

APLICAÇÃO DO APRENDIZADO DE MÁQUINAS PARA CLASSIFICAÇÃO DE COOPERATIVAS DE CRÉDITO EM RISCO DE ENCERRAMENTO

ANA CAROLINA DE ALCANTARA OLIVEIRA
UNIVERSIDADE FEDERAL DE SÃO CARLOS (UFSCAR)

FLÁVIO LEONEL DE CARVALHO
UNIVERSIDADE FEDERAL DE SÃO CARLOS (UFSCAR)

APLICAÇÃO DO APRENDIZADO DE MÁQUINAS PARA CLASSIFICAÇÃO DE COOPERATIVAS DE CRÉDITO EM RISCO DE ENCERRAMENTO

1 INTRODUÇÃO

As cooperativas de crédito, entidades formadas por pessoas com interesses econômicos em comum, têm como objetivo atingir metas econômicas e sociais dos cooperados e contribuir para o desenvolvimento da localidade à qual estão inseridas (McKillop; Wilson, 2011). Desse modo, desempenham um papel fundamental no combate aos desequilíbrios socioeconômicos (Anakpo *et al.*, 2023).

Apesar de não terem como objetivo a obtenção de lucro, as cooperativas necessitam de mecanismos que garantam sua sustentabilidade financeira e operacional, como controle de liquidez, solvência e administração eficiente de seus ativos financeiros (Silva *et al.*, 2017). No entanto, mesmo com a implementação de mecanismos de gestão, a estrutura organizacional das cooperativas pode apresentar riscos à sua sobrevivência, especialmente devido a conflitos de interesse na concessão de créditos (Jansen; Maehler; Wegner, 2018; Paula *et al.*, 2019).

A busca pela maximização dos resultados individuais dos cooperados poderá ser viabilizada por meio da aplicação de taxas de juros mais atrativas aos cooperados, em detrimento da solidez financeira e sobrevivência das cooperativas de crédito (Carvalho *et al.*, 2015), além da busca pela maximização da distribuição de sobras. Nesse contexto, destacam-se o risco de crédito, de liquidez, operacional e de mercado, exigindo gerenciamento efetivo para assegurar os objetivos centrais e a qualidade contínua dos serviços ofertados aos cooperados (Santos *et al.*, 2020).

Esses problemas agravam a inadimplência e podem ameaçar a estabilidade financeira dessas organizações. Assim, a supervisão e monitoramento da saúde financeira das entidades financeiras nacionais são essenciais para a sustentabilidade do sistema financeiro (Karma; Made; Ni, 2022). Para isso, o modelo CAMEL (Capital Adequacy, Assets, Management Capability, Earnings, and Liquidity) é amplamente utilizado (Dadhich, *et al.*, 2022). No entanto, dado que essa metodologia utiliza um elevado número de indicadores financeiros e uma grande quantidade de informações, sua análise pode ser difícil de se realizar.

Desta forma, é essencial encontrar os métodos de monitoramento mais simples ou os indicadores financeiros mais relevantes para a previsão de dificuldades financeiras (Liang *et al.*, 2017). Uma abordagem promissora é o uso de aprendizado de máquina para analisar os indicadores CAMEL e determinar quais indicadores são mais importantes para identificar cooperativas com maior risco de descontinuidade.

Assim, seria possível tomar medidas preventivas para minimizar os riscos e garantir a sustentabilidade dessas organizações a longo prazo. A determinação dos indicadores mais pertinentes para categorizar as cooperativas com maior potencial de descontinuidade pode resultar em um monitoramento mais eficaz e auxiliar no processo de tomada de decisões durante o processo de gestão.

Desta forma, este estudo procura identificar os indicadores CAMEL mais relevantes para classificar as cooperativas entre aquelas com maior risco de encerramento. Espera-se que os resultados possam contribuir para a melhorar o processo de monitoramento de riscos dessas entidades, fortalecendo o sistema cooperativista. Além disso, os resultados podem ter impacto positivo na sustentabilidade e na confiança no setor financeiro, influenciando decisões internas das cooperativas e práticas de gestão de riscos mais eficazes.

2 REVISÃO BIBLIOGRÁFICA

De acordo com o Relatório de Economia Bancária (2020), em dezembro de 2020 havia cerca de 847 cooperativas de crédito singulares em operação no Brasil, sendo cerca de 222 independentes e outras 625 pertencentes a cinco cooperativas centrais (Ailos, Cecoop, Credisis, Uniprime e Cecrers).

No Brasil, as cooperativas de crédito são organizadas em três níveis principais; as cooperativas singulares, que realizam operações diretamente com seus associados, podendo estes serem pessoas físicas ou jurídicas; as cooperativas centrais, que prestam diversos serviços como administração de recursos de terceiros e consultorias técnicas para as cooperativas singulares; e as confederações de centrais ou sistemas cooperativos, que administram as filiadas e realizam operações que excedem a capacidade das cooperativas singulares e centrais, especialmente relacionadas a serviços de tecnologia da informação (Graça-Brasil *et al.*, 2020).

Outro nível de cooperativas são as confederações de centrais ou sistemas cooperativos, cuja atuação está voltada para a administração das filiadas e de operações que excedam a capacidade das cooperativas singulares e centrais, principalmente relacionados com a prestação de serviços de tecnologia da informação (Graça-Brasil *et al.*, 2020).

A Resolução 4.434, de 5 de agosto de 2015, estabelece que os serviços restritos aos associados das cooperativas de crédito incluem a captação de depósitos à vista e a prazo, compensação de cheques, operações de créditos, prestação de garantias e aplicação de recursos no mercado financeiro. No entanto, elas também oferecem serviços ao público geral, como cobrança, custódia, serviços de recebimentos e pagamentos, cartões de crédito, seguros e consórcios, distribuição de recursos do crédito rural e distribuição de cotas de fundos de investimento (BACEN, 2020).

As cooperativas de crédito se destacam em relação aos bancos comerciais pelo seu caráter participativo e social, onde os associados são simultaneamente clientes e sócios. Eles participam das decisões e da distribuição das sobras ao final do período (Sanchez; Moreira; Fontanini, 2020). Esse modelo de gestão demanda um acompanhamento constante do desempenho financeiro e social, não só para comparar com outras instituições dentro e fora do segmento cooperativista, mas também para garantir a viabilidade econômica dessas organizações, que têm objetivos sociais e econômicos a serem alcançados (Gonçalves; Bressan; Souza, 2023).

O desempenho financeiro obtido pelas cooperativas de crédito pode ser avaliado por meio de indicadores econômicos como PEARLS e CAMELS, que, através da análise dos dados dos balancetes, identificam a viabilidade econômica e as taxas de risco que impactam a continuidade dessas organizações (Silva; Santos; Ranciaro, 2023).

Além das atividades financeiras, as cooperativas de crédito buscam promover resultados sociais nas regiões onde estão inseridas (Araújo; Carmona, 2015). Para medir o desempenho social, autores como Araújo; Carmona (2015), Cavinato; Capitani (2023) e Fried *et al.*, (1993) sugerem o uso do método DEA (*Data Envelopment Analysis*), que agrega diferentes variáveis a um único indicador, apontando os níveis de eficiência social das instituições.

Uma das principais diferenças entre as cooperativas de crédito e os bancos tradicionais é a participação igualitária no controle societário. Segundo o Banco Central (2019), cada associado das cooperativas tem direito a um voto nas decisões do conselho, independentemente do valor investido no capital da cooperativa. Além disso, a distribuição de lucros ou prejuízos, chamados de sobras ou perdas nas cooperativas, é proporcional ao montante operado na instituição, e não há incidência de impostos como imposto de Renda das Pessoas Jurídicas (IRPJ) e contribuição Social sobre Lucro Líquido (CSLL) sobre essas distribuições (BACEN, 2019).

Apesar dos benefícios, esse modelo de gestão apresenta riscos, como a possibilidade de busca da maximização dos benefícios aos cooperados, além da capacidade econômico-financeira da cooperativa (Carvalho *et al.*, 2015). Esse potencial conflito de interesses, onde os clientes e proprietários são os mesmos indivíduos, pode levar a decisões que favorecem menores taxas de juros e maior retorno das sobras, o que pode impactar negativamente a sobrevivência das organizações (Morais; Martínez, 2018).

Os principais riscos enfrentados pelas cooperativas de crédito incluem risco de crédito, risco de liquidez, risco operacional e o risco de mercado, tornando necessário o gerenciamento efetivo das operações, para garantir a qualidade dos serviços ofertados aos cooperados (Santos *et al.*, 2020). Adicionalmente, há riscos relacionados ao porte da instituição, a baixa adesão nos primeiros anos de operação, e a fatores sociais e econômicas que podem influenciar diretamente a sobrevivência das cooperativas devido a influência de seus associados sobre a receita da organização (Canassa; Zancan; Costa, 2022).

Essa crise global impactou significativamente o sistema financeiro. O Relatório de Economia Bancária (2021) indica que durante a pandemia houve uma queda acentuada do indicador Lerner, refletindo uma redução na margem de poder de mercado, com queda de preços e o aumento dos custos marginais. No entanto, o cooperativismo apresentou crescimento superior ao restante do Sistema Financeiro Nacional, com maior adesão de associados e aumento na carteira de crédito, passando de 2,74% para 5,1% de todo SFN (BACEN, 2021).

O acompanhamento do desempenho financeiro das cooperativas de crédito, assim como de outras instituições, pode ser realizado por meio das demonstrações de resultados e indicadores que quantificam o desempenho dessas instituições (Silva, 2017). Segundo Silva (2017), o crescimento das instituições cooperativas é vital não apenas para sua sobrevivência, mas também pelo papel que desempenham nas comunidades onde atuam, oferecendo acesso ao crédito a municípios sem cobertura bancária, com custos baixos e incentivos para novos empreendimentos.

Diferentemente das instituições financeiras tradicionais, as cooperativas de crédito mantêm um relacionamento mais próximo com seus associados, buscando democratizar a disponibilidade de crédito e a distribuição de renda (Espich *et al.*, 2021). Além disso, desempenham um papel significativo no desenvolvimento social e ambiental das regiões onde estão inseridas, muitas vezes integrando essas responsabilidades em sua missão empresarial, contribuindo para a melhoria da qualidade de vida de seus associados (Souza; Schmidt, 2020).

Para Cavinato (2023), as cooperativas de crédito desempenham um papel crucial na disponibilização de serviços financeiros a pequenas comunidades, especialmente em cidades com até 10 mil habitantes. Essas atividades sociais beneficiam não apenas cooperados e colaboradores, mas também investidores privados e outros agentes que podem se beneficiar do desenvolvimento econômico e social destas regiões (Martínez-Campillo; Fernández-Santos; Sierra-Fernández, 2016).

Diferentes estudos (Silva, Santos, Ranciaro, 2023; Simamora *et al.*, 2023; Vieira, Bressan V., Bressan A, 2018; Almehdawe *et al.*, 2021) abordam a respeito da importância da análise de desempenho das operações nas cooperativas de crédito, e embora não tenham como objetivo a obtenção de lucros, as cooperativas têm a necessidade de gerir seus recursos com eficiência visando apresentar aos cooperados maiores rendimentos de seus investimentos e melhores taxas para operações, compreendendo resultados que subsidiem sua função social e econômica e que garantam sua sobrevivência no mercado financeiro.

A metodologia CAMEL gera um volume bastante elevado de indicadores financeiros, um total de 33 índices contábeis. Esse volume de informação pode gerar dificuldades no processo de análise. Uma forma de diminuir essa dificuldade é o emprego de aprendizado de máquinas para a compreensão de características das organizações. Segundo Nandi, Jana e Das

(2023), o uso de machine learning vem sendo muito empregado na previsão de cenários no mercado financeiro, tendo um potencial significativo no processamento de dados.

Assim, é fundamental identificar os indicadores financeiros mais importantes para a classificação e identificação das cooperativas com possibilidade de dificuldades no futuro. Para classificar as cooperativas de crédito com maior risco de descontinuidade, o uso de técnicas de aprendizado de máquina pode ser útil.

3 METODOLOGIA

Foram utilizados dados financeiros dos balanços patrimoniais e informações cadastrais fornecidas pelo Banco Central do Brasil (Bacen) entre 2018 e 2022 com o objetivo de identificar os indicadores mais relevantes para a classificação das cooperativas de crédito entre aquelas com risco de encerramento. Para calcular os indicadores CAMEL, os dados financeiros foram coletados usando o padrão COSIF, considerando apenas os resultados semestrais de julho e dezembro. Os CNPJs foram responsáveis por conectar os dados cadastrais aos financeiros. A variável dependente foi criada a partir das informações cadastrais de dezembro de 2022, indicando se a cooperativa estava ativa (autorizada e em atividade) ou encerrada nesta data.

Posteriormente, os dados foram consolidados, removendo-se as linhas com ausência de informações, e foram calculados os 31 indicadores financeiros conforme a metodologia CAMEL, utilizando as fórmulas de Resende (2012) e implementadas por Iurovski, Nascimento e Carvalho (2022), para os cinco grupos de avaliação da metodologia CAMEL.

Silva *et al.* (2017), pontuam que os indicadores CAMEL foram desenvolvidos para avaliar o desempenho das instituições financeiras e para auxiliar no monitoramento e acompanhamento dessas entidades.

Considerando a natureza dos dados, que se trata de índices financeiros calculados com base em informações contábeis, muitas vezes, quando não há informações nos denominadores, os indicadores podem ter valores extremos ou distorcidos devido à natureza desses dados. Isso pode prejudicar a análise estatística e a interpretação dos resultados. Para minimizar esse impacto e garantir uma análise mais consistente e representativa, os dados foram winsorizados. A winsorização limita os valores extremos a um percentil específico (Mahat; Ch'ng, 2020), melhora a tendência central dos dados e diminui os outliers causados por condições extremas ou erros de medição. Essa técnica foi aplicada para tratar os valores extremos em 1%.

A base de dados inicial continha toda a população de cooperativas de crédito brasileiras. No entanto, após excluir cooperativas não singulares, paralisadas, em intervenção ou canceladas, além de observações com erros ou valores inválidos (NaN) em algum dos indicadores CAMEL, a amostra final foi composta por 8.852 observações de 952 cooperativas que operaram no período, das quais 792 estavam ativas em dezembro de 2022.

Os indicadores foram então classificados conforme a adequação de capital (C), a qualidade dos ativos (A), a qualidade da gestão (M), a rentabilidade (E) e a liquidez (L).

Os indicadores financeiros relacionados à adequação de capital (C) objetivam avaliar a capacidade da instituição de cobrir suas perdas e riscos com seus recursos próprios, garantindo sua solvência e estabilidade financeira. Esses índices financeiros avaliam a solidez da instituição financeira, baseando-se na relação entre ativos e capital próprio. Isso reflete a capacidade da instituição de cobrir possíveis perdas com recursos próprios (Varga; Bánkuti, 2021).

Os indicadores associados à qualidade dos ativos (A), avaliam a qualidade e a composição dos ativos da instituição, incluindo empréstimos e investimentos, para determinar o risco associado a esses ativos. Eles analisam a capacidade dos ativos de gerar receita e identificam possíveis riscos, como inadimplência ou desvalorização (Varga; Bánkuti, 2021),

além de monitorar seu desempenho ao longo do tempo (Iurovski; Nascimento; Carvalho, 2022).

Os indicadores CAMEL associados à qualidade da gestão (M), avaliam a eficiência operacional e a gestão de custos das instituições financeiras, medindo a proporção das despesas administrativas e operacionais em relação às receitas e ativos, e monitorando a evolução das operações de crédito e receitas operacionais. Dessa forma, examinam as práticas de gestão e governança da instituição, incluindo sua eficácia na gestão de riscos, conformidade regulatória e tomada de decisões estratégicas (Resende, 2012).

O grupo de indicadores associados à rentabilidade e lucratividade (E), está focado na eficiência operacional e na capacidade de gerar receita das instituições financeiras. Esses indicadores utilizam as contas de resultados credoras e devedoras em comparação com a Receita Operacional e os ativos da instituição como base para análise (Iurovski; Nascimento; Carvalho, 2022).

Os indicadores de liquidez (L) têm como objetivo avaliar a capacidade da instituição financeira de cumprir suas obrigações de curto prazo com mínima perda. Quanto maior o índice, maior é a capacidade da organização de transformar seus ativos em dinheiro rapidamente (Varga; Bánkuti, 2021).

Após o cálculo de cada indicador, procedeu-se à identificação em cada grupo do indicador de maior impacto na previsão de encerramento das atividades das cooperativas. Para isso, utilizou-se o método Random Forest, um algoritmo de aprendizado de máquina, reconhecido por sua versatilidade e eficácia em várias tarefas de classificação, previsão e regressão (Liu; Wang; Zhang, 2012; Schonlau; Zou, 2020). O Random Forest foi implementado para classificar os indicadores financeiros por meio de árvores de decisão independentes construídas com amostras aleatórias, contribuindo para reduzir o viés tendencioso do modelo.

Este método desempenha um papel fundamental no treinamento do comportamento dos dados. Nele, algumas amostras dos dados de treino são selecionadas, formando as árvores da floresta, cada árvore é formada por características aleatórias que contribuem para a escolha do melhor modelo preditivo (Leite; Moraes; Lopes, 2021). Após a identificação dos indicadores mais relevantes para a previsão do encerramento das cooperativas pelo Random Forest, esses indicadores foram submetidos ao algoritmo KNN.

O KNN (K Nearest Neighbors) é um método de classificação de aprendizagem supervisionada não paramétrica, que utiliza dados próximos para fazer previsões ou classificações (Bansal; Goyal; Choudhary, 2022). Uma vantagem importante deste método é sua capacidade de lidar eficazmente com conjuntos de dados grandes, fornecendo resultados precisos, rápidos e robustos (Shokrzade *et al.*, 2021). Ele opera com a suposição de que dados semelhantes geralmente estão próximos uns dos outros. A escolha do KNN foi motivada pelo seu modo de classificação, que leva em conta o comportamento da amostra, tornando-o adequado para o processamento de dados financeiros (Santos; Costa, 2023).

Segundo Nguyen, *et al.* (2021), os vetores gerados direcionam o comportamento de cada conjunto de dados, com cada árvore representando uma nova subdivisão e selecionando aleatoriamente entre as K melhores possibilidades. Cada árvore faz previsões, e a classe de maior importância é definida como padrão da modelagem. Isso reduz o impacto de decisões tomadas com dados únicos e aleatórios, minimizando o viés tendencioso do modelo (Moraes; Martínez, 2015).

Para este estudo, 80% da população foi separada como grupo de treino, enquanto os demais foram utilizados como teste. Esse sistema pressupõe que dados semelhantes tendem a estar próximos uns dos outros (Bastos, Nascimento, Laretto, 2013). Diferentemente de outros algoritmos que se baseiam em modelos diferentes, o KNN é considerado o método mais adequado para o processamento de dados financeiros (Qi, Gu, Wang, 2021).

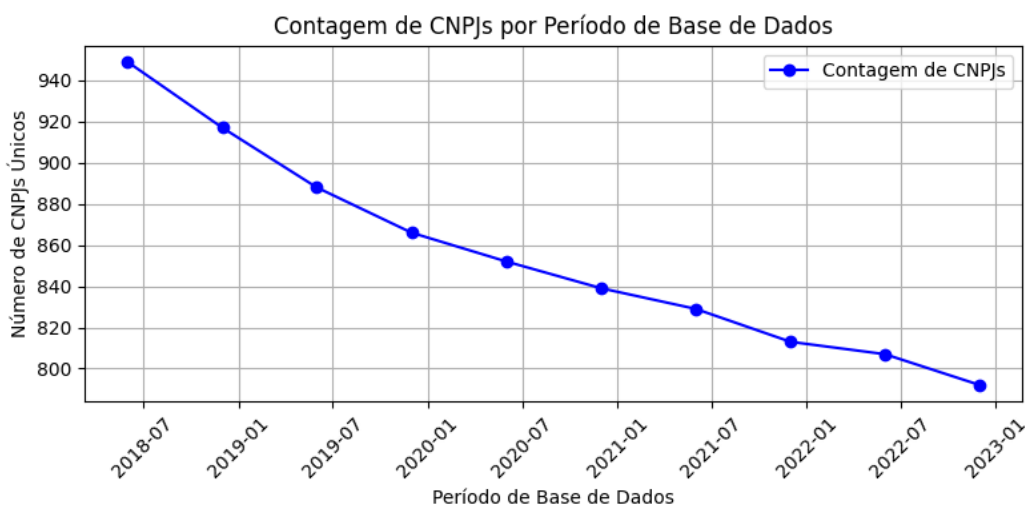
Ao final deste processo, os resultados obtidos serão analisados e apresentados, contribuindo para uma melhor compreensão do potencial de encerramento das cooperativas de crédito e subsidiando a tomada de decisão por parte dos reguladores e investidores do setor financeiro.

4 RESULTADOS E DISCUSSÃO

4.1 Análise descritiva e teste de médias

O estudo analisou dados semestrais de cooperativas de crédito brasileiras de 2018 a 2022, totalizando 952 cooperativas singulares e 8.552 observações. A quantidade de cooperativas ativas diminuiu ao longo dos semestres, partindo de 949 no primeiro semestre de 2018 para 792 no segundo semestre de 2022, como ilustrado na Figura 1. No primeiro semestre de 2018, a amostra era composta por 949 cooperativas de crédito singulares.

Figura 1– Número de cooperativas ativas por semestre



Fonte: Produção própria

Os resultados obtidos nos indicadores CAMEL foram classificados entre as cooperativas que operaram entre 2018 e 2022, agrupando esses dados entre as cooperativas que estavam ativas em dezembro de 2022, ou seja, que continuaram suas atividades, e aquelas que encerraram suas atividades antes dessa data. O teste de normalidade de Shapiro-Wilk mostrou que nenhuma das variáveis segue uma distribuição normal, exigindo o uso de métodos não paramétricos para comparar as médias. Ao aplicar o teste de Mann-Whitney U, verificou-se que, com exceção de C02, todos os índices CAMEL apresentaram diferenças estatisticamente significativas entre os dois grupos.

Os indicadores de Adequação do Capital (C01 a C05), com exceção de C02, foram estatisticamente significativos para distinguir as cooperativas que estariam ativas das encerradas em dezembro de 2022. A variável C01 apresentou uma média de 4.847 para cooperativas ativas contra 3.916 para encerradas, indicando que ambos os grupos de cooperativas utilizaram menos recursos próprios do que o recomendado pelo Banco Central (entre 6 e 12).

Os indicadores de Qualidade dos Ativos (A01 a A05) revelaram que as cooperativas ativas apresentaram maior renda de intermediação financeira (A01), crescimento de ativos (A04) e maior proporção de depósitos em relação aos ativos (A05), enquanto as encerradas mostraram maior proporção de créditos de alto risco (A02 e A03).

Os indicadores de qualidade da gestão (M) avaliam principalmente a eficiência operacional e a gestão de custos. Eles monitoram despesas administrativas e operacionais, a evolução das operações de crédito e receitas operacionais, fornecendo insights sobre a eficácia na gestão de riscos. Observou-se que os indicadores M01, M04 a M07 apresentaram valores médios superiores para o grupo de cooperativas ativas em 2022. Esses índices estão associados à eficiência operacional (M01), crescimento das operações de crédito (M04), evolução das receitas operacionais (M05), aumento das despesas operacionais (M06) e custo de gerenciamento dos ativos (M07). Com exceção dos indicadores M02 e M03, que relacionam receitas e custos administrativos, todos os demais índices evidenciam um desempenho superior para as cooperativas cujo status em dezembro de 2022 é *ativa*.

Os indicadores de rentabilidade e lucratividade (E) estão focados na eficiência operacional e na capacidade de gerar lucros (sobras) das instituições financeiras. Observe que os indicadores E01, E02, E07, E08, E11 e E12, que proporcionam uma visão mais abrangente do desempenho financeiro global, considerando retorno sobre capital próprio, retorno sobre ativos e rentabilidade operacional, são, em média, superiores para as cooperativas cujo status no final de 2022 seria *ativa*. Por outro lado, os índices E03, E04, E05 e E06, que focam na eficiência e rentabilidade específicas das operações de crédito e intermediação financeira, são superiores para as cooperativas com status futuro de *encerradas*. Essas observações sugerem que as diferenças nos resultados financeiros podem estar relacionadas à ênfase na gestão global de recursos e ou na sustentabilidade operacional ao longo do tempo.

Os indicadores de liquidez (L) como L03, L04 e L05 têm denominadores que podem resultar em erros de cálculo quando são muito baixos ou zero, por isso, foram removidos da amostra. Os indicadores de liquidez L01 e L02 comparam, respectivamente, as disponibilidades com o total de ativos, e a proporção dos ativos de curto prazo em relação ao total de ativos. Os resultados mostraram uma situação contraditória: o indicador L01 foi superior para o grupo de cooperativas cujo status futuro seria encerrado, enquanto o L02 foi mais alto para o grupo de cooperativas ativas. O indicador L02 representa o percentual de ativos de rápida liquidez (ativos de curto prazo) representando a facilidade em que a instituição financeira tem para mitigar os riscos da operação, enquanto o L01 trata-se da liquidez imediata (Iurovski; Nascimento; Carvalho, 2022).

Em síntese, os resultados indicam diferenças significativas nos indicadores CAMEL entre cooperativas ativas em 2022 e aquelas que encerram suas atividades entre 2018 e 2022. A próxima fase do estudo empregará o método Random Forest para identificar, dentre esses, os indicadores financeiros mais relevantes na classificação de cooperativas em risco de encerramento.

4.2 Random Forest

O Random Forest é conhecido por sua capacidade de identificar variáveis significativas e avaliar a importância de cada uma para uma classificação precisa de dados (Janitza; Tutz; Boulesteix, 2016). O objetivo deste estudo foi determinar as variáveis da metodologia CAMEL mais importante para a classificação das cooperativas de crédito entre aquelas com maior risco de encerramento.

Este método, que funciona baseado em um conjunto de árvores de decisão, apresenta resultados robustos em operações de classificação (Belgiu; Drăguț, 2016). Proporcionando uma análise confiável e precisa, ele determina a classe de um dado de entrada por meio do voto majoritário de cada árvore individual.

Como pode-se observar na Tabela 1 o nível de acurácia do modelo com todas as variáveis foi de 93%, indicando que 93% das classificações das cooperativas em relação à sua situação operacional futura (ativa ou encerrada) foram corretas. A matriz de confusão evidencia

que o modelo foi bastante preciso em classificar as cooperativas ativas – precisão de 0.94, recall de 0.99 e f1-score de 0.96. Esses valores indicam um excelente desempenho na identificação correta dos casos da Classe 0, ou seja, cooperativas não encerradas no futuro. No entanto, o desempenho do modelo foi significativamente inferior, com uma precisão de 0.77, recall de 0.20 e f1-score de 0.32, para o grupo de cooperativas encerradas. Esses resultados indicam que o modelo tem dificuldades em identificar corretamente as cooperativas de créditos encerradas, resultando em muitos falsos negativos, conforme matriz de confusão. Isso sugere a necessidade de técnicas adicionais para lidar com o desequilíbrio de classes (a quantidade de cooperativas da classe ativas é superior a quantidade de cooperativas que encerraram suas atividades no período de análise, cooperativas encerradas) e melhorar a identificação das cooperativas encerradas (o que será realizado nas próximas seções).

Tabela 1. Resultados do Modelo Random Forest

Métrica	Ativas	Encerradas	Macro Média	Média Ponderada	Geral
Acurácia					0.929
Precisão	0.94	0.62	0.78	0.91	
Recall	0.99	0.19	0.59	0.93	
F1-Score	0.96	0.29	0.63	0.91	
Suporte	1579	132	1711	1711	1711
Matriz de Confusão					
- Verdadeiro Positivo	1564	25			
- Falso Positivo	15	107			

Fonte: Produção própria

Importante ressaltar que o emprego da metodologia Random Forest buscou medir a importância de cada variável em relação a condição futura das cooperativas (ativa ou encerrada em dezembro de 2022). Desse modo, por essa metodologia de aprendizado de máquinas é possível obter o peso relativo de cada indicador no modelo de classificação. Quanto maior o valor atribuído a uma variável, mais significativa é sua contribuição para identificar as cooperativas com risco de encerramento. Assim, objetivou-se identificar quais características financeiras são mais importantes para a classificação da situação operacional futura dessas organizações.

Os resultados podem ser verificados na Tabela 2, a importância relativa das variáveis é indicada pela Importância e pelo Ranking. Quanto maior o valor da “importância”, mais significativa será a variável para o modelo. Assim, observa-se que as variáveis foram agrupadas conforme metodologia CAMEL e, tanto o ranking geral de cada uma delas (terceira coluna), quanto o ranking relativo (sexta coluna), possibilitam identificar as variáveis de maior importância.

Observa-se que as variáveis mais influentes no modelo foram M02, E11, C03, M07 e C04. Isso sugere que fatores como a rentabilidade do Patrimônio Líquido (E11), taxa de crescimento do Patrimônio Líquido Ajustado (C03), custos associados ao gerenciamento de ativos (M07) e a evolução do patrimônio, que representa os recursos próprios cobrindo os riscos das aplicações (C04), desempenham papéis críticos na previsão do encerramento de cooperativas. Essas variáveis são, portanto, indicadores-chave que podem oferecer insights valiosos para estratégias de gestão e mitigação de riscos no setor cooperativista.

Tabela 2. Ranking de Importância das Variáveis Utilizando Random Forest

Variável	Importância	Ranking	Variável	Importância	Ranking
C01:	0.029	21	C03:	0.043	1
C02:	0.037	6	C04:	0.038	2
C03:	0.043	3	C02:	0.037	3
C04:	0.038	5	C05:	0.037	4

C05:	0.037	7	C01:	0.029	5
A01:	0.033	10	A03:	0.037	1
A02:	0.034	9	A02:	0.034	2
A03:	0.037	8	A01:	0.033	3
A04:	0.030	19	A04:	0.030	4
A05:	0.027	26	A05:	0.027	5
M01:	0.027	27	M02:	0.047	1
M02:	0.047	1	M07:	0.038	2
M03:	0.031	17	M04:	0.032	3
M04:	0.032	12	M03:	0.031	4
M05:	0.026	28	M06:	0.028	5
M06:	0.028	23	M01:	0.027	6
M07:	0.038	4	M05:	0.026	7
E01:	0.033	11	E11:	0.044	1
E02:	0.028	24	E01:	0.033	2
E03:	0.032	13	E03:	0.032	3
E04:	0.032	16	E04:	0.032	4
E05:	0.030	20	E10:	0.031	5
E06:	0.026	29	E05:	0.030	6
E07:	0.025	30	E12:	0.029	7
E08:	0.023	31	E02:	0.028	8
E09:	0.027	25	E09:	0.027	9
E10:	0.031	18	E06:	0.026	10
E11:	0.044	2	E07:	0.025	11
E12:	0.029	22	E08:	0.023	12
L01:	0.032	15	L02:	0.032	1
L02:	0.032	14	L01:	0.032	2

Fonte: Elaboração própria

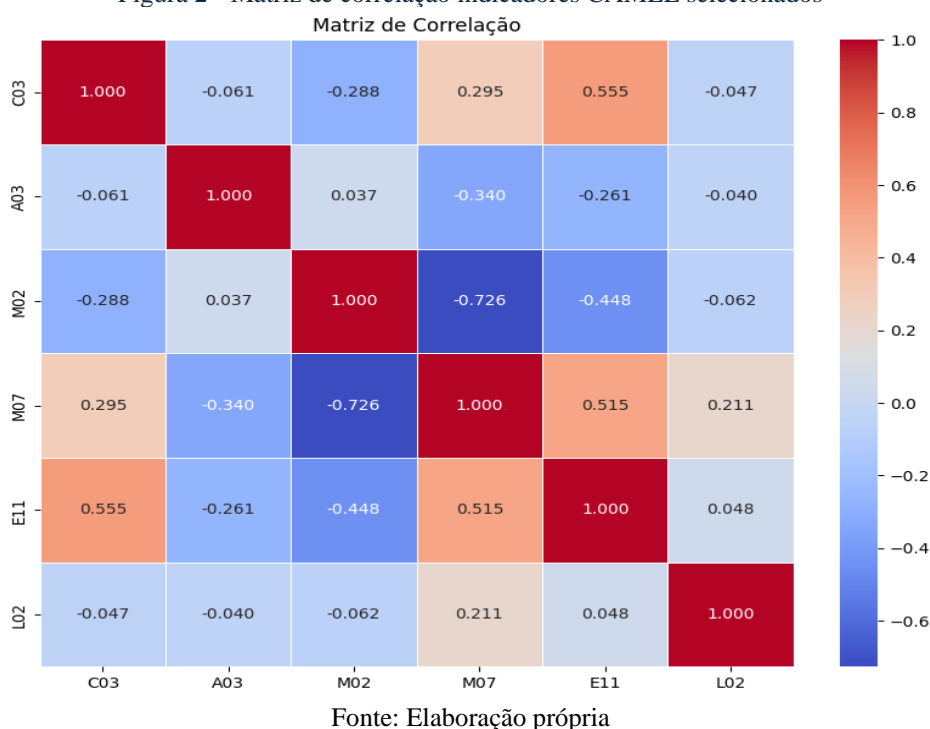
Por meio da classificação de importância dos indicadores utilizando Forest Random, identificou-se que as variáveis C02 e C03 são altamente correlacionadas, desse modo, optou-se por excluir a variável C04, quinta variável mais relevante, mas altamente correlacionada com a primeira variável em termos de importância, e acrescentar as variáveis A03, que mede a proporção não provisionada de ativos de alto risco em relação ao Patrimônio Líquido ajustado, e L02, que apresenta a proporção dos ativos de curto prazo em relação ao total de ativos às próximas etapas do estudo.

Desta forma, mesmo considerando os desafios mencionados anteriormente e apresentados na matriz de confusão, e considerando que esta etapa do estudo visava identificar as variáveis mais importantes, optou-se por selecionar essas variáveis conforme a metodologia Random Forest e analisá-las conjuntamente utilizando o modelo K-Vizinhos Mais Próximos (KNN).

4.3 K-Vizinhos Mais Próximos (KNN)

Na seção anterior, por meio da metodologia de aprendizado de máquinas Random Forest, foi possível identificar as variáveis mais relevantes para a determinação da situação operacional futura das cooperativas de crédito. Por meio dessa metodologia selecionaram-se as variáveis C03, A03, M02, M07, E11 e L02. Assim, os indicadores de maior importância, anteriormente citados, foram plotados em uma matriz de correlação, conforme Figura 2, com o intuito de identificar o grau de relação entre os selecionados visando a exclusão de indicadores com características semelhantes.

Figura 2 - Matriz de correlação indicadores CAMEL selecionados



Observa-se que os índices E11 e C03, que representam, respectivamente, a rentabilidade operacional em relação ao Patrimônio Líquido e a variação da taxa de crescimento do Patrimônio Líquido ajustado, apresentaram um nível de correlação de 55%. A relação entre E11 e M07, que mede o custo associado ao gerenciamento do Ativo Total, foi de 0,515. Por outro lado, as variáveis M02 e M07, que fazem parte do grupo de variáveis associados à qualidade da gestão, apresentaram correlação negativa de 0,726. No entanto, enquanto o índice M02 tem como finalidade avaliar se as receitas de serviços cobrem totalmente os custos administrativos, o indicador M07, como já mencionado, mede o custo associado ao gerenciamento de todos os ativos. Todas as demais variáveis apresentaram nível de correlação em módulos inferiores a 50%.

O modelo *K-Nearest Neighbors* (KNN) foi aplicado para prever as cooperativas com risco de encerramento com base nas variáveis selecionadas. Verifica-se na Tabela 3 que a acurácia foi de 91.8% ao comparar dados de cooperativas que encerraram suas atividades entre 2018 e 2022, das que se mantiveram em funcionamento após 2022. Para definir de forma mais precisa a acurácia dos dados, utilizou-se a F1-score, métrica que calcula a média harmônica entre precisão e recall (Kumar; Gupta, 2015), neste caso, o resultado obtido foi de 90% de assertividade do modelo KNN para previsão de encerramento das atividades das cooperativas de crédito no Brasil.

Tabela 3. Resultados do Modelo KNN

Métrica	Ativas	Encerradas	Macro Média	Média Ponderada	Geral
Acurácia					0.918
Precisão	0.93	0.46	0.69	0.89	
Recall	0.98	0.15	0.57	0.92	
F1-Score	0.96	0.23	0.59	0.90	
Suporte	2364	202	2566	2566	2566
Matriz de Confusão					
- Verdadeiro Positivo	2327	37			
- Falso Positivo	171	31			

Fonte: Produção própria

Os altos valores de acurácia e de F1-score, indicam que o modelo tem um bom desempenho geral em prever corretamente se uma cooperativa estará encerrada no futuro ou não. No entanto, uma análise mais detalhada das métricas de classificação revela algumas limitações importantes. A performance para a classe de cooperativas encerradas foi inferior, com precisão de 0.46 e recall de 0.15. A precisão de 46% indica que, do total de cooperativas classificadas como encerradas, somente 46% realmente estavam encerradas em dezembro de 2022. E o recall de 0,15 indica que, entre todas as cooperativas que realmente encerraram, o modelo foi capaz de identificar corretamente apenas 15%. A matriz de confusão reflete essa dificuldade, mostrando um número significativo de falsos negativos. Para melhorar a capacidade de identificar cooperativas que estão em risco de encerramento é necessário considerar outras técnicas estatísticas.

5 CONSIDERAÇÕES FINAIS

O objetivo deste estudo foi identificar os indicadores contábeis mais importantes para classificar as cooperativas de crédito singulares brasileiras entre aquelas com maior risco de encerramento. Para isso utilizou-se os indicadores CAMEL e técnicas de aprendizado de máquinas. A base de dados foi construída com base nas demonstrações financeiras disponíveis no sítio do Banco Central do Brasil e em conformidade com o padrão COSIF, além das informações cadastrais dessas entidades. Utilizaram-se dados financeiros de 2018 a 2022, bem como informações sobre a situação operacional das cooperativas de crédito em dezembro de 2022.

Os resultados dos testes não paramétricos de diferenças de média mostraram que 30 dos 31 indicadores CAMEL analisados apresentavam diferenças médias estatisticamente significativas nos anos anteriores ao encerramento. Por esse resultado, é possível concluir que o desempenho das cooperativas que estariam encerradas no futuro já apresentava indícios de problemas operacionais em períodos anteriores.

O método Random Forest possibilitou ranquear os indicadores financeiros por seu nível de importância geral, comparando todos os indicadores, e importância específica, comparando somente os indicadores de cada uma das dimensões CAMEL. Assim, pode-se selecionar os mais importantes para a análise. A amostra foi dividida em um grupo teste formado por 20% da amostra total e um grupo de treinamento com 80% restante. O nível de acurácia geral do modelo foi de 0,929, indicando que foi possível identificar corretamente a situação das cooperativas do grupo teste em ativas ou encerradas em aproximadamente 93% dos casos.

No entanto, somente 46% das cooperativas classificadas como encerradas no grupo teste estavam realmente encerradas ao final do período de análise. Enquanto 93% das cooperativas classificadas como ativas estavam realmente ativas. Como resultado, embora o modelo tenha um desempenho geral positivo, com 92,90% de acerto em média, ele não é eficaz na identificação de cooperativas com risco de encerramento, objetivo principal do estudo. Assim, pode-se concluir que o objetivo do Random Forest, que era encontrar os indicadores financeiros mais relevantes, foi atingido, mas essa metodologia é ineficaz na classificação das cooperativas em risco de encerramento

A taxa de crescimento do patrimônio líquido ajustado (C03), o percentual não provisionado do risco de crédito (A03), a cobertura das despesas administrativas pelas receitas de serviços (M02), os custos relacionados ao gerenciamento de ativos (M07), a rentabilidade operacional em relação ao patrimônio líquido (E11) e a liquidez determinada pela relação entre ativos de crédito e ativos totais (L02) foram as variáveis mais relevantes identificadas pelo modelo de aprendizado de máquinas Random Forest.

O método K-Vizinhos Mais Próximos (KNN) foi utilizado com o intuito investigar a possibilidade de, com base nas variáveis mais relevantes determinadas pelo Random Forest, identificar as cooperativas de crédito em risco de encerramento. A precisão do modelo foi alta (91,8%), mesmo com somente os seis indicadores financeiros mencionados anteriormente. No entanto, a precisão em relação ao grupo de cooperativas classificadas como encerradas foi de apenas 46%. Pode-se concluir que a utilização de seis indicadores atinge um nível de informação semelhante ao obtido pelo uso da totalidade dos indicadores CAMEL. Além disso, por conta do elevado percentual de erro na classificação ou identificação das cooperativas de crédito em risco de encerramento, é possível verificar a necessidade de utilização de outros modelos ou técnicas para a previsão de encerramento de cooperativas de crédito singulares.

6 REFERÊNCIAS

ALMEHDAWE, E. *et al.* Factors affecting Canadian credit unions' financial performance. **Agricultural finance review**, v. 81, n. 1, p. 51–75, 2021.

ANAKPO, G. *et al.* Sustainability of Credit Union: A Systematic Review of Measurement and Determinants. **Journal of African Business**, p. 1–22, 9 mar. 2023.

ARAÚJO, E. A.; CARMONA, C. U. M. Efficiency of microcredit institutions: An application of DEA/VRS to the Brazilian case. **Producao**, v. 25, n. 3, p. 701–712, 2015.

BANSAL, M.; GOYAL, A.; CHOUDHARY, A. A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short-Term Memory algorithms in machine learning. **Decision Analytics Journal**, v. 3, p. 100071, jun. 2022.

BASTOS, D.; NASCIMENTO, P.; LAURETTO, M. Proposta e Análise de Desempenho de Dois Métodos de Seleção de Características para Random Forests. In: **Anais do IX Simpósio Brasileiro de Sistemas de Informação**. SBC, 2013. p. 49-60.

BELGIU, M.; DRĂGUȚ, L. Random Forest in remote sensing: A review of applications and future directions. **ISPRS journal of photogrammetry and remote sensing: official publication of the International Society for Photogrammetry and Remote Sensing**, v. 114, p. 24–31, 2016.

BRASIL (2019). Banco Central do Brasil. **Relatório de Economia Bancária 2019**.

Disponível

em:

<https://www.bcb.gov.br/content/publicacoes/relatorioeconomiabancaria/REB_2019.pdf>.

BRASIL (2020). Banco Central do Brasil. **Relatório de Economia Bancária 2020**.

Disponível

em:<https://www.bcb.gov.br/content/publicacoes/relatorioeconomiabancaria/REB_2020.pdf>

BRASIL (2021). Banco Central do Brasil. **Relatório de Economia Bancária 2021**.

Disponível

em:<https://www.bcb.gov.br/content/publicacoes/relatorioeconomiabancaria/REB_2021.pdf>

CANASSA, B. J.; ZANCAN, F. COSTA, D. R. Credit Union Life Cycle and Membership: Evidence from Brazilian Credit Unions. **Revista Contabilidade, Gestao E Governanca**, v. 25, n. 1, 2022.

CARVALHO, F. L. DE *et al.* Exit and Failure of Credit Unions in Brazil: A Risk Analysis. **Revista Contabilidade & Finanças**, v. 26, p. 70–84, 2015.

- CAVINATO, N. R.; CAPITANI, D. H. D. Eficiência social das cooperativas de crédito rural no Brasil: uma análise sob a ótica regional. **Economia e Sociedade**, v. 32, n. 2, p. 429–455, maio 2023.
- DADHICH, M. *et al.* Analytical study of financial wellbeing of selected public and private sector banks: A CAMEL approach. **2021 Emerging Trends in Industry 4.0 (ETI 4.0)**. Índia, 2021.
- ESPICH, D. *et al.* Estudo bibliográfico sobre o tema eficiência em cooperativas de crédito. **Revista de Gestão e Organizações Cooperativas**, v. 8, n. 15, p. 01-31, 2021.
- FRIED, H. O.; KNOX LOVELL, C. A.; EECKAUT, P. V. Evaluating the performance of US credit unions. **Journal of Banking & Finance**, v. 17, n. 2-3, p. 251–265, abr. 1993.
- GONÇALVES, C. M.; BRESSAN, V. G. F.; SOUZA, G. H. D. Camel e Pearls: comparativo entre metodologias de análise de desempenho financeiro de cooperativas de crédito. **Reunir Revista de Administração Contabilidade e Sustentabilidade**, [s.l.], v. 13, n. 1, p. 17-38, 22 mar. 2023
- GRAÇA-BRASIL, L., *et al.* Cooperativas de Crédito e a Pandemia: Uma análise de possíveis impactos no Sistema Cooperativo Financeiro. **Revista Paraense de Contabilidade-RPC**, v. 5, n. 1, p. 36-50, 2020.
- IUROVSCHI, Y. Z.; NASCIMENTO, R. P.; CARVALHO, F. L. Desempenho Financeiro de Bancos Brasileiros em períodos de crise: Avaliação a partir dos indicadores Camel. **Contabilometria**, v. 9, n. 2, 2022.
- JANITZA, S.; TUTZ, G.; BOULESTEIX, A.-L. Random Forest for ordinal responses: Prediction and variable selection. **Computational statistics & data analysis**, v. 96, p. 57–73, 2016.
- JANSEN, A. C.; MAEHLER, A. E.; WEGNER, D. Cooperative Governance and the Legitimacy Dilemma: A Case Study in a Credit Cooperative. **Revista Ibero-Americana de Estratégia**, v. 17, n. 3, p. 61–80, 13 set. 2018.
- KARMA; MADE, I.; NI. A Health Level and Risk Rating Monitoring System of Village Credit Institutions. **International journal of management, IT and social sciences**, v. 9, n. 6, p. 788–799, 30 set. 2022.
- KUMAR, S.; GUPTA, P. Comparative analysis of intersection algorithms on queries using precision, recall and F-score. **International Journal of Computer Applications**, v. 130, n. 7, p. 28-36, 2015.
- LEITE, D. R. A.; MORAES, R. M. DE; LOPES, L. W. Método de Aprendizagem de Máquina para Classificação da intensidade do desvio vocal utilizando Random Forest. **Journal of Health Informatics**, Brasil, v. 12, 2021.
- LIANG, D. *et al.* A novel classifier ensemble approach for financial distress prediction. **Knowledge and Information Systems**, v. 54, n. 2, p. 437–462, 11 maio 2017.
- LIU, Y.; WANG, Y.; ZHANG, J. New Machine Learning Algorithm: Random Forest. **Information Computing and Applications**, v. 7473, p. 246–252, 2012.
- MARTÍNEZ-CAMPILLO, A.; FERNÁNDEZ-SANTOS, Y. PILAR SIERRA-FERNÁNDEZ, M. How Well Have Social Economy Financial Institutions Performed During the Crisis Period? Exploring Financial and Social Efficiency in Spanish Credit Unions. **Journal of Business Ethics**, v. 151, n. 2, p. 319–336, 11 maio 2016.

- MAHAT, N. I.; CH'NG, C. K. Winsorize tree algorithm for handling outlier in classification problem. **International journal of operational research**, v. 38, n. 2, p. 278, 2020.
- MCKILLOP, D.; WILSON, J. O. S. Credit Unions: A Theoretical and Empirical Overview. **Financial Markets, Institutions & Instruments**, v. 20, n. 3, p. 79–123, 27 jun. 2011.
- MEYER, P. A.; PIFER, H. W. Prediction of Bank Failures. **The Journal of Finance**, v. 25, n. 4, p. 853–868, 1970.
- MORAES, R. M.; MARTÍNEZ, L. Computational Intelligence Applications for Data Science. **Knowledge-Based Systems**, v. 87, p. 1–2, out. 2015.
- NANDI, B.; JANA, S.; DAS, K. P. Machine learning-based approaches for financial market prediction: A comprehensive review. **Journal of AppliedMath**, v. 1, n. 2, p. 134, 2023.
- NGUYEN, J.-M. *et al.* Random Forest of perfect trees: concept, performance, applications and perspectives. **Bioinformatics (Oxford, England)**, v. 37, n. 15, p. 2165–2174, 2021.
- PAULA, D. A. V. *et al.* Estimating credit and profit scoring of a Brazilian credit union with logistic regression and machine-learning techniques. **RAUSP Management Journal**, v. 54, n. 3, p. 321–336, 8 jul. 2019.
- QI, M.; GU, Y.; WANG, Q. Internet financial risk management and control based on improved rough set algorithm. **Journal of Computational and Applied Mathematics**, v. 384, p. 113179, mar. 2021.
- RESENDE, L. L. **Análise do desempenho econômico-financeiro: um estudo ex ante e ex post diante da fusão Itaú Unibanco**. 2012. 1 v. Dissertação (Mestrado) - Contabilidade e Controladoria, Departamento de Ciências Contábeis, Universidade Federal de Minas Gerais, Belo Horizonte, 2012.
- SANCHES, A. C.; MOREIRA, V. R.; FONTANINI, C. A. C. Riscos em operações de troca no agronegócio: Análise de modelo de gerenciamento de riscos para cooperativas. **Revista de Gestão e Organizações Cooperativas**, v. 7, n. 14, p. 01-17, 2020.
- SANTOS, L. S. Z. *et al.* Risco de crédito e eficiência técnica nas cooperativas de crédito brasileiras. **Cadernos EBAPE.BR**, v. 18, n. 4, p. 956–973, out. 2020.
- SANTOS, T. R.S.; COSTA, O. L.V. Sistema de tomada de decisão no mercado de ações utilizando aprendizado de máquina. In: **Anais do II Brazilian Workshop on Artificial Intelligence in Finance**. SBC, 2023. p. 25-36.
- SCHONLAU, M.; ZOU, R. Y. The random forest algorithm for statistical learning. **The Stata Journal: Promoting Communications on Statistics and Stata**, v. 20, n. 1, p. 3–29, mar. 2020.
- SHOKRZADE, A. *et al.* A novel extreme learning machine based KNN classification method for dealing with big data. **Expert Systems with Applications**, v. 183, p. 115293, nov. 2021.
- SILVA, A.; SANTOS, J. F. dos; RANCIARO, A. Performance of Brazilian credit unions: An analysis from PEARLS indicators. **RAM. Revista de Administração Mackenzie**, v. 24, n. 1, p. eRAMR230057, 2023.
- SILVA, T. P. da; *et al.* Financial and economic performance of major Brazilian credit cooperatives. **Contaduría y Administración**, v. 62, n. 5, p. 1442–1459, dez. 2017.
- SIMAMORA, D. H.; MAKALIWE, W. A.; MARDANUGRAHA, E.; MURAD, Z. K. N. Factors Affecting the Financial Performance of Credit Union in Indonesia. **Indonesian Journal of Multidisciplinary Science**, v. 2, n. 4, p. 2365–2376, 31 jan. 2023.

SOUZA, D. M.; SCHMIDT, D. Comparativo de desempenho econômico-financeiro e social entre cooperativas de crédito e bancos comerciais privados não cooperativos. **Revista Eletrônica de Ciências Contábeis**, v. 9, n. 2, p. 79-100, 2020.

VARGA, J; BÁNKUTI, G. Ranking methodology for Islamic banking sectors-Modification of the conventional CAMELS method. **Banks and Bank Systems**, v. 16, n. 1, p. 36-51, 2021.

VIEIRA, L.K.; BRESSAN, V.G.F.; BRESSAN, A.A. Diversification and Performance of Credit Unions. Individual Behaviors and Technologies for Financial Innovations. **Springer, Cham.**, p. 239–262, 27 jul. 2018.