

APRENDIZADO DE MÁQUINA PARA CLASSIFICAÇÃO DO RATING DE EMPRESAS DO MERCADO

Anderson Mitterhofer Iung¹, Leonardo Domingues¹, and Renato de Viveiros Lima¹

¹Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)

Dezembro, 2020

RESUMO

Em meados do século XIX iniciou-se, nos Estados Unidos, a atividade de qualificação de crédito de empresas, prática que se consolidou e, hoje, dominada pelas conhecidas *The Big Three*, as três maiores agências de classificação de risco de crédito do mundo: Moody's, Fitch e S&P. As decisões relacionadas às operações de crédito, investimentos, relações comerciais etc. estão sujeitas à análise e detecção dos riscos de crédito dos agentes envolvidos. Sendo assim, uma classificação do risco de crédito, levando-se em conta a solvabilidade das empresas, é muito importante. Os critérios de classificação, adotados pelas principais agências, apresentam padrões semelhantes e são largamente utilizados. Em geral, as análises das agências de riscos cobrem grandes empresas e são disponibilizadas através de pagamento, limitando o acesso. O artigo aborda a utilização de técnicas de aprendizado de máquina para emular a classificação de rating de crédito das grandes agências de risco, com uma base de dados de empresas do mercado americano. A literatura apresenta estudos a respeito da classificação dos títulos privados (*bonds*) emitidos pelas empresas, as quais se mostraram ferramentas promissoras, pois pode se extrair conhecimento implícito dos algoritmos de aprendizado e complementar com a informação dos especialistas. Desta forma, o artigo propõe utilizar de técnicas bem conhecidas de aprendizado de máquina como Redes Neurais, SVM e *Random Forest* para classificar o rating das empresas. Os resultados obtidos demonstraram bastante potencial e, que melhores resultados podem ser obtidos através de um conjunto de dados mais significativo ou aplicações de outras técnicas, como por exemplo, classificação hierárquica.

Palavras-chaves: Rating, Empresas, Redes neurais, SVM, Random forest, Aprendizado de máquina

1 Introdução

A conceituação de risco é caracterizada de diversas maneiras. Pode ser tratado como a relação risco e retorno de um investimento qualquer, risco de algum sinistro para companhias seguradoras, ou, também, a incerteza relacionada a qualquer evento. Os riscos enfrentados pelas instituições financeiras se encaixam amplamente nas seguintes amplas categorias: o risco de mercado, risco de crédito, risco de liquidez, risco operacional e o risco sistêmico, conforme observado por Duffie e Singleton (2012).

Jarrow (2009) destaca que o risco de crédito e o risco de liquidez costumam andar juntos e que modelagem conjunta destes riscos é uma área importante e relativamente inexplorada que merece muito mais atenção.

A extinção da incerteza em um tempo à frente não pode ser realizada através de uma gestão de risco, mas sim a mitigação dos acontecimentos futuros. De forma geral, a gestão de risco monitora e controla a probabilidade de perda, mas não a elimina.

Crédito, no setor bancário, é comumente dito como a permissão que as pessoas físicas ou jurídicas acessem recursos específicos com um acordo que impõe condições predefinidas tanto para o credor quanto para o tomador (MARQUES; GARCÍA; SÁNCHEZ, 2013). Normalmente, o risco de crédito é relacionado com base na competência de reembolso, capital, condições de empréstimos, histórico de crédito e garantias associadas.

A respeito do estudo de crédito, em sua essência, as informações de risco são comumente agrupadas em algumas categorias amplas. Alguns escritores, como mencionado em Durand (1941), falavam dos três C's do crédito - caráter, capacidade e capital; outros se referem ao risco moral, risco financeiro e risco cíclico. Hodgman (1960) diz que a atratividade de um empréstimo ou investimento para um credor está amarrado à múltiplos fatores,

tais como: a taxa de juros, risco, possíveis benefícios de um relacionamento de longo prazo com o cliente e assim por diante.

Nos primórdios da classificação do risco de crédito, uma época em que não haviam modelos robustos, as primeiras ações vieram das agências de classificação, como a John Moody & Company e a Fitch Publishing Company, no início do século XX, onde elas apresentaram seus manuais e livros abordando questões estatísticas e informações sobre ações e títulos. Em 1924, a Fitch divulgou seu sistema de classificação a que se tornou a base para classificações em todo o setor e é conhecida até hoje.

A partir da metade do século XX houve uma globalização do sistema bancário, acarretando numa competição entre as instituições financeiras, sendo, então, um motivo para estudar os problemas deste sistema, além do desenvolvimento de uma metodologia de gerenciamento de risco de crédito para os bancos comerciais, pois as dificuldades na modelagem do risco de crédito surgem devido ao fato da inadimplência da empresa não ser um fenômeno frequente, mas ocorrer principalmente de forma inesperada, conforme observado por Klieštík e Cúg (2015).

O sucesso de uma operação de crédito individual depende da capacidade de uma equipe de detectar e rejeitar riscos de crédito ruins. Consequentemente, para conseguir separar os bons e maus pagadores, surge a necessidade da classificação de crédito. Este é um processo de constituição de um sistema de classificação para credores de crédito para diferenciar tomadores de crédito atuais ou potenciais. Os primeiros estudos foram realizados na década de 60, por Hodgman (1960) e Smith (1964). Não se deve levar apenas em consideração a capacidade de pagamento da dívida, mas também a duração das relações individuais com os clientes, como ilustrado por Efimenko et al. (1970).

De acordo com Crouhy, Galai e Mark

(2001), um método de classificação de crédito tem três grupos de risco diferentes: o grau de investimento, que considera os clientes mais seguros no que diz respeito à força de reembolso; o segundo se refere ao grau especulativo e; o grau de rejeição.

Uma revisão sistemática dos modelos utilizados para classificação de crédito foi levantada e será descrita na Seção 2.

2 Revisão da Literatura

A literatura oferece diversos modelos quantitativos a respeito de risco de crédito, desde os modelos de pontuação de crédito do Beaver (1968) e Altman (1968), os estruturais de primeira geração de Merton (1974) e seus derivados (GESKE, 1977; BLACK; COX, 1976; VASICEK, 1984; KIM; RAMASWAMY; SUNDARESAN, 1993; LONGSTAFF; SCHWARTZ, 1995). Jarrow, Lando e Turnbull (1997) divulgaram o primeiro modelo de crédito de forma reduzida. No final da década de 1990, alguns bancos desenvolveram dois tipos de modelos de value at risk de crédito (VaR): modelos de modo de não pagamento e modelos de marcação a mercado. Os modelos de value at risk de crédito mais importantes são Credit Metrics, Credit Portfolio Manager da KMV e Credit Risk +, conforme destacado por Klieštík e Cúg (2015).

Altman, Haldeman e Narayanan (1977) destacam os cinco incrementos das últimas duas décadas do século XX que coagiram os bancos a ajustar seu risco de crédito por meio de instrumentos de classificação de crédito mais eficazes: aumento do número de falências, tendência à desintermediação, mais competitividade nas margens sobre os créditos, a redução do valor dos ativos reais e o crescimento dos instrumentos extrapatrimoniais com exposição ao risco de inadimplência inerente, que incluem derivativos de risco de crédito.

dito.

Por meio dos modelos, as instituições financeiras classificam os tomadores de empréstimo, avaliando seu desempenho financeiro e / ou não financeiro. Um pretendente que caracterizado fortemente como um candidato a reembolsar, é classificado de forma positiva para ser um tomador de crédito. Caso contrário, é visto como indigno de crédito. As classificações do tomador podem ser feitas de diversas maneiras, com diferentes finalidades, como destacado por Gül, Kabak e Topcu (2018).

O crescente interesse pelo risco de crédito pode ser imposto a diferentes fatores que surgiram desde queda do sistema financeiro dos Estados Unidos na década de 1980, a crise asiática no final dos anos 90, e se intensificou na crise do subprime nos Estados Unidos e, recentemente, o desastre de empréstimos da crise de crédito europeia. A última grande recessão de 2007-2008 teve como resultado a exposição das principais agências de classificação, as quais sofreram rígidas críticas por conta de suas estimativas erradas (MCLEAN; NOCERA, 2011). Os modelos usados na indústria antes de 2007 eram muito simples, onde havia a predominância de modelos estruturais modificados, sendo que estes são de natureza estática e não capturam a estrutura dinâmica do risco de crédito. Portanto, não é surpreendente que os profissionais e acadêmicos deem atenção cada vez maior à precificação de risco de crédito.

As pressões oriundas do mercado de capitais para as companhias atingirem maiores lucros, juntamente com a facilidade que a tecnologia nos trouxe para acessar e modificar dados dentro das organizações, acarretaram em fraudes nas demonstrações contábeis, conforme observado em Albrecht et al. (2015). Como consequência, a detecção de fraudes tem se tornado uma ferramenta importante para minimizar estas ocorrências.

Desta forma, as pesquisas de diversas

áreas acadêmicas têm buscado construir modelos de classificação de crédito mais eficientes, dinâmicos, utilizando técnicas como inteligência artificial ou algoritmos genéticos, conforme visto em Abdou e Pointon (2011), Addo, Guegan e Hassani (2018), Assef e Steiner (2020), Bhatore, Mohan e Reddy (2020), Gül, Kabak e Topcu (2018), Jagtiani e Lemieux (2019), Khandani, Kim e Lo (2010) Li e Zhong (2012) e; Sadatrasoul et al. (2013). Muitas técnicas diferentes, como tomada de decisão por critérios múltiplos, análise envoltória de dados e lógica fuzzy ainda estão sendo investigadas em pesquisas acadêmicas, de acordo com Gül, Kabak e Topcu (2018). Segundo Bhatore, Mohan e Reddy (2020), estas ferramentas, frequentemente são utilizadas para (i) Pontuação de crédito (ii) Previsão ativos sem desempenho (NPA) (iii) Detecção de fraude (transacional e não transacional) e outros casos de uso. Khandani, Kim e Lo (2010) acreditam que as previsões de aprendizado de máquina são mais adaptáveis e capazes de mensurar a dinâmica dos ciclos de crédito, bem como os níveis absolutos das taxas de inadimplência.

Vives (2017) notou que nos últimos anos as Fintechs estão inovando em alguns serviços financeiros, como a avaliação da capacidade de crédito dos solicitantes de empréstimos, aperfeiçoando a interface entre os clientes e seus prestadores de serviços. Além disso, Jagtiani e Lemieux (2019) observaram que as Fintechs medem a qualidade de crédito em poucos segundos, através de suas plataformas, aplicando várias técnicas de aprendizado de máquina que consomem dados estruturados e não estruturados.

Conforme destacado em Assef e Steiner (2020) e Sadatrasoul et al. (2013), os principais métodos de mineração de dados utilizados para a classificação de risco de crédito são os associados a regressão logística, regressão probit, análise de vizinho mais próximo (KNN), redes Bayesianas, Redes Neurais Ar-

tificiais (RNA), árvores de decisão, algoritmos genéticos, ensemble, máquinas de vetores de suporte (SVM), entre muitos outros. Assef e Steiner (2020) ainda afirmam que as RNAs é o método de inteligência artificial mais utilizado na última década e que para trabalhos futuros, poderiam ser feitos estudos de modelos híbridos. Esta análise também foi vista em Abdou e Pointon (2011). Por outro lado, Sadatrasoul et al. (2013), Bhatore, Mohan e Reddy (2020), Li e Zhong (2012) e Abdou e Pointon (2011) indicam que o método ensemble, SVM e RNAs são as técnicas mais usadas recentemente. Além disso, Sadatrasoul et al. (2013) citam várias possíveis investigações futuras e dentre elas a checagem de métodos híbridos contendo “Classificação + Clustering” que podem identificar e extrair grupos de candidatos bons e ruins em potencial.

Addo, Guegan e Hassani (2018) se concentraram na pontuação de risco de crédito, examinando o impacto do uso de diferentes modelos de aprendizado para identificar inadimplências das empresas, utilizando modelos como random forest, gradient boosting machine e deep learning.

Abdou e Pointon (2011) observaram que a classificação da importância das variáveis utilizadas nos modelos de pontuação é negligenciada na maioria dos trabalhos de pesquisa publicados sobre pontuação de crédito, sendo necessário realizar investigações nesse sentido. Além disso, a modelagem através de uma perspectiva comportamental dos inadimplentes pode ser estudada, para identificar os fatores da inadimplência.

Apesar de todos os estudos ilustrados ao longo do texto, podemos observar no trabalho de Naili e Lahrichi (2020) que o extenso trabalho realizado nos últimos anos relacionados ao risco de crédito permanece não solucionado e que a inadimplência ainda não é totalmente compreendida. Nesse sentido, de forma resumida, Naili e Lahrichi (2020) indicaram os seguintes pontos para estudos futuros: apro-

fundar os estudos em relação as variáveis utilizadas para a estrutura dos bancos; analisar o impacto da governança corporativa nos níveis de NPA dos bancos; estudar o impacto das práticas regulatórias e de supervisão sobre os níveis de NPA dos bancos; dentre outros.

Desta forma, a proposta do trabalho é aplicar os estágios da metodologia KDD (Knowledge Discovery in Databases), conforme apresentado em Fayyad, Piatetsky-Shapiro e Smyth (1996), para classificar empresas conforme seu risco de crédito, a partir dos dados coletados das demonstrações financeiras padronizadas (DFP), que é um documento eletrônico encaminhado à SEC (U.S. Securities & Exchange Commission) periodicamente, conforme descrito a seguir:

1. Data/Selection: para o desenvolvimento do trabalho será necessário ter acesso às demonstrações financeiras das empresas que serão classificadas de acordo com o seu risco de crédito;

2. Preprocessing: nesta etapa do processo serão desenvolvidos algoritmos para verificar a qualidade dos dados disponíveis, se existem dados faltantes etc.;

3. Transformation: após o acesso aos dados das demonstrações financeiras das empresas e o pré-processamento para obter uma base de dados confiável, serão aplicadas técnicas para obtenção dos índices financeiros e cálculo do rating de crédito de cada empresa, com base em uma série de parâmetros que serão desenvolvidos;

4. Data Mining: com base nos ratings de crédito calculados na etapa Transformation, serão desenvolvidas técnicas para classificação das empresas, conforme rating de crédito atribuído. Técnicas de visualização de dados poderão ser aplicados nesta fase;

5. Interpretation/Evaluation: após a aplicação de todas as etapas anteriores, poderão ser desenvolvidos relatórios com base nas informações e resultados. Como, por exemplo, evolução do risco de crédito de determinada

indústria, de uma empresa específica, como uma crise econômica afeta o risco de crédito das empresas, etc.

As classificações de crédito da empresa são normalmente muito caras de se obter, uma vez que exigem que agências como a Standard & Poor's e Moody's invistam grande quantidade de tempo e recursos humanos para realizar uma análise profunda do status de risco da empresa com base em vários aspectos que vão desde a competitividade estratégica até o nível operacional detalhes. Como resultado, nem todas as empresas podem pagar as classificações de crédito atualizadas anualmente dessas agências, o que torna a previsão da classificação de crédito bastante valiosa para a comunidade de investidores.

Por fim, o objetivo de classificar empresas conforme seu risco de crédito utilizando a metodologia KDD aplicada no conjunto de dados financeiros publicados pelas empresas. Neste trabalho, a base de dados dos demonstrativos financeiros são de empresas americanas, que possuem a obrigação de publicá-los na Comissão de Valores Mobiliários dos Estados Unidos (*SEC - Securities and Exchange Commission*).

3 Metodologia

Nesta seção serão abordadas as técnicas utilizadas para pré-processamento e mineração dos dados, bem como a caracterização do dataset utilizado.

3.1 Caracterização do dataset

O dataset utilizado neste trabalho foi gerado a partir de um arquivo contendo os dados das demonstrações financeiras de empresas do mercado norte americano, compreendendo o período de 2010-2018, que foram captados no Simfin (2020) e, de um arquivo contendo os ratings de crédito de empresas,

atribuídos pelas principais agências de riscos (DATA.WORLD, 2020). O dataset final, compostos pela interseção das duas bases, contém 3.122 registros e 73 atributos, que contém os dados originais e os indicadores financeiros mais utilizados na análise fundamentalista das companhias (PIZZOLATO, 2017).

3.2 Pré-processamento

O pré-processamento resume-se em: coleta dos dados, análise dos dados coletados, verificação da qualidade, eliminação de observações com dados faltantes, combinação dos resultados contábeis das empresas com o seus respectivos ratings de crédito, normalização dos dados, definição das classes e partição dos dados, em conjuntos de treinamento e teste.

A base original de demonstrações contábeis, coletada em Simfin (2020), contou com cerca de 1 milhão de registros, tendo como chave de identificação a combinação dos campos ticker da empresa e ano. No entanto, este arquivo contém dados de outras entidades que não são empresas de capital aberto e, portanto, fora do escopo deste trabalho. Sendo assim, os registros com estas características foram eliminados, reduzindo a base para, aproximadamente 16.000 registros e 45 atributos. Todos os atributos deste arquivo são oriundos do Balanço Patrimonial (BP), Demonstrativo de Resultados do Exercício (DRE) e Demonstrativo de Fluxo de Caixa (DFC).

O arquivo de ratings, coletado em Data.World (2020) possui cerca de 8 milhões de registros, tendo como chave de identificação o nome da empresa, ano, agência e o rating atribuído. Mais uma vez, foi necessário trabalhar a base para aproveitar apenas os registros do par empresa-ano que poderiam ser encontrados no arquivo de demonstrações financeiras, chegando a um total de 56.155 re-

gistros. O cruzamento dos dois arquivos, necessário para gerar o dataset desejado, exigiu um tratamento para associar o nome da empresa ao seu ticker. Finalmente, foi produzido um dataset de 3.122 registros e 45 atributos originais.

A esta base foram acrescentados 28 novos atributos, chamados indicadores, calculados a partir dos atributos originais. Esses indicadores são definidos na literatura de Finanças (PIZZOLATO, 2017) e se dividem em Liquidez (capacidade de pagamento das obrigações de curto prazo), Endividamento (proporção do capital de terceiros no financiamento da empresa), Eficiência (rotações de estoque, contas a receber, contas a pagar, ativo imobilizado e capital de giro) e Rentabilidade (margens e retornos sobre investimentos).

Desta forma, todos os dados usados na predição do rating são numéricos e objetivos, baseados em documentos públicos e acessíveis a qualquer investidor. Não foram considerados dados macro-econômicos, categóricos ou qualitativos, como índice de governança corporativa, por exemplo.

As empresas, que compõem o dataset, possuem ratings atribuídos em uma ampla faixa de D até AAA, faixa esta, amplamente utilizada pelas principais agências de risco, como a Standard & Poor's, Moody's Corporation, Fitch Ratings e Egan-Jones Ratings Company (EJR), totalizando 22 tipos de classificadores.

Os registros do dataset final não contém representates em todas as classes e, utilizar este amplo espectro, poderia prejudicar muito o processo de classificação. Por fim, classificou-se todo o conjunto de dados em cinco classes: Extremamente Fraca (EW), Fraca (W), Regular (R), Forte (S) e Extremamente Forte (ES), conforme apresentado na Tabela 1. A distribuição das classes é indicada na Tabela 2.

Descrição	Classe		S&P	Moody's	Fitch	EJR
Condição financeira extremamente forte	Extremamente Forte	5	AAA	Aaa	AAA	AAA
A capacidade de cumprir os compromissos financeiros é forte			AA+	Aa1	AA+	AA+
			AA	Aa2	AA	AA
			AA-	Aa3	AA-	AA-
Sólida condição financeira, mas suscetível a mudanças adversas nos negócios / economia	Forte	4	A+	A1	A+	A+
Exibe proteção adequada, mas é suscetível a mudanças adversas nos negócios / economia			A	A2	A	A
			A-	A3	A-	A-
			BBB+	Baa1	BBB+	BBB+
Tem capacidade para cumprir obrigações financeiras, enfrenta grandes incertezas em curso	Regular	3	BBB	Baa2	BBB+	BBB
			BBB-	Baa3	BBB-	BBB-
			BB+	Ba1	BB+	BB+
BB			Ba2	BB	BB	
Tem capacidade para cumprir as obrigações financeiras que as condições adversas de negócios, econômicas e financeiras irão prejudicar	Frac	2	BB-	Ba3	BB-	BB-
			B+	B1	B+	B+
			B	B2	B	B
CCC+			Caa1	CCC+	CCC+	
Vulnerável ao não pagamento; condições de negócios favoráveis necessárias para cumprir as obrigações	Extremamente Fraca	1	CCC	Caa2	CCC	CCC
			CCC-	Caa3	CCC-	CCC-
			CC	Ca	CC	CC
C			C	C	C	
Solicitada proteção contra falência			D	D	D	D
Em default						

Tabela 1: Definição das classes utilizadas.

	Labels	Freq.
1	Extremamente Frac	5
2	Frac	33
3	Regular	961
4	Forte	1.854
5	Extremamente Forte	269
Total		3.122

Tabela 2: Dados por classe.

3.3 Redes Neurais

Pode-se dizer que uma rede neural é um algoritmo que tenta imitar a obscura estrutura de funcionamento do cérebro humano, refletindo seu comportamento em uma máquina que aprende situações de causa-efeito. Uma característica marcante das redes neurais é a sua total independência do conhecimento

da natureza fenomenológica do processo que se vai abordar, de modo que tal abordagem torna-se interessante quando as regras de um determinado processo não são determinísticas ou, são muito complexas, como, por exemplo, problemas onde ocorrem processos altamente não-lineares. Tais problemas, podem ser resolvidos com o uso de redes neurais artificiais (RNAs), pois os neurônios de uma RNAs apresentam resposta não-linear e são altamente conectados, podendo a rede mapear relações multivariáveis não-lineares com eficiência.

Redes neurais artificiais são compostas de vários elementos computacionais simples (nodos ou neurônio) que interagem localmente (altamente conectada). A arquitetura destes modelos é especificada pelas características do neurônio, topologia da rede e algoritmo

de treinamento.

Os neurônios em RNAs são processadores bastante simples inspirados por seus similares biológicos (neurônios cerebrais). Verifica-se que o neurônio artificial realiza seus cálculos baseados em suas informações de entrada. Ele faz o somatório do produto entre os vetores de entrada A e os pesos W_j , subtrai a ativação residual interna e então passa este resultado para uma forma funcional, $f(\cdot)$, ou seja:

$$f(W_j A - Th_j) = f\left(\sum_{n=1}^n (W_{i,j} a_i - Th_j)\right) \quad (1)$$

Esta forma funcional poderia ser qualquer função como logarítmica, exponencial, raiz quadrada, sigmoidais, etc.

O aprendizado (ou treinamento) de uma rede neural é o processo pelo qual os parâmetros livres de uma rede neural são ajustados por meio de uma forma continuada de estímulo pelo ambiente externo, sendo o tipo específico de aprendizado definido pela maneira particular como ocorrem os ajustes dos parâmetros livres (BRAGA, 2000).

Aprendizado supervisionado é utilizado para treinamento da rede neural quando esta é utilizada como ajustadora de funções multivariáveis e não-lineares, como o caso a ser tratado neste trabalho. Neste tipo de aprendizado, necessariamente, pressupõe-se a existência de um supervisor, ou um especialista, o qual é responsável por estimular as entradas da rede por meio de padrões de entrada e observar a saída calculada pela mesma, comparando-a com a saída desejada. Como a resposta da rede é função dos valores atuais do seu conjunto de pesos, estes são ajustados num processo de otimização baseada em mínimos quadrados de forma a aproximar a saída da rede da saída desejada.

Para realização deste processo que, na realidade, matematicamente é um processo de otimização, existem alguns algoritmos já esta-

belecidos que apresentam características peculiares nas suas utilizações.

Dentre os algoritmos citados na literatura (DOROFKI et al., 2012; PAOLA; SCHOWENGERDT, 1995; GEVREY; DIMOPOULOS; LEK, 2003; EL-MIHOUB et al., 2006), pode-se destacar, pela frequência de utilização, dois algoritmos baseados em métodos de otimização diferentes: o de retropropagação do erro (*backpropagation*), com a regra do delta generalizado, que é um método de gradiente descendente bastante utilizado, inclusive neste trabalho; e o método de Levenberg-Marquardt, baseado no método quasi-Newton. Por se tratarem de algoritmos que envolvem o cálculo de derivadas, supõem-se a utilização de funções contínuas, principalmente as sigmóides. Neste trabalho utilizou-se várias funções de transferências, mas a que apresentou resultados mais consistentes foi a *relu* do *Keras model*.

As redes neurais podem ser utilizadas em qualquer situação em que o objetivo é o de determinar uma variável desconhecida ou atributo a partir de observações conhecidas ou medições registradas (por exemplo, várias formas de regressão, classificação, e as séries de tempo), onde não há uma quantidade suficiente de dados históricos e, onde há existe uma relação tratável subjacente ou um conjunto de relações. A rede neural pode contribuir para outras técnicas no sentido da capacidade de aprendizagem a partir de exemplos (histórico de dados).

3.4 SVM

Diferente das Redes Neurais, o método *Supporting Vector Machine* (SVM) segue a abordagem "ansiosa" (eager), na qual o modelo é construído em tempo real, durante a etapa de treinamento. A técnica básica do SVM é encontrar a reta cuja distância para o elemento mais próximo de cada classe seja a maior possível, chamada maximização de

margem.

Quando a classificação envolve várias classes, usa-se uma função kernel que adapta a função linear a um espaço de n dimensões, conforme explicado em Yu e Sungchul (2012). Desta forma, este modelo tende a funcionar bem em datasets com alto número de dimensões, sendo independente da distribuição de classes e resultando em uma solução única, diferente das Redes Neurais, o que se aplica bem ao problema tratado neste trabalho.

3.5 Random Forest

O método Random Forest também segue a abordagem "ansiosa", sendo uma evolução do tradicional método de árvores de decisão por reduzir a variabilidade das árvores individuais. É um dos métodos mais usados atualmente por se tratar de um método de aprendizado e predição não paramétrica e devido à sua capacidade de prover resultados com maior acurácia com fácil utilização (Tuselman, Sinkovics e Pishchulov (2015)).

A base deste método é acrescentar uma camada de aleatoriedade a cada interação, substituindo a busca pelo melhor atributo a ser definido como nó, nas árvores de decisão simples, pela busca de um conjunto de melhores atributos para definir cada nó. Segundo Liaw e Wiener (2002) este método tende a apresentar melhor performance quando comparado com outros classificadores, além de ser robusto contra *overfitting*.

4 Resultados

A linguagem de programação aplicada foi o *R* para a utilização das técnicas citadas na seção 3. Seus resultados serão apresentados por partes, como segue:

4.1 Redes Neurais

O modelo utilizado para treinamento dos dados foi o *Keras*. Este é um API de aprendizado profundo escrito originalmente em Python, e recentemente adicionado na documentação *R*. *Keras* é executada na plataforma de aprendizado de máquina *TensorFlow*.

Redes neurais *backpropagation* têm sido extremamente populares por sua capacidade única de aprendizagem, conforme citado anteriormente, e mostraram ter um bom desempenho (HUANG et al., 2004; GOLBAYANI; FLORESCU; CHATTERJEE, 2020).

Uma rede neural *backpropagation* típica consiste em uma estrutura de três camadas: nós da camada de entrada, nós da camada de saída e nós da camada oculta. Em nosso estudo, usamos variáveis financeiras oriundas dos balanços contábeis como nós de entrada e o resultado da classificação como nós da camada de saída, conforme classificação indicada na Tabela 1.

A melhor acurácia obtida foi de 72,76%, utilizando 70% dos dados para treinamento, com apenas as variáveis oriundas do balanço contábil (excluindo os indicadores), com uma rede neural com estrutura de cinco nós, nas camadas escondidas, com 100 neurônios em cada camada oculta, *dropout* de 2,5% (ajuda a evitar *overfitting*), 100 épocas, distribuição normal para os pesos iniciais das camadas, função de ativação *relu*, otimizador *RMSprop*, cálculo da média dos quadrados dos erros entre rótulos e previsões, tamanho do lote com 25 amostras e divisão de validação de 25%. A matriz de confusão obtida é ilustrada na Tabela 3.

As métricas de avaliação utilizadas neste trabalho foram a precisão (P), cobertura (C) e *F-measure* (*F-M*). Estas são indicadas na Tabela 4.

	EW	W	R	S	ES
EW	1	4	4	0	0
W	0	0	0	0	0
R	0	1	138	49	4
S	0	6	145	513	37
ES	0	0	0	5	29

Tabela 3: Matriz de confusão.

	EW	W	R	S	ES
P	11%	0%	72%	73%	85%
C	100%	0%	48%	90%	41%
F-M	20%	-	58%	81%	56%

Tabela 4: Métricas de avaliação.

Em termos de acurácia, o modelo pode ser aceitável, mas em relação às outras métricas, o modelo desenvolvido não atinge as expectativas. Uma possibilidade verificada foi a diminuição das classes, na tentativa de balancear estas, mas as soluções encontradas foram piores. O modelo proposto tem potencial de classificar o rating das empresas a partir das demonstrações financeiras padronizadas (DFP), mas o conjunto de dados utilizados carece de amostras, principalmente nas classes Extremamente Fraca e Fraca.

4.2 SVM

Com o método SVM foram executados testes com três variações do dataset original: os dados originais (base Brutos), indicadores financeiros (base Indicadores) calculados a partir dos dados originais e calculados segundo a literatura de Finanças (Pizzolato (2017)) e os dados originais com redução de dimensionalidade, na qual foram excluídos os atributos redundantes, que são funções de outros atributos, colunas totalizadoras e de resultados (base Redução).

Para os três datasets foi executado o SVM variando as diferentes opções do kernel: linear, polinomial, radial e sigmoid. Não se verificou grande alteração nos resultados, tanto

para variações no dataset quanto nos parâmetros da função.

A melhor acurácia encontrada foi igual a 69% para o dataset Redução com kernel radial, enquanto a pior acurácia encontrada foi de 49%, para o dataset Redução com kernel sigmoid. Os demais testes apresentaram resultados intermediários. Cabe observar que para este método foi necessário realizar a normalização dos atributos nos datasets Brutos e Redução, caso contrário os resultados foram piores pela natural disparidade dos valores entre as diferentes empresas de diversos setores.

O dataset Indicadores apresentou uma acurácia maior (68%) que a base Brutos (61%), uma vez que os indicadores já são naturalmente normalizados. Esta ideia já tinha sido observada em Auria e Moro (2008) para o caso de análise de crédito de clientes de um determinado banco.

4.3 Random Forest

Neste trabalho o Random Forest foi o método que apresentou os melhores resultados de acurácia, acima de 80% nos três datasets testados (Brutos, Indicadores e Redução).

Nos três testes o número de árvores foi fixado em 100, enquanto o parâmetro *mtry* foi calculado como a raiz quadrada do número de atributos, conforme recomendado por Liaw e Wiener (2002). O melhor resultado foi acurácia igual a 87% para a base Brutos, superiores aos 84% encontrados para a base Redução e os 80% da base Indicadores.

Vale observar que variações feitas em cima dos parâmetros *nrtree* e *mtry* não apresentaram melhoras relevantes no resultado.

A tentativa de downsampling também não produziu resultado esperado, uma vez que a baixa quantidade de registros na classe EW fez com que o dataset de treino ficasse limitado a cerca de 0,01 dos registros da base original.

Da mesma forma, a normalização dos da-

dos das bases Brutos e Redução não melhorou o resultado obtido.

4.4 Outros

Breves testes foram realizados com o método Naive Bayes, porém, como não se obteve resultados melhores do que os já encontrados, de forma que os resultados foram descartados.

5 Conclusões e trabalhos futuros

As redes neurais apresentaram resultados razoáveis na acurácia do modelo, mas apresentou performance abaixo da técnica de Random Forest. O modelo apresentou potencial na classificação do rating das empresas a partir dos dados da DFP mas, como o conjunto apresenta poucos elementos nas classes Extremamente Fraca e Fraca, é possível que o modelo proposto apresente melhores resultados com um conjunto de dados maiores, com mais representates nestas classes, possibilitando maior aprendizado dos algoritmos.

A abordagem por SVM alcançou os piores resultados em termos de acuracidade.

A técnica de Radom Forest apresentou os melhores resultados, alcançando 87%. O que a torna, altamente promissora para continuação dos estudos nesta área.

Importante ressaltar, que dificilmente um modelo baseado em aprendizado alcançará uma acurácia próxima de 100%, visto que nas metodologias utilizadas pelas agências de rating aspectos qualitativos, projeções etc. que não estão disponíveis nos documentos contábeis, são utilizados na atribuição dos ratings.

Para trabalhos futuros vislumbra-se a aplicação da metodologia de classificação hierárquica, para melhorar a acurácia dos modelos propostos.

Referências

ABDOU, H. A.; POINTON, J. Credit scoring, statistical techniques and evaluation criteria: a review of the literature. **Intelligent systems in accounting, finance and management**, Wiley Online Library, v. 18, n. 2-3, p. 59–88, 2011.

ADDO, P. M.; GUEGAN, D.; HASSANI, B. Credit risk analysis using machine and deep learning models. **Risks**, Multidisciplinary Digital Publishing Institute, v. 6, n. 2, p. 38, 2018.

ALBRECHT, W. S.; ALBRECHT, C. O.; ALBRECHT, C. C.; ZIMBELMAN, M. F. **Fraud examination**. [S.l.]: Cengage Learning, 2015.

ALTMAN, E. I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. **The journal of finance**, JSTOR, v. 23, n. 4, p. 589–609, 1968.

ALTMAN, E. I.; HALDEMAN, R. G.; NARAYANAN, P. Zetatm analysis a new model to identify bankruptcy risk of corporations. **Journal of banking & finance**, Elsevier, v. 1, n. 1, p. 29–54, 1977.

ASSEF, F. M.; STEINER, M. T. A. Ten-year evolution on credit risk research: a systematic literature review approach and discussion. **Ingeniería e Investigación**, v. 40, n. 2, 2020.

AURIA, L.; MORO, R. A. Support vector machines (svm) as a technique for solvency analysis. **DIW Berlin Discussion Papers ISSN electronic edition 1619-4535**, n. 811, 2008. Disponível em: <<http://ssrn.com/abstract=1424949>>.

BEAVER, W. H. Alternative accounting measures as predictors of failure. **The accounting review**, JSTOR, v. 43, n. 1, p. 113–122, 1968.

- BHATORE, S.; MOHAN, L.; REDDY, Y. R. Machine learning techniques for credit risk evaluation: a systematic literature review. **Journal of Banking and Financial Technology**, Springer, p. 1–28, 2020.
- BLACK, F.; COX, J. C. Valuing corporate securities: Some effects of bond indenture provisions. **The Journal of Finance**, Wiley Online Library, v. 31, n. 2, p. 351–367, 1976.
- BRAGA, A. d. P. **Redes neurais artificiais: teoria e aplicações**. [S.l.]: Livros Técnicos e Científicos, 2000.
- CROUHY, M.; GALAI, D.; MARK, R. Prototype risk rating system. **Journal of banking & finance**, Elsevier, v. 25, n. 1, p. 47–95, 2001.
- DATA.WORLD. **Data World - Credit Ratings history data**. 2020. Disponível em: <<https://data.world/muni-finance/credit-ratings-history-data>>.
- DOROFKI, M.; ELSHAFIE, A. H.; JAAFAR, O.; KARIM, O. A.; MASTURA, S. Comparison of artificial neural network transfer functions abilities to simulate extreme runoff data. **International Proceedings of Chemical, Biological and Environmental Engineering**, v. 33, p. 39–44, 2012.
- DUFFIE, D.; SINGLETON, K. J. **Credit risk: pricing, measurement, and management**. [S.l.]: Princeton university press, 2012.
- DURAND, D. **Risk elements in consumer installment financing**. [S.l.]: National Bureau of Economic Research, New York, 1941.
- EFIMENKO, L. V.; SHINDINA, T. A.; TABAKOVA, E. V.; GAVRILOVA, A. V. Credit risk prediction to individuals. **The Journal of Internet Banking and Commerce**, Research and Reviews, 1970.
- EL-MIHOUB, T. A.; HOPGOOD, A. A.; NOLLE, L.; BATTERSBY, A. Hybrid genetic algorithms: A review. **Engineering Letters**, v. 13, n. 2, p. 124–137, 2006.
- FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37–37, 1996.
- GESKE, R. The valuation of corporate liabilities as compound options. **Journal of Financial and quantitative Analysis**, JSTOR, p. 541–552, 1977.
- GEVREY, M.; DIMOPOULOS, I.; LEK, S. Review and comparison of methods to study the contribution of variables in artificial neural network models. **Ecological modelling**, Elsevier, v. 160, n. 3, p. 249–264, 2003.
- GOLBAYANI, P.; FLORESCU, I.; CHATTERJEE, R. A comparative study of forecasting corporate credit ratings using neural networks, support vector machines, and decision trees. **The North American Journal of Economics and Finance**, Elsevier, v. 54, p. 101251, 2020.
- GÜL, S.; KABAK, Ö.; TOPCU, I. A multiple criteria credit rating approach utilizing social media data. **Data & Knowledge Engineering**, Elsevier, v. 116, p. 80–99, 2018.
- HODGMAN, D. R. Credit risk and credit rationing. **The Quarterly Journal of Economics**, MIT Press, v. 74, n. 2, p. 258–278, 1960.
- HUANG, Z.; CHEN, H.; HSU, C.-J.; CHEN, W.-H.; WU, S. Credit rating analysis with support vector machines and neural

- networks: a market comparative study. **Decision support systems**, Elsevier, v. 37, n. 4, p. 543–558, 2004.
- JAGTIANI, J.; LEMIEUX, C. The roles of alternative data and machine learning in fintech lending: evidence from the lendingclub consumer platform. **Financial Management**, Wiley Online Library, v. 48, n. 4, p. 1009–1029, 2019.
- JARROW, R. A. Credit risk models. **Annu. Rev. Financ. Econ.**, Annual Reviews, v. 1, n. 1, p. 37–68, 2009.
- JARROW, R. A.; LANDO, D.; TURNBULL, S. M. A markov model for the term structure of credit risk spreads. **The review of financial studies**, Oxford University Press, v. 10, n. 2, p. 481–523, 1997.
- KHANDANI, A. E.; KIM, A. J.; LO, A. W. Consumer credit-risk models via machine-learning algorithms. **Journal of Banking & Finance**, Elsevier, v. 34, n. 11, p. 2767–2787, 2010.
- KIM, I. J.; RAMASWAMY, K.; SUNDARESAN, S. Does default risk in coupons affect the valuation of corporate bonds?: A contingent claims model. **Financial management**, JSTOR, p. 117–131, 1993.
- KLIEŠTIK, T.; CÚG, J. Comparison of selected models of credit risk. **Procedia economics and finance**, Elsevier Science Publishing Company, Inc., v. 23, p. 356–361, 2015.
- LI, X.-L.; ZHONG, Y. An overview of personal credit scoring: techniques and future work. Scientific Research Publishing, 2012.
- LIAW, A.; WIENER, M. **Classification and Regression by randomForest**. 2002.
- LONGSTAFF, F. A.; SCHWARTZ, E. S. A simple approach to valuing risky fixed and floating rate debt. **The Journal of Finance**, Wiley Online Library, v. 50, n. 3, p. 789–819, 1995.
- MARQUES, A.; GARCÍA, V.; SÁNCHEZ, J. S. A literature review on the application of evolutionary computing to credit scoring. **Journal of the Operational Research Society**, Springer, v. 64, n. 9, p. 1384–1399, 2013.
- MCLEAN, B.; NOCERA, J. **All the devils are here: The hidden history of the financial crisis**. [S.l.]: Penguin, 2011.
- MERTON, R. C. On the pricing of corporate debt: The risk structure of interest rates. **The Journal of finance**, JSTOR, v. 29, n. 2, p. 449–470, 1974.
- NAILI, M.; LAHRICHI, Y. The determinants of banks' credit risk: Review of the literature and future research agenda. **International Journal of Finance & Economics**, Wiley Online Library, 2020.
- PAOLA, J. D.; SCHOWENGERDT, R. A. A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery. **International Journal of remote sensing**, Taylor & Francis, v. 16, n. 16, p. 3033–3058, 1995.
- PIZZOLATO, N. D. **Introdução à Contabilidade Gerencial**. [S.l.]: Publit, 2017.
- SADATRASOUL, S. M.; GHOLAMIAN, M.; SIAMI, M.; HAJIMOHAMMADI, Z. Credit scoring in banks and financial institutions via data mining techniques: A literature review. **Journal of AI and Data Mining**, Shahrood University of Technology, v. 1, n. 2, p. 119–129, 2013.

SIMFIN. **Simfin - Dados contábeis das empresas americanas**. 2020. Disponível em: <<https://simfin.com/>>.

SMITH, P. F. Measuring risk on consumer instalment credit. **Management Science**, INFORMS, v. 11, n. 2, p. 327–340, 1964.

TUSELMAN, H.; SINKOVICS, R. R.; PISHCHULOV, G. Towards a consolidation of worldwide journal rankings – a classification using random forests and aggregate rating via data envelopment analysis. **Omega**, n. 51, p. 11–23, 2015. Disponível em: <<http://dx.doi.org/10.1016/j.omega.2014.08.002>>.

VASICEK, O. A. **Credit valuation**. [S.l.]: KMV Corporation, March, 1984.

VIVES, X. The impact of fintech on banking. **European Economy**, Europeye srl, n. 2, p. 97–105, 2017.

YU, H.; SUNGCHUL, K. **SVM Tutorial: Classification, Regression and Ranking**. 2012. Disponível em: <https://www.researchgate.net/profile/Sungchul_Kim/publication/229010211/_SVM_Tutorial_Classification_Regression_and_Ranking/links/53e2a7060cf2b9d0d8329b99/SVM-Tutorial-Classification-Regression-and-Ranking.pdf>.