esenogho, et al. (2022)	2022	1	284807	30	284.315	492	10-Fold Cross Validation
An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine	Taha et al. (2020)	2020	2	284.807 94.683	31 20	284.315 92.589	492 2.094	5-Fold Cross Validation
Efficient Resampling for Fraud Detection During Anonymized Credit Card Transactions with Unbalanced Datasets	Mrozek, et al. (2020)	2020	1	284.807	30	284.315	492	5-Fold Cross Validation
Exploratory analysis and implementation of machine learning techniques for predictive assessment of fraud in banking systems	Lellis, et al. (2022)	2022	1	6.362.620	11	6.354.407	8.213	70% conjunto treino 30% conjunto teste
Fraud Detection in Banking Data by Machine Learning Techniques	Hashemi, et al. (2022)	2022	1	284.807	30	284.315	492	5-Cross Fold Validation
Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost	Ileberi, et al. (2021)	2021	2	284.807 24.357.143	30; (Sem informação)	284.315 24.357.143	492 23.757	(Sem informação)
Sequential credit card fraud detection: A joint deep neural network and probabilistic graphical model approach	Forough, et al. (2020)	2020	1	284.807 374.823	30 17	284.315 360.792	492 14.031	70% Conjunto Treino 20% Conjunto Teste 10% Validação

Fonte: Elaboração própria

Pensaríamos que quanto maior o *dataset*, maior seria o número de transações fraudulentas, contudo podemos comprovar que tal pensamento, não corresponde à

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

verdade. Desde logo, se compararmos o *dataset* analisado pelos autores Forough, et. al. (2020), das 373.823 mil transações analisadas, 14.031 mil constituíram transações fraudulentas, o que, por outro lado, o *dataset* dos autores Lellis et.al. (2022), das 6.354.407 milhões de transações analisadas, apenas 8.213 mil transações foram consideradas como fraudulentas. O mesmo acontece com o *dataset* dos autores Taha et. al. (2020) que com 92.589 transações analisadas, 2.094 mil foram transações fraudulentas, um valor bastante superior, quando comparando com o *dataset* dos autores Lellis, et. al. (2022), ou até com o *dataset* dos autores Esenogho, et. al. (2022), que entre as 284.807 mil transações analisadas, 492 transações constituíram fraude.

Necessário salientar uma particularidade deste último *dataset* de 284.807 mil transações, na medida em que o mesmo é abordado em 6 dos 8 registos da Tabela 4.1. Numa análise mais aprofundada e na tentativa de perceber o porquê da sua constante utilização, deparamo-nos que este mesmo provém de fonte pública ou *Open Data*, mais concretamente da plataforma Kaggle, sendo estudado e escrutinado por vários investigadores, com intuito de compreender como se comportam os vários algoritmos de deteção de fraude quando aplicados a este *dataset*. Para além disso, este mesmo *dataset*, encontra-se anonimizado, necessário face à problemática dos dados sensíveis contidos nos cartões de crédito, e as informações pessoais dos vários clientes das instituições financeiras. Daí a dificuldade em encontrar um *dataset* com dados reais, sem violar nenhuma violação legal, possivelmente respondendo à questão de ter sido tão usado por diversos autores. Aliado a esta razão, o facto de ser *Open Source*, e já ter sido alvo de escrutínio por outros usuários da plataforma Kaggle, permite uma comparação com outros trabalhos e projetos.

Como podemos verificar na Tabela 4.1, a divisão deste mesmo *dataset* é realizada de diferentes formas, seja por *10-Fold Cross Validation*, *5-Fold Cross Validation*, 70% conjunto treino, 20% conjunto teste e 10% de validação, e numa situação, nem possuímos informação quanto à respetiva divisão. O mesmo iremos verificar posteriormente na Tabela 4.2, que tanto as técnicas de equilíbrio dos dados, assim como os algoritmos utilizados, apesar de algumas semelhanças, diferem na generalidade, mesmo com a

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

escolha deste *dataset*. Face ao exposto, podemos concluir que a análise e escolha do pódio dos diferentes algoritmos não será colocada em causa pela escolha deste *dataset* pelos diferentes autores.

4.2 Algoritmos utilizados

A Tabela 4.2, apresenta o resumo dos algoritmos identificados nos registos analisados resultantes da RSL.

Tabela 4.2 - Algoritmos utilizados nos registos analisados

Nome dos Registos Bibliográficos	Técnicas de Equilibrio de Dados	Métricas de Classificação	Algoritmos analisados	Algoritmos seleccionados
A Closer Look into the Characteristics of Fraudulent Card Transactions	Aleatório com um Ratio 5:1	Precision Recall F1-Score AUC-ROC AUC-PR	NB AD/DT RF MLP	**MLP** **RF**
A Neural Network Ensemble with Feature Engineering for Improved Credit Card Fraud Detection	SMOTE-ENN	Recall Specificity AUC	AdaBoost LSTM DT/AD SVM MLP LSTM Ensemble	**LSTM Ensemble**
An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine	Desiquilirado e posterior aplicação do 5-Fold Cross Validation	Precision Recall Accuracy AUC F1-Score	RL SVM KNN DT/AD RF NB LightGBM com Bayesian Hyper-parameter	**LightGBM com Bayesian Hyper-parameter**
Efficient Resampling for Fraud Detection During Anonymized Credit Card Transactions with Unbalanced Datasets	RUS SMOTE	AUC F1-Score Precision Recall Accuracy	RL RF KNN SGD	**RF** **RL**
Exploratory analysis and implementation of machine learning techniques for predictive assessment of fraud in banking systems	RUS SMOTE ADASYN	AUC Accuracy	RL KNN NB Perceptron	**RL** **KNN**
Fraud Detection in Banking Data by Machine Learning Techniques	Bayesian Optimization	Accuracy Precision Recall F1-Score AUC MCC	RL LightGBM XGBoost CatBoost MVEL ANN	**LightGBM-XGBoost** **XGBoost**
Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost	SMOTE SMOTE-Adaboost	Accuracy Recall Precision MCC AUC	SVM RL DT/AD RF XGBoost ET AdaBoost	**RF-AdaBoost** **ET-AdaBoost** **DT-AdaBoost**
Sequential credit card fraud detection: A joint deep neural network and probabilistic graphical model approach	Seq-US Random-US ADASYN-OS SMOTE-OS	Precision Recall F1-Score AUC-ROC AUC-PR	HMM MEM CRF RNN LSTM GRU ANN LSTM-CRF	**LSTM-CRF**

Fonte: Elaboração própria

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Quanto aos algoritmos e respetivas técnicas utilizadas ao longo dos diferentes registos, segundo a Tabela 4.2, selecionamos os melhores algoritmos de acordo com as conclusões finais elaboradas por cada um dos autores, tendo em consideração cada registo individualmente. Isto é, analisamos cada um dos registos, observamos as conclusões de cada autor desse registo, e salientamos na coluna de "Algoritmos selecionados" os algoritmos que foram indicados como os melhores, por esses mesmos autores.

Somos da opinião, que esta é a melhor análise a ser feita, já que cada registo, segue variáveis diferentes, *datasets* diferentes (pelo menos em dois registos), assim como técnicas de equilíbrio, e os próprios algoritmos analisados são distintos dos demais.

Neste sentido, perante esta análise individual de cada registo, realizamos uma contagem dos algoritmos selecionados na totalidade dos registos e atribuímos assim o respetivo pódio. Contudo, enfrentamos uma problemática na respetiva contagem. Deveríamos contabilizar o algoritmo LSTM Ensemble e LSTM-CRF como um único algoritmo LSTM? O mesmo acontece com o algoritmo LightGBM-XGBoost e o LightGBM com Bayesian Hyperparameter, na sua ponderação como apenas como algoritmo LightGBM. Assim como no caso do algoritmo RF, como RF-Adaboost. São algoritmos diferentes e devem ser contabilizados apenas como um algoritmo? Ou apesar de estar combinado (híbrido) com outro algoritmo, ou parametrizado, devemos considerar como algoritmos distintos?

Ora, confrontando cada um dos registos de que são alvo estes algoritmos, conseguimos verificar na Tabela 4.2, que no caso do LightGBM com *Bayesian Hyperparameter*, e o LightGBM-XGBoost, apesar de sofrerem ambos parametrização, no primeiro caso, os dados foram equilibrados com *5-Fold Cross Validation*, enquanto o segundo consiste num algoritmo híbrido, em que diferença entre ambos, não se encontra meramente na técnica de equilíbrio dos dados, mas no próprio algoritmo em si. Neste sentido, somos da opinião de tratar estes algoritmos como diferentes pelas suas características intrínsecas aplicadas ao caso concreto de cada registo.

Face a esta conclusão, conseguimos observar pela Tabela 4.2, que os algoritmos RF e RL são maioritariamente escolhidos dentre os registos, contudo não é uma vantagem clara e

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

incontestável, por apenas terem sido selecionados duas vezes em diferentes registos. Neste sentido, se considerarmos todos como algoritmos únicos e diferentes dos demais, então todos, com exceção do RF e RL, foram escolhidos pelo menos uma vez.

Poderíamos concluir pela atribuição do pódio a estes dois algoritmos, contudo, será que podemos atribuir o pódio dos "melhores" já que esta vantagem não é clara, e em que de acordo com a situação em concreto (variáveis, *dataset*, técnicas utilizadas, etc.), este mesmo pódio poderá alterar-se?

Somos da opinião que outras particularidades terão de ser tidas em consideração na escolha do pódio dos algoritmos, especialmente aquelas que deveriam ser tidas como essenciais quando analisamos estes mesmos algoritmos na perspetiva bancária. Neste sentido, analisemos.

4.3 Métricas de classificação

O modelo de deteção de fraude bancário, como podemos observar na Figura 1.5, consiste num modelo híbrido, que para além da necessidade de ter de conseguir detetar corretamente as transações fraudulentas ou *True Positive* (TP), tem de conseguir possuir um valor baixo de falsos positivos ou *False Positive* (FP), isto é, transações que foram consideradas como fraudulentas, mas que na verdade não o eram. Esta necessidade ocorre da parte manual/humana, o valor tem de ser baixo o suficiente para atribuir a estas equipas o tempo necessário para mitigar, investigar e descartar ou não, a respetiva transação como fraudulenta ou não.

Se compararmos na Tabela 4.2 a métrica de classificação mais usada, iremos verificar que a AUC foi claramente a mais predominante. Esta consiste numa métrica de performance bastante conhecida, que tem como objetivo representar graficamente a capacidade do classificador binário, através do ratio dos verdadeiros positivos (TPR) sobre o ratio dos falsos positivos (FPR).

Contudo, num *dataset* desequilibrado como é o caso da fraude, a curva ROC pode esconder a má performance do algoritmo em causa. Há quem defenda que a utilização do

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

AUC-PR em vez do AUC, já que o primeiro é mais sensível ao desequilíbrio dos dados que o segundo (Makki, 2019, p. 93016).

Mas então qual será a certa? Ou a mais adequada para a particularidade de um *dataset* de fraude, aplicada à realidade bancária?

A performance de um classificador pode ser medida de várias formas como podemos verificar pela Tabela 4.3:

Tabela 4.3 - Métricas de classificação e respetivos campos de atuação

Metric	Description	
Accuracy Detection rate	Overall correctness of the classifier	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision Hit rate	When a fraud is detected, how often was the detection correct? i.e.: TP/predicted yes	$\frac{TP}{TP+FP}$
Recall TPR Sensitivity	When there was a fraud, how often did the classifier manage to detect it? i.e.: TP/actual yes	$\frac{TP}{TP+FN}$
FPR	Ratio of fraud detected incorrectly i.e.: FP/actual no	$\frac{FP}{FP+TN}$
F1 F-score F-measure	Harmonic mean of precision and recall	$\frac{2 \cdot sensitivity \cdot precision}{sensitivity + precision}$
G-mean	balance between classification performances on both the majority and minority classes	$\sqrt{sensitivity \cdot specificity}$
AuROC	Area under the ROC (graph of all possible true and false positive hit rates)	$\int ROC$
Cohen Kappa	Score of the agreement between the prediction and the actual	$\frac{total\ accuracy - random\ accuracy}{1 - random\ accuracy}$
Mathew	Score of the correlation between the prediction and the actual	$\frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP \cdot FP)(TP \cdot FN)(TN \cdot FP)(TN \cdot FN)}}$

Fonte: Adaptado de Kulatilleke, 2017, p. 5

Ora, um elevado nível de TP/TPR (*Precision*/*Recall*) é sempre desejável, contudo nem sempre é possível, já que só atingimos uma elevada *Precision* à custa de um baixo nível de *Recall* e vice-versa, verificando uma relação invertida (Kulatilleke, 2017, p.5).

Ao contrário da *Precision*, o *Recall* é independente do número de amostras negativas, isto é, se o modelo classificar todas as amostras como positivas, o *Recall* será 1.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Neste sentido, que a *Precision* por ela própria, seria a melhor nas situações de fraude bancária, já que tem consideração o número de falsos positivos (Kulatilleke, 2017, p.5). Mas isto por si só não é verdade, nenhum destes métodos pode ser usado exclusivamente para confirmar a competitividade e qualidade dos respetivos algoritmos (Makki, 2019, p. 93016).

Consequentemente havendo uma *Precision* alta e um *Recall* baixo, significa que o modelo é bom a prever a classe positiva, contudo apenas prevê uma pequena percentagem dos resultados positivos possíveis, o modelo estaria a prever fracamente. Contrariamente significaria que uma *Precision* baixa e um *Recall* alto, apesar do modelo prever corretamente a classe positiva, previu também um número avultado de negativos como positivos. Ora, aplicando esta realidade ao contexto bancário, apesar da *Precision* ser um bom indicador na deteção de fraude, numa componente de gestão das equipas internas bancárias, deixaria algo a desejar, já que o número avultado de transações classificadas como positivas que na verdade são falsas, seria incomportável. Neste sentido, o ideal será sempre valor elevados em ambos os classificadores. Poderíamos também recorrer ao *F1-Score* para originar uma média entre estes dois classificadores em último ratio.

A verdade é que quando comparando com a Tabela 4.2, de entre os 8 registos, 7 utilizaram juntamente o *Recall* e *Precision*, e apenas 5 o F1-Score. Somos da opinião que o registo "*Exploratory analysis and implementation of machine learning techniques for predictive assessment of fraud in banking systems",* peca por não representar a classificação do *Recall* e *Precision*, mesmo demonstrando o comportamento do AUC, o ratio do TP e FN não parece ser suficiente para classificação da performance do algoritmo RL como o indicado para a deteção de fraude, especialmente quando aplicamos esta realidade ao contexto bancário.

Não obstante, o algoritmo RL foi escolhido por outros registos, não desmerecendo o seu valor enquanto algoritmo de eleição. Neste sentido, mantemos as conclusões tidas anteriormente quanto ao pódio dos dois algoritmos da Tabela 4.2.

Por último, quando ao facto do AUC-PR ser segundo alguns autores, preferível em contexto de dados desequilibrados como o caso dos *datasets* de fraude, não tivemos em

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

consideração tal afirmação, já que nos 8 registos selecionados, apesar de algumas referências a randomização dos dados ou não equilíbrio dos mesmos, foram utilizadas técnicas de equilíbrio ao longo de todos os registos. Neste sentido, não criticamos a escolha da métrica AUC como a mais escolhida pelos respetivos registos.

4.4 O tempo de processamento

O tempo no caso da deteção de fraude transacional é fulcral e necessário em duas perspetivas diferentes. Em primeiro lugar, o algoritmo de deteção de fraude, tem de ser capaz de detetar milhões de transações diariamente, de uma forma célere e definir as mesmas como fraudulentas ou não, de uma forma eficaz (tempo na ótica de rápido processamento), e a consequente necessidade de um baixo número de falsos positivos, para possibilitar a correta investigação das equipas internas, sem a influência da pressão do avultado número de transações (tempo operacional) (Figura 1.5).

Nas transações bancárias é essencial o tempo de execução dos respetivos algoritmos. Para além da sua capacidade na deteção de fraude transacional, a morosidade de execução e a capacidade de processar quantidades avultadas de transações diariamente, faz com que, o tempo, constitua uma característica da performance do respetivo algoritmo (Can, 2020, p. 166016-166017).

Ora, se formos a confrontar os 8 registos da RSL, apenas um, fez referência a esta problemática (Tabela 4.4). Ora, confrontando a Tabela, os autores *Can et al* (2020), verificaram que o tempo de análise variava consoante cada algoritmo analisado, sendo o melhor algoritmo a DT com 0.06 milésimos de segundo, e em último o algoritmo RF com 5.54 milésimos de segundo, perfazendo uma média de 6 milésimos de segundo o tempo de processamento de cada transação, utilizando como referência 15 milhões de transações (valor diário estipulado pelos autores). Não obstante, os autores também decidiram verificar o tempo de execução de cada um dos algoritmos, apenas com 1 milhão de instâncias ou transações, verificando um tempo médio de 0.8 milésimos de segundo por transação.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Pode parecer pouco tempo se analisarmos cada transação individualmente. Mas se formos a ter em consideração o funcionamento operacional de uma equipa interna bancária, poderemos dizer o mesmo?

Ora, se tivermos em conta a média de tempo de execução tendo como referência 15 milhões de transações, 6 milésimos de segundo corresponde a 0.006 segundos, que após multiplicarmos por 15 milhões irá originar, 90000 segundos. Se convertermos este último valor em minutos/horas, iremos obter 1500 minutos ou 25 horas.

Neste sentido, se tivermos em consideração o *dataset* mais utilizado dos 8 registos (Tabela 4.1), em que das 284.807 transações, 492 foram consideradas fraudulentas, podemos concluir (com as devidas limitações), que o número de transações fraudulentas levadas à devida investigação por parte de uma equipa interna, será bastante penoso ou até insustentável.

Como tal, comporta um verdadeiro desafio determinar um algoritmo de deteção de fraude com a capacidade suficiente de selecionar as transações fraudulentas sem colocar em causa o correto funcionamento dos serviços bancários.

Salientar que o tempo transacional pode variar, para menor ou maior tempo, tendo em conta as variáveis, as técnicas utilizadas no equilíbrio dos dados, e o *hardware* respetivo, adotado por cada instituição bancária. Contudo, apenas com a referência de um único artigo relativamente a esta problemática na bibliografia selecionada, não é possível realizar uma correta comparação de algoritmos, ficando apenas o alerta da necessidade desta problemática para futuros estudos.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Tabela 4.4 - Tempos de execução e hardware utilizado nos registos analisados

Nome dos Registos Bibliográficos	Autores	Algoritmos analisados	Algoritmos seleccionados	Tempo de execução dos algoritmos	Hardware utilizado
A Closer Look into the Characteristics of Fraudulent Card Transactions	Baris Can; Ali Gokhan Yavuz; Elif. M. Karsligil; M. Amac Guvensan;	NB AD/DT RF MLP	**MLP** **RF**	NB = 0.52 ms RF = 5.54 ms DT = 0.06 ms MLP =1.43 ms	IBM Power 9 160 CPU 1.1 TB memória
A Neural Network Ensemble with Feature Engineering for Improved Credit Card Fraud Detection	Ebenezer Esenogho; Ibomoye Domor Mienye; Theo G.Swart; Kehinde Aruleba; George Obaido;	AdaBoost LSTM DT/AD SVM MLP LSTM Ensemble	**LSTM Ensemble**	Não foi realizado	Não foi realizado
An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine	Altyeb Altaher Taha; Sharaf Jameel Malebary;	RL SVM KNN DT/AD RF NB LightGBM com Bayesian Hyper-parameter	**LightGBM com Bayesian Hyper-parameter**	Não foi realizado	Não foi realizado
Efficient Resampling for Fraud Detection During Anonymized Credit Card Transactions with Unbalanced Datasets	Petr Mrozek; John Panneerselvam; Ovidiu Bagdasar;	RL RF KNN SGD	**RF** **RL**	Não foi realizado	Não foi realizado
Exploratory analysis and implementation of machine learning techniques for predictive assessment of fraud in banking systems	Miguel Ângelo de Lellis; Claudio de Souza Rocha Junior; Diogo Ferreira de Lima Silva; Marcos Pinto de Castro Junior; Igor Pinheiro de Araújo Costa; Carlos Francisco Simões Gomes; Marcos dos Santos;	RL KNN NB Perceptron	**RL** **KNN**	Não foi realizado	Não foi realizado
Fraud Detection in Banking Data by Machine Learning Techniques	Seyedeh Khadijeh Hashemi; Seyedeh Leili Mirtaheri; Sergio Greco;	RL LightGBM XGBoost CatBoost MVEL ANN	**LightGBM-XGBoost** **XGBoost**	Não foi realizado	Não foi realizado
Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost	Emmanuel Ileberi; Yanxia Sun; Zenghui Wang;	SVM RL DT/AD RF XGBoost ET AdaBoost	**RF-AdaBoost** **ET-AdaBoost** **DT-AdaBoost**	Não foi realizado	Não foi realizado
Sequential credit card fraud detection: A joint deep neural network and probabilistic graphical model approach	Javad Forough; Saeedeh Momtazi;	HMM MEM CRF RNN LSTM GRU ANN LSTM-CRF	**LSTM-CRF**	Não foi realizado	Não foi realizado

Fonte: Elaboração própria

4.5 Discussão

Partilhamos a opinião que apenas poderemos responder à questão controversa, de qual ou quais ou melhores algoritmos de deteção de fraude na perspetiva bancária, se tivermos em consideração as particularidades que a banca, enquanto instituição, enfrenta.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Face às problemáticas e particularidades que a banca enfrenta, será que podemos atribuir na mesma o pódio aqueles dois algoritmos (RF e RL - Tabela 4.2) Será que as métricas de classificação utilizadas pelos diferentes autores, são as mais indicadas tendo em conta a perspetiva bancária? E haveria outros fatores que deveriam ter sido analisados pelos autores? Analisemos.

Os 8 registos da RSL tentam, com diferentes *datasets*, variáveis, técnicas de equilíbrio de dados, técnicas de divisão dos respetivos *datasets*, e métricas de classificação, encontrar o melhor, ou melhores algoritmos de deteção de fraude transacional. Acontece que, apesar de 3 ou 4 algoritmos serem predominantemente escolhidos pela bibliografia, e escolherem maior parte deles, métricas de classificação aconselhadas anteriormente, não podemos concluir, nem confirmar, que estes são os melhores para detetar fraude, pelo menos na perspetiva bancária.

Ora, como conseguimos observar na Tabela 4.1 e na Tabela 4.4, os *datasets* possuem dimensões diferentes, e foram abordados de formas diferentes, em que apenas um dos registos teve em consideração a possível problemática dos modelos híbridos de gestão de fraude, sendo esta a gestão de fraude operacional que ocorre transversalmente nas instituições bancárias. Somos da opinião, que apesar de toda bibliografia ser benéfica para aprofundar o conhecimento sobre esta problemática, pouca é aquela que analisa intrinsecamente as dificuldades práticas na deteção de fraude diária por parte destas instituições, e consequentemente as suas equipas internas. Podemos verdadeiramente afirmar que estes 2 algoritmos (RF e RL - Tabela 4.2) são os melhores de deteção de fraude na perspetiva bancária? Podemos concluir que no caso concreto de cada um dos registos, demarcaram-se pela sua excelente capacidade de deteção de fraude transacional, contudo, apenas esta mera análise, não é suficiente para responder à nossa questão. Seria uma afirmação, no mínimo dúbia, sem primeiro aprimorar a eventual componente prática. A resposta a esta questão, aguarda uma maior investigação e escrutínio aplicado à realidade bancária.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

5 CONCLUSÃO

Nesta dissertação tentámos responder à questão controversa: Qual ou quais os melhores algoritmos de deteção de fraude transacional em contexto bancário? Infelizmente, essa resposta não é, na nossa opinião, alcançada.

Contudo, isto não é o fim, mas sim o início. Atualmente estamos a entrar na era dourada da análise preditiva. A era de extrair informação dos dados e atribuir um sentido, está a tornar-se, cada vez mais, no foco principal de muitos negócios (Subramanian, 2014, p. 161), e a banca não é exceção.

Cada vez surgem mais instituições financeiras que, partilham a necessidade de utilizar a IA para combater o crime financeiro. Muitas destas, recorrem a empresas especializadas na área, como é o caso da empresa portuguesa, Feedzai. Esta última, possui um modelo capaz de atribuir um score de risco e automaticamente identificar crime financeiro, em tempo real em menos de três milésimos de segundo, mas para além disso, possuem ferramentas próprias capazes de investigar padrões e ter em consideração as recomendações dadas pelos operacionais/analistas, para detetar a prática do crime financeiro que o modelo anterior não foi capaz de fazer (Ferreira, J. 2019).

Observamos assim, vários formas de abordar a problemática do crime financeiro, mais concretamente a fraude transacional. Pode ser implementada internamente nas instituições bancárias, possuindo um corpo laboral e mecanismos próprios para o efeito, ou recorrer a empresas especializadas, dedicadas única e exclusivamente a estas problemáticas.

Também é necessário realçar que, na presente dissertação a análise das transações fraudulentas foi realizada sem levantar a particularidade do tempo-real (apenas algumas referências, sem aprofundar a questão), mas sim, transações em bloco ou *batch*, e mesmo esta última, apenas um registo teve esta característica em consideração. Neste sentido, como seria a performance do tempo de execução dos 8 registos analisados seria a mesma em tempo-real e em bloco? Qual seria o mais vantajoso? Outros algoritmos ocupariam o pódio? Apenas outras contribuições futuras nesta matéria o dirão.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

REFERÊNCIAS

Albrecht, S., Albrecht C., Albrecht C., Zimbelman, M. (2019). *Fraud Examination*. Cengage Learning, Inc.

Baesens, B., Vlasselaer V., Verbeke, W. (2015). *Fraud Analytics: Using Descriptive, Predictive, and Social Network Techniques – A Guide to Data Science for Fraud Detection*. Wiley.

B-on. (2023). https://www.b-on.pt/

Can, B., Yavuz, A., Karsligil, E., Guvensan, M. (2020). *A Closer Look Into the Characteristics of Fraudulent Card Transactions*. IEEE.

Charles Sturt University. (2023). *Literature Review: Systematic literature reviews*. https://libguides.csu.edu.au/review/Systematic

Deloitte. (2021). *Fraud Survey Portugal 2021: Clear and focused attention*. *https://www2.deloitte.com/content/dam/Deloitte/pt/Documents/fraudsurvey/Deloitte-Fraud-Survey-2021-Empresas.pdf*

Dill, A. (2020). *Bank Regulation, Risk Management, and Compliance – Theory, Practice, and Key Problem Areas*. Informa Law from Routledge

Dimensions.ai. (2023). https://www.dimensions.ai/

Esenogho, E., Mienye, I., Swart, T., Aruleba, K., Obaido, G. (2022). *A Neural Network Ensemble With Feature Engineering for Improved Credit Card Fraud Detection*. IEEE

Fernández, A., García, S., Galar, M., Prati, R., Krawczyk B., Herrera, F. (2018*), Learning from Inbalanced Data Sets*. Springer.

Ferreira, J. (2019). *How to Detect Fraud in Less Than Three Milliseconds*. Feedzai. https://feedzai.com/blog/how-to-detect-fraud-in-less-than-three-milliseconds/

Forough, J. & Momtazi S. (2020). *Sequential credit card fraud detection: A joint deep neural network and probabilistic graphical model approach*. John Wiley & Sons Inc.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Gee, S., (2015). *Fraud and Fraud Detection – A Data Analytics Approach*. John Wiley & Sons Inc.

Hashemi, S., Mirtaheri, S., Greco S. (2022). *Fraud Detection in Banking Data by Machine Learning Techniques*. IEEE

Ileberi, E., Sun, Y., Wang Z. (2021). *Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost*. IEEE.

Isson, J. (2018). *Unstructured Data Analytics – How to Improve Customer Acquisition, Customer Retention, and Fraud Detection and Prevention*. John Wiley & Sons Inc.

Kassem, R. & Higson A. (2012). *The New Fraud Triangle Model*. Scholarlink Research Institute Journals.

Kuhn., S., N. (2022). *Técnicas de aprimoramento de equidade em aprendizado de máquina: Uma revisão sistemática de literatura*. (Dissertação de mestrado, Instituto Superior de Contabilidade e Administração de Coimbra, Coimbra, Portugal)

Kulatilleke, G. & Samarakoon, S. (2017). *Empirical study of Machine Learning Classifier Evaluation Metrics behavior in Massively Imbalanced and Noisy data*. (Dissertação de mestrado, Queen Mary University of London, Londres).

Makki, S., Assaghir, Z., Taher, Y., Haque, R., Hacid, M., Zeineddine, H. (2019). *An Experimental Study With Imbalanced Classification Approaches for Credit Card Fraud Detection*. IEEE

Megdad, M., Abu-Nasser, B., Abu-Nasser, S. (2022). *Fraudulent Financial Transactions Detection Using Machine Learning*. IJAISR

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. (2009). *Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement*. BMJ. https://www.bmj.com/content/339/bmj.b2535

Moreira, M., Junior C., Silva, D., Junior M., Costa, I., Gomes C., Santos, M. (2022). *Exploratory analysis and implementation of machine learning techniques for predictive assessment of fraud in banking systems*. Elsevier B.V.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Mrozek, P., Panneerselvam, J., Bagdasar, O. (2020). *Efficient Resampling for Fraud Detection During Anonymised Credit Card Transactions with Unbalanced Datasets*. IEEE.

Nesvijevskaia, A., Ouillade, S., Guilmin, P., Zucker, J. (2020). *The accuracy versus interpretability trade-off in fraud detection model*. Cambridge University Press.

Oxford Learner's Dictionaries. (2023). https://www.oxfordlearnersdictionaries.com/definition/american_english/fraud

Saporta, G. & Maraney S. (2022). *Pratical Fraud Prevention: Fraud and AML Analytics for Fintech and eCommerce, Using SQL and Python*. O'Reilly.

Subramanian, R. (2014). *Using Technology to Combat Losses*. John Wiley & Sons Inc.

Taha, A. & Malebary, S. (2020). *An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine*. IEEE.

Vousinas, G. (2019). *Advancing theory of fraud: the S.C.O.R.E model*. Journal of Financial Crime.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

APÊNDICES

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

APÊNDICE 1. Algoritmos de ML identificados no texto:

- **Regressão Logística** (RL): a RL consiste num modelo de regressão em que a variável de resposta Y é categórica. O RL permite-nos estimar a probabilidade de uma determinada resposta categórica, baseando-se em um ou mais variáveis preditoras X. Logo, estima a probabilidade de um acontecimento se realizar, isto é, a probabilidade de sucesso. Neste caso, estamos perante uma regressão logística binária, em que a probabilidade está contida no intervalo [0,1], em que a regressão logística usa a curva logística ou *sigmoid*, para representar a relação entre estas duas variáveis. Considerando a sua facilidade de interpretação e potencial de generalização, a RL tem sido largamente utilizada para a deteção de fraude transacional.
- ***K-Nearest Neighbours*** (KNN): consiste num algoritmo que é usualmente utilizado para a classificação e para a regressão. É um simples algoritmo que consiste em usar os K pontos mais próximos daquele que queremos prever. Portanto, é um algoritmo que depende da regra da distância para a classificação, usando os K vizinhos mais próximos de uma nova amostra. Mas como determinamos esta regra?

  Há duas formas:

  1) Votação Simples: A nova observação é anexada à classe da maioria dos K pontos vizinhos. Ou seja, se a maioria dos "vizinhos" votar para essa classe, a nossa observação é determinada de acordo com votação maioritária;
  2) Distância ponderada: Ocorre na mesma uma votação, contudo é atribuído um maior peso no voto aos vizinhos mais próximos do ponto que queremos prever.

  Portanto este algoritmo depende de três fatores: a distância é usada para determinar os vizinhos mais próximos; qual a regra da distância é usada para a classificação dos K vizinhos mais próximos; o número de vizinhos que são considerados para classificar a nova amostra.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Numa aplicação à deteção da fraude, o algoritmo KNN determina as classes em que as novas transações vão ser anexadas dependendo da regra, seja pelo ponto (K) da transação estar próxima de K vizinhos, ou a uma classe maioritária de K pontos vizinhos. Depende da regra de decisão que determine a nova transação (e o seu K ponto) como fraudulenta ou não.

- ***Naive Bayes*** (NB): consiste num algoritmo de deteção de fraude, que se baseia no Teorema de Bayes, em que toma uma determinada decisão de acordo com a maior probabilidade. A probabilidade de Bayes estima uma probabilidade desconhecida através de valores conhecidos. Também permite que o conhecimento e lógica prévia ser aplicada a casos incertos ou desconhecidos. Adaptando ao caso da fraude, estamos perante um método para determinar a probabilidade de uma determinada hipótese ser verdadeira, ou seja, um método que demonstra a probabilidade de existir fraude ou não.
- ***Perceptron***: consiste num algoritmo comumente utilizado nas classificações binárias, e que se assemelha e tentar replicar o neurónio humano. Ora num único neurónio, o modelo *Perceptron* deteta se a função é um input ou não, e classifica cada a função, atribuindo-a à classe respetiva. Existe dois tipos de *Perceptron*, o multicamadas e única camada. Neste caso, estamos perante um modelo único, com uma única camada, que é definida pela sua capacidade de classificar inputs lineares.
- ***Support Vector Machine*** (SVM): consiste num algoritmo de *Machine Learning* supervisionado, que pode ser tanto usado para casos de classificação ou regressão. Contudo, é mais utilizado nos casos de classificação. Neste algoritmo nós distribuímos cada dado de um determinado objeto como um ponto num espaço *n*-dimensional (em que *n* é onúmero de variáveis/classes que possuímos), sendo que o valor de cada variável, é ovalor de uma determinada coordenada. Então a ideia base do algoritmo SVM é dividir as classes dentro desse espaço *n*-dimensional, classificando assim os dados de treino, tendo em conta um determinado critério. Aplicando à situação da fraude transacional, o algoritmo

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

SVM é usado para categorizar as transações como fraudulentas ou não. O SVM analisa o padrão da transação anterior e quando uma nova transação ocorre, que desvia do padrão anterior ou do espaço n-dimensional que esse determinado objeto/transação tinha sido anteriormente colocado, esta nova transação é conotada como uma transação fraudulenta.

- ***Decision Tree/Árvore de Decisão*** (DT/AD): consiste numa representação gráfica de uma árvore, em que esta mesma é usada para tomar decisões. Foi criada como um classificador de aprendizagem indutiva, originando uma estrutura de uma árvore que tenta separar as classes em subgrupos mutuamente exclusivos em cada nó da árvore, até que determinadoobjetivo seja atingido.
- ***Random Forest*** (RF): consiste numa nova abordagem de DT quetentar resolver um problema de *overfitting* (usualmente causado pelos algoritmos iniciais de DT), através da combinação de várias DT utilizando vários subgrupos de conjunto treino e uma seleção aleatória dos mesmos. Logo, este último consiste numa randomização das várias árvoresde decisão.
- ***Extreme Gradient Boosting*** (XGBoost): consiste numa variação da técnica de regressão e classificação conotada de *Gradient Boosting*, que produz um modelo de previsão, através do agregado de vários modelos preditivos mais fracos, tipicamente originários de árvores de decisão (DT). Contudo, o que faz o *XGBoost*, diferente do *Gradient Boosting*, é que as árvores são construídas em paralelo em vez de uma forma sequencial como acontece no *Gradient Boosting*, controlando melhor o *over-fitting* das amostras, originando uma melhor performance.
- ***Extra Tree*** (ET): consiste numa abreviação do termo *Extremely Randomized Trees*, que de forma similar ao RF, cria várias árvores de decisão, mas a criação de amostras ou *sampling* para cada árvore ocorre de forma aleatória, e usa todo o conjunto treino para treinar cada árvore de regressão.
- ***Stochastic Gradient Descent*** (SGD): consiste num algoritmo de otimização, que tem em consideração o peso de cada ponto de informação, calculando-o individualmente. Em vez de analisar todos os pontos de informação de uma só

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

vez, calcula individualmente e verifica o gradiente, atualizando os pesos com essa análise individual. O processo continua até atingir o melhor valor otimizado.

- ***Hidden Markov Model*** (HMM): consiste num modelo estatístico utilizado para descrever a relação probabilística entre uma sequência de observações e a sequência de variáveis que gerem os dados observados, contudo eles não diretamente observáveis ou determináveis. É comummente utilizado nas situações em que o sistema ou processo que gera estas observações é desconhecido ou "escondido".
- ***Maximum Entropy Markov*** (MEM): consiste na utilização do algoritmo HMM para prever um determinado output, observando uma determinada sequência, mas incorporando a multinominal RL (mais conhecida como *Maximum Entropy*), que possibilita a liberdade de extrair o tipo e variáveis que queremos da sequência observada.
- ***Conditional Random Fields*** (CRF): consiste num modelo gráfico que é largamente utilizado em várias áreas, seja na saúde, reconhecimento de imagem, processamento de linguagem, etc. O seu principal objetivo é prever sequências, utilizando para o efeito o exemplo do seu dado vizinho, sendo este essencial para calcular uma determina sequência.
- ***Recurrent Neural-Work*** (RNN): consiste num tipo de algoritmo que usa dados sequencias ou series temporais de dados. É comumente utilizado para problemas ordinais ou temporais, seja por exemplo, reconhecimento de voz, captação de imagem, etc. Este algoritmo é distinguido pela sua "memória", já que adquire informação do seus inputs passados para influenciar o input atual e correspondente output. Este algoritmo depende sempre dos inputs passados, ao contrário dos algoritmos de *Deep Learning* em que os inputs e outputs são independentes um do outro.
- ***Long Short-Term Memory*** (LSTM): consistem num tipo de algoritmo de Deep Learning, consistindo numa variação de RNN, que é capaz de processar uma quantidade inteira de dados, e não apenas pequenos dados como por exemplo imagens. A vantagem que possui face ao RNN, é que consegue selecionar dentro

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

dessa quantidade inteira de dados, qual a informação que é considerada essencial, daquela que pode ser descartada.

- ***Gated Recurrent Unit*** (GRU): consiste num algoritmo de *Deep Learning*, sendo também uma variação do algoritmo RNN. Apesar de muito semelhante ao LSTM e ao RNN, a vantagem que possui é que usa menor memória de processamento, sendo neste sentido mais rápido que o LSTM, contudo é menos preciso quando é utilizado *datasets* com grandes sequências.
- ***Artificial Neural Network*** (ANN): consiste num modelo computacional que tem como objetivo replicar a forma como as células neurais funcionam. Consiste em várias camadas de nódulos, que contém uma camada de input, uma ou mais camadas "escondidas", e uma camada de output. Cada nódulo, liga ao outro, e tem um peso e limite associado. Se o output de cada nódulo individual está acima do limite associado, o nódulo é ativado, enviando informação para a próxima camada da rede neural. Ao contrário, nenhuma informação é capturada pela próxima camada.
- ***LightGBM:*** consiste num algoritmo de árvores de decisão com intuito de aumentar a eficácia do modelo e reduzir o tempo de memória utilizado. Este algoritmo divide em forma de histograma o conjunto de dados, numa forma a otimizar o processamento dos mesmos. Por exemplo no caso de termos milhões de dados, este algoritmo reduz essa quantidade, dividindo esse mesmo conjunto e reduzindo o número de linhas a serem analisadas.
- ***CatBoost:*** consiste num algoritmo que trata tanto problemas de regressão como de classificação. É um algoritmo que possui uma grande quantidade de parâmetros que otimizam as variáveis na fase de processamento das mesmas, não sendo necessário uma otimização posterior. É um algoritmo bastante utilizado quando analisamos dados categóricos.
- ***Long Short-Term Memory - CRF*** **(**LSTM-CRF**)**: Este modelo consiste numa única camada do algoritmo LSTM e uma única camada do algoritmo CRF. Na medida em que a camada LTSM recebe uma determinada sequência de transações como input, e providência uma sequência de outputs após processar as mesmas.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Este LSTM output, consiste numa sequência que demonstra a probabilidade de cada transação ser fraudulenta ou não, que depois é "alimentada" à camada CRF como input. O objetivo da camada CRF é selecionar a sequência fraudulenta ou não transacional, tendo em conta a sua capacidade de predição. Ou seja, a camada LSTM encontra probabilidade de cada transação ser fraudulenta ou não, tendo em conta um determinado comportamento e transações passada, já que, a camada CRF toma a decisão final, tendo em conta a probabilidade e as dependências entre sucessivas predições.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

APÊNDICE 2. Métricas de ML identificadas no texto:

- ***Accuracy*:** Consiste no ratio entre o número de predições corretas e o número total da amostra.
- **AUC-PR:** consiste numa métrica utilizada para avaliar a performance de modelos de classificação binários, especialmente quando as classes não estão equilibradas. Ao contrário da curva ROC e AUC, que testa o *True Positive Rate* contra o *False Positive Rate*, a curva PR testa a *Precision* contra o *Recall*.
- **AUC-ROC:** consiste numa métrica de avaliação da performance do classificador binário. A curva ROC consiste num teste entre o *True Positive Rate* contra o *False Positive Rate*, em que quanto mais próxima esta curva estiver do valor 1 melhor performance o classificador terá, contrariamente quanto mais próximo se encontrar do 0, pior performance terá. Neste sentido, a AUC, ou *Area Under the Curve*, representa a performance geral do modelo, em que demonstra a probabilidade de uma amostra aleatória positiva ser caracterizada como positiva, do que uma amostra aleatória negativa de ser caracterizada como negativa.
- **F1-Score:** Corresponde a uma média entre a *Precision* e o *Recall*, e o seu resultado varia entre 0 e 1. Indica o quão preciso que o classificador é, isto é, quantos dados/instâncias classificou corretamente. Esta métrica tenta encontrar um balanço entre a *Precision* e o *Recall*.
- ***Mathhews Correlation Coefficient*** (MCC): a técnica MCC é utilizada para medir a qualidade da classificação, e que pode variar entre – 1 e + 1. O quanto mais próximo o MCC estiver do +1, maior será a qualidade da classificação.
- **Precision:** Consiste no número de resultados positivos corretos dividido pelo número total de resultados positivos pelo classificador.
- **Recall:** Consiste no número de resultados positivos corretos dividido pelo número total de todas as amostras relevantes (ou seja, todas as amostras que deveriam ter sido identificadas como positivas, por exemplo).

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

- **Sensivity/True Positive Rate:** Corresponde à proporção de dados positivos que foram corretamente identificados como positivos, tendo em conta todos os resultados positivos.
- **Specificity/False Positive Rate:** Corresponde à proporção de dados negativos que foram corretamente identificados como negativos, tendo em conta todos os resultados negativos.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

APÊNDICE 3. Técnicas de equilíbrio de *datasets* identificadas no texto:

- **Adasyn (Adaptive Synthetic):** consiste numa técnica de equilíbrio de dados, na medida em que gere dados/observações sintéticas de acordo com o nível de dificuldade em aprender as observações da classe minoritária, isto é, quanto mais difícil for de aprender as observações minoritárias, mais observações sintéticas são geradas**.**
- **Aleatório com um Ratio:** consiste numa técnica de equilíbrio de dados, em que por exemplo, por cada 5 dados da classe maioritária, é atribuído um dado da classe minoritária, e assim consequentemente.
- **Bayesian Optimization:** Consiste numa técnica utilizada na determinação de qual ou quais os melhores parâmetros para atingir uma determinada performance de um algoritmo. Tem em consideração parâmetros já utilizados e através dessa memória, descarta e tem em consideração outros, até atingir aqueles que melhor performance atribuem ao algoritmo. Podemos aplicar esta técnica no equilíbrio dos dados, especialmente quando o *dataset* não é equilibrado, esta técnica irá determinar quais os parâmetros/variáveis que podem mitigar e ultrapassar o problema do desequilíbrio.
- **Oversampling:** consiste numa técnica de equilíbrio de dados, em que mantemos a classe maioritária, mas aumentamos a dimensão da classe minoritária.
- **Seq-US (Sequence-Aware Undersampling):** Este modelo mantém o padrão sequencial antes de uma determinada transação ser reconhecida como fraude, pois essa sequência possui informação essencial que permitiu descodificar determinada transação como fraude (*critical instances*). Após a retenção desta informação, retira do *undersampling* as sequências que não foram essenciais na captura desta transação fraudulenta (*safe instances*).
- **SMOTE (Synthetic Minority Oversampling Technique):** consiste numa técnica criada especificamente para nos casos de datasets não equilibrados, ou

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

classes não equilibradas, serão gerados amostras sintéticas para a classe minoritária, com o objetivo de equilibrar ambas as classes.

- **Undersampling:** consiste numa técnica de equilíbrio de dados, em que mantemos a classe minoritária, mas reduzindo a dimensão da classe maioritária.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt