

Aplicação das técnicas de *machine learning* na categorização de despesas de fluxo de caixa: uma pesquisa-ação

*Application of machine learning techniques in the categorization of cash flow expenses:
an action research*

 Felipe Torres Dâmaso¹  Sonia Rosa Arbues Decoster²  Leandro d'Avila da Silva³
 Julielton Costa de Brito⁴

Faculdade Fipecafi, São Paulo, SP, Brasil

Resumo

A gestão financeira desempenha um papel fundamental na estabilidade e no crescimento das empresas. A categorização inadequada de despesas no fluxo de caixa pode acarretar consequências negativas, como relatórios financeiros imprecisos, dificuldades na previsão do fluxo de caixa e problemas na identificação de áreas com custos excessivos ou ineficientes (Silva e Navarro & Valverde, 2023). O objetivo deste relato técnico é apresentar a aplicação de uma ferramenta desenvolvida com base em técnicas de *machine learning* para resolver o problema da categorização incorreta de despesas em uma planilha de fluxo de caixa de uma empresa familiar alagoana do setor varejista de artigos e armarinhos. O método adotado foi a pesquisa-ação, que, no ambiente organizacional, busca frequentemente solucionar problemas de natureza técnica. Devido às inconsistências nas categorias atribuídas manualmente pelos funcionários, foi desenvolvida uma solução utilizando bibliotecas Python para análise de texto e classificação de dados. Modelos de Regressão Logística e *Random Forest* foram aplicados para automatizar a correção das categorias. Como resultado, a aplicação dessas técnicas permitiu uma melhora na precisão da categorização das despesas, alcançando uma acurácia de 94% com o modelo *Random Forest*. Este estudo evidencia a eficácia da integração de *machine learning* em processos financeiros, demonstrando como essas tecnologias podem contribuir para maior precisão e eficiência, reduzindo erros e otimizando a gestão empresarial.

Palavras-chave: fluxo de caixa, análise textual, machine learning, pesquisa-ação

Notas dos autores

Conflito de interesse: Os autores não declararam nenhum potencial conflito de interesse.

Autor correspondente: Sonia Rosa Arbues Decoster - sonia.decoaster@fipecafi.org

Cite como

American Psychological Association (APA)

Dâmaso, F. T., Decoster, S. R. A., Silva, L. d., & Brito, J. C. (2025, jan./abr.). Aplicação das técnicas de machine learning na categorização de despesas de fluxo de caixa: uma pesquisa-ação. *Revista Inovação, Projetos e Tecnologias - IPTEC*, São Paulo, 13(1), 1-26, e27432. <https://doi.org/10.5585/2025.27432>

¹ Possui graduação em Ciências Contábeis - Faculdades da Seune, MBA em Gestão Financeira, Controladoria e Auditoria pela (FGV), MBA International Financial Reporting Standards (IFRS) pela Fipecafi, Mestrando pela Fipecafi em Finanças e Controladoria com foco em pesquisa em Tecnologia Inteligência Artificial e Machine Learning. felippeconsult@gmail.com.br

² Doutora (2015) e Mestre (2008) em Administração de Empresas pela Faculdade de Economia, Administração e Contabilidade (FEA) da Universidade de São Paulo (USP). Pós-graduada em Administração pela Escola de Administração de Empresas de São Paulo - FGV e graduada em Matemática pela Pontifícia Universidade Católica de São Paulo. sonia.decoaster@fipecafi.org

³ Mestrando pela Fipecafi em Finanças e Controladoria com foco de pesquisa em Contabilidade Financeira e Sustentabilidade, pós-graduado em Finanças e Contabilidade Empresarial pela FAE Business school, graduado em Ciências Contábeis e Administração de Empresas pela Univali - Universidade do Vale do Itajaí. ldavila1971@gmail.com

⁴ Possui graduação em Ciências Econômica - Fundação Nilton Lins, MBA em Finanças & Banking Faculdade Uninove, Especialização em Controladoria, Contabilidade e Finanças pela Fipecafi, Mestrando pela Fipecafi em Controladoria e Finanças com foco em Contabilidade financeira. julielton.brito@fipecafi.edu.br

*Application of machine learning techniques in the categorization of cash flow expenses:
an action research*

Abstract

Financial management plays a fundamental role in the stability and growth of companies. Inadequate categorization of expenses in the cash flow can lead to negative consequences, such as inaccurate financial reports, difficulties in cash flow forecasting, and challenges in identifying areas with excessive or inefficient costs (Silva e Navarro & Valverde, 2023). The objective of this technical report is to present the application of a tool developed based on machine learning techniques to address the issue of incorrect expense categorization in the cash flow spreadsheet of a family-owned retail business in Alagoas, specializing in haberdashery and related goods. The method adopted was action research, which often aims to solve technical problems within organizational environments. Due to inconsistencies in the categories manually assigned by employees, a solution was developed using Python libraries for text analysis and data classification. Logistic Regression and Random Forest models were applied to automate the correction of categories. As a result, the application of these techniques improved the accuracy of expense categorization, achieving 94% accuracy with the Random Forest model. This study highlights the effectiveness of integrating machine learning into financial processes, demonstrating how these technologies can contribute to greater accuracy and efficiency, reducing errors and optimizing business management.

Keywords: cash flow, textual analysis, machine learning, action research

Introdução

A gestão financeira eficiente é fundamental para o sucesso e a sustentabilidade de qualquer empresa, especialmente em organizações familiares, onde a formalização de processos pode ser menos rigorosa. Nesses contextos, a gestão financeira desempenha um papel crucial na garantia de estabilidade e crescimento, pois decisões financeiras bem informadas são essenciais para a alocação adequada de recursos, o controle de custos e o planejamento estratégico (Moraes & Oliveira, 2014). A ausência de uma gestão financeira robusta pode resultar em problemas significativos, como a incapacidade de prever fluxos de caixa, dificuldades na obtenção de crédito e até mesmo a falência.

A categorização inadequada de despesas acarreta diversas consequências negativas, incluindo relatórios financeiros imprecisos, dificuldades na previsão de fluxo de caixa e desafios na identificação de áreas com custos excessivos ou ineficientes (Silva e Navarro & Valverde, 2023). Estudos indicam que, em empresas familiares, onde os processos frequentemente apresentam menor formalização, a precisão na categorização de despesas é ainda mais crítica para a saúde financeira da organização (Moraes & Oliveira, 2014). Nesses casos, a categorização tradicionalmente é realizada manualmente pelos funcionários, um

processo suscetível a erros humanos, especialmente em ambientes onde não há um sistema rigoroso de verificação. Tal abordagem não apenas consome tempo, mas também pode introduzir inconsistências significativas nos registros financeiros, comprometendo a confiabilidade dos relatórios e a eficácia da gestão financeira (Géron, 2019).

Com o avanço da tecnologia e a crescente disponibilidade de ferramentas de análise de dados, as técnicas de *machine learning* têm se mostrado promissoras para automatizar e aprimorar a precisão de processos repetitivos e propensos a erros. *Machine learning*, uma subárea da inteligência artificial, permite que sistemas computacionais aprendam a partir de dados e realizem previsões ou decisões sem a necessidade de programação explícita para cada cenário (Russell & Norvig, 2020). No contexto da categorização de despesas, essas técnicas podem automatizar o processo, reduzindo erros humanos e aumentando a eficiência.

Este estudo foca em uma empresa varejista de artigos e armarinhos localizada em Alagoas, na qual foi identificado um problema recorrente na categorização de despesas em suas planilhas de fluxo de caixa. Embora os históricos das despesas estivessem corretamente registrados, as categorias atribuídas frequentemente apresentavam imprecisões, resultando em análises financeiras menos precisas e dificultando a tomada de decisões. Assim, o problema central foi definido da seguinte forma: "*Como o desenvolvimento de uma ferramenta baseada em técnicas de machine learning pode auxiliar na categorização correta de despesas?*"

Neste contexto, o objetivo deste estudo é desenvolver uma solução baseada em *machine learning* para corrigir a categorização de despesas de maneira automatizada e precisa. Por meio de uma abordagem de pesquisa-ação, a proposta é implementar a solução na empresa mencionada, integrando tecnologia e processos internos. Segundo Thiollent (2009), a pesquisa-ação é um tipo de pesquisa social de base empírica que visa tanto a ação quanto a resolução de problemas coletivos. Essa metodologia frequentemente busca resolver problemas técnicos, como a introdução de novas tecnologias ou a melhoria do fluxo de informações.

Inicialmente, as planilhas manuais foram digitalizadas, e os dados inconsistentes foram tratados com técnicas de Processamento de Linguagem Natural (NLP). Os modelos preditivos foram treinados utilizando os dados ajustados, automatizando a categorização de despesas. Com o uso de bibliotecas Python, como NLTK (*Natural Language Toolkit*) e Scikit-Learn, os dados foram pré-processados, vetorizados e empregados no treinamento de modelos de Regressão Logística (RL) e *Random Forest* (RF). Essas bibliotecas são amplamente reconhecidas pela comunidade de ciência de dados devido à sua eficiência e flexibilidade (Géron, 2019). Por fim, a ferramenta foi integrada à rotina financeira, com validações realizadas pela equipe administrativa para ajustar eventuais discrepâncias.

O estudo de Dumitrescu et al. (2022) destaca que os métodos de classificação binária podem ser divididos em duas categorias: estatísticos e de aprendizado de máquina. Entre os métodos de aprendizado de máquina, os métodos *ensemble* baseados em árvores de decisão, como o *Random Forest*, apresentam desempenho superior em contextos como a pontuação de crédito, devido à sua alta capacidade preditiva (Montevechi et al., 2024). Adicionalmente, o RF previne o *overfitting* e é robusto em relação a dados ausentes (Halteh et al., 2018; Petropoulos et al., 2020). Conforme Zöller e Huber (2021), entre os métodos supervisionados, o RF destaca-se por sua precisão preditiva ao combinar diversas árvores de decisão.

A implementação dessas técnicas pode servir como modelo para outras empresas familiares que enfrentam desafios semelhantes na categorização de despesas. A automação desse processo não só melhora a eficiência, mas libera os funcionários para se concentrarem em tarefas mais estratégicas, aumentando o valor agregado ao negócio. A aplicação de *machine learning* na gestão financeira é uma tendência crescente, com o potencial de revolucionar a forma como as empresas gerenciam recursos e tomam decisões (Russell & Norvig, 2020).

Este artigo está estruturado em cinco seções. Além desta introdução, o referencial teórico explorando os temas de gestão financeira, fluxo de caixa, técnicas de *machine learning* e suas aplicações. Na terceira seção, os procedimentos metodológicos são apresentados. A análise dos resultados é detalhada na quarta seção e, por fim, a última seção traz as considerações finais, com limitações da pesquisa e sugestões para estudos futuros.

Referencial teórico

Gestão financeira nas empresas

A gestão financeira é essencial para o sucesso e a sustentabilidade de qualquer empresa. Essa área envolve a administração eficiente dos recursos financeiros da organização para alcançar seus objetivos estratégicos. A importância de uma gestão financeira eficaz é destacada por sua capacidade de auxiliar as empresas a manter a solvência, maximizar lucros e otimizar a alocação de recursos. Estudos indicam que práticas sólidas de gestão financeira são cruciais para a resiliência das empresas em tempos de crise econômica, permitindo que elas se adaptem rapidamente às mudanças do mercado (Harris, Opp & Opp, 2020).

Um estudo de Kumar, Charles e Mishra (2016) revelou que a implementação de práticas avançadas de gestão financeira, como a análise de *big data* e o uso de inteligência artificial, pode melhorar significativamente a eficiência e a precisão nas decisões financeiras. Essas

tecnologias permitem que as empresas analisem grandes volumes de dados em tempo real, identificando tendências e padrões que seriam impossíveis de detectar manualmente.

Conceito de fluxo de caixa e sua importância

Conforme Gitman e Zutter (2017), o “fluxo de caixa é o fator determinante básico do valor de uma organização, cuja meta é assegurar a solvência da empresa, cumprindo as obrigações financeiras nas datas previstas e gerando fluxo de caixa positivo para os seus proprietários”. A importância do fluxo de caixa foi amplamente evidenciada durante a crise financeira de 2008 e a pandemia de COVID-19, períodos em que muitas empresas enfrentaram desafios significativos para manter sua liquidez. De acordo com um estudo de Harris et al. (2020), empresas que monitoram de forma contínua seu fluxo de caixa possuem maior capacidade de sobreviver a crises financeiras, uma vez que conseguem identificar rapidamente problemas de liquidez e adotar medidas corretivas. Além disso, uma gestão eficiente do fluxo de caixa pode auxiliar as empresas a identificar padrões de gasto, prever problemas de liquidez e planejar melhor suas necessidades de financiamento de curto e longo prazo (Koralun-Bereźnicka, 2013). O monitoramento frequente do fluxo de caixa permite que as empresas atinjam seus objetivos estratégicos, proporcionando previsões mais precisas, avaliação da liquidez e da solvência da organização (Hendriksen & Van Breda, 2015).

Técnicas utilizadas na classificação de texto

A classificação de texto envolve a atribuição de categorias predefinidas a textos não rotulados com base em um conjunto de dados de treinamento. Neste estudo, o modelo estatístico da Regressão Logística (RL) foi utilizado como alternativa ao algoritmo de *machine learning* Random Forest (RF), considerando que ambos são amplamente reconhecidos pela eficácia em tarefas de classificação. A Tabela 1 apresenta um descritivo explicativo sobre essas técnicas, detalhando suas características principais e aplicações específicas.

Tabela 1

Técnicas de estatística e aprendizado de máquina utilizadas na classificação de texto

<p>Regressão Logística (RL)</p>	<p>Modelo estatístico amplamente utilizado para prever a probabilidade de um evento binário. É reconhecida por sua eficiência em aplicações como a previsão de inadimplência e pela facilidade na interpretação de resultados (Hosmer et al., 2013). Um exemplo relevante é o estudo de Coelho et al. (2021), no qual a RL foi aplicada para avaliar o risco de crédito. Os autores destacam a capacidade do modelo em lidar com variáveis categóricas e contínuas, facilitando a análise de risco em instituições financeiras. Essa característica torna a RL uma ferramenta versátil e útil em contextos que exigem previsões confiáveis baseadas em múltiplas variáveis.</p>
<p>Random Forest- RF (Floresta Aleatória)</p>	<p>Técnica de aprendizado de máquina do tipo ensemble, que combina os resultados de múltiplas árvores de decisão para produzir um modelo preditivo mais robusto e preciso. Ele é amplamente reconhecido por seu desempenho superior, especialmente em conjuntos de dados grandes e complexos (Lima & Delen, 2020). Machine learning, ou aprendizado de máquina, é um subcampo da inteligência artificial dedicado ao desenvolvimento de algoritmos capazes de aprender a partir de dados, reproduzindo funções cognitivas humanas relacionadas aos processos de aprendizagem e resolução de problemas, com o objetivo de alcançar resultados "ótimos" (Lima & Delen, 2020). Estudos recentes, como o de Altieri, Apro e Basile (2021), compararam algoritmos de classificação, incluindo RF, Árvore de Decisão e RL, na detecção de fake news. O RF destacou-se pela sua eficácia em identificar padrões em grandes volumes de dados. Ele opera construindo diversas árvores de decisão, cada uma treinada em um subconjunto diferente dos dados. As previsões individuais das árvores são combinadas para produzir uma previsão mais robusta e precisa (Avelar et al., 2022). Essa abordagem reduz a variância do modelo, tornando-o ideal para problemas de classificação complexos.</p>
<p>Processamento de Linguagem Natural (Natural Language Processing - NLP)</p>	<p>Subárea da inteligência artificial que visa a interação entre computadores e humanos por meio da linguagem natural. O objetivo do NLP é desenvolver métodos e algoritmos que permitam aos computadores entender e processar a linguagem humana, realizando tarefas como tradução automática, análise de sentimentos, geração de texto e reconhecimento de fala (Chopra, Prashar & Sain, 2013). A análise de sentimentos é uma aplicação comum do machine learning em classificação de texto. Essa técnica categoriza opiniões expressas em textos como positivas, negativas ou neutras. Maia e Salton (2022) implementaram um agente inteligente para classificar sentimentos em textos no idioma português do Brasil, combinando técnicas de machine learning com NLP. Segundo Bondielli e Marcelloni (2019), o uso de machine learning aliado ao NLP pode ser crucial no combate à desinformação nas mídias sociais, onde o compartilhamento de conteúdo falso é frequente. O NLP capacita algoritmos de machine learning a compreender e processar dados textuais, utilizando técnicas como vetorização de texto e remoção de stopwords para preparar os dados para análise. A vetorização transforma o texto em um conjunto de palavras, calculando a frequência de ocorrência de cada termo. Este processo é fundamental para construir modelos preditivos eficazes, que podem ser aplicados em tarefas como a classificação de notícias falsas.</p>

Fonte: Elaborado pelos autores a partir de Hosmer et al., 2013, Coelho et al. (2021), (Lima e Delen, 2020), Altieri, Apro e Basile (2021), (Avelar et al., 2022), Chopra, Prashar & Sain, 2013), Maia e Salton (2022) e Bondielli e Marcelloni (2019).

2.3 Métricas de desempenho de modelos de machine learning

A avaliação de modelos de *machine learning* é uma etapa fundamental para assegurar que os algoritmos desenvolvidos estão desempenhando conforme o esperado. A eficácia de um modelo é mensurada por meio de diversas métricas que quantificam sua precisão, capacidade preditiva e robustez. Conforme apresentado na Tabela 2, as principais métricas utilizadas na avaliação de modelos de *machine learning* incluem acurácia, *recall*, F1-score, matriz de confusão e a área sob a curva ROC (AUC-ROC). Cada uma dessas métricas desempenha um

papel específico na análise do desempenho do modelo, fornecendo informações complementares sobre sua eficácia em diferentes cenários e conjuntos de dados.

Tabela 2

Métricas de Desempenho de Modelos de Machine Learning

<p>Acurácia ou Taxa de Casos Corretamente Classificados</p>	<p>Indica a performance geral do modelo. Dentre todas as classificações, quantas o modelo classificou corretamente. A acurácia é a proporção de previsões corretas feitas pelo modelo em relação ao total de previsões.</p> $\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN}$
<p>Precisão ou Taxa de Positivos Previstos</p>	<p>Dentre as classificações de classe Positivo que o modelo fez, quantas estão corretas.</p> $\text{Precisão} = \frac{VP}{VP + FP}$
<p>Recall/Sensibilidade ou Taxa de Verdadeiros Positivos</p>	<p>Dentre todas as situações de classe Positivo como valor esperado, quantas estão corretas.</p> $\text{Revocação} = \frac{VP}{VP + FN}$ <p>O recall, também conhecido como sensibilidade, é a proporção de verdadeiros positivos (eventos corretamente previstos) em relação ao total de verdadeiros positivos e falsos negativos.</p>
<p>Especificidade</p>	<p>Mede a proporção de casos negativos corretamente classificados por negativos .</p> $\text{Especificidade} = \frac{VN}{VN + FP}$
<p>F1-Score</p>	<p>Média harmônica entre precisão e recall.</p> $F1 = 2 * \frac{\text{Precisão} * \text{Revocação}}{\text{Precisão} + \text{Revocação}}$ <p>O F1-score é a média harmônica da precisão e do recall, fornecendo uma única métrica que balanceia os trade-offs entre as duas. É especialmente útil em situações de desequilíbrio de classes, pois combina a precisão e o recall em uma métrica única que pode ser otimizada (Prati et al., 2008; Santos et al., 2019).</p>
<p>Área Sob a Curva ROC (AUC-ROC)</p>	<p>A AUC-ROC é uma métrica amplamente utilizada para avaliar o desempenho de classificadores. A curva ROC (<i>Receiver Operating Characteristic</i>) é um gráfico que mostra a relação entre a taxa de verdadeiros positivos e a taxa de falsos positivos para diferentes pontos de corte de um modelo preditivo. A área sob essa curva (AUC) quantifica a habilidade do modelo em distinguir entre as classes. Um valor de AUC próximo de 1 indica um excelente desempenho, enquanto um valor de 0,5 sugere que o modelo não tem capacidade discriminativa (Prati et al., 2008; Santos et al., 2019).</p>
<p>Matriz de Confusão</p>	<p>A matriz de confusão é uma ferramenta que fornece uma visão detalhada do desempenho de um modelo de classificação, mostrando as previsões corretas e incorretas separadas por cada classe. A matriz é composta por quatro elementos principais: verdadeiros positivos (VP), verdadeiros negativos (VN), falsos positivos (FP) ou Erro Tipo 1 e falsos negativos (FN) ou Erro Tipo 2. Esses elementos permitem calcular métricas adicionais, como precisão, recall e F1-score, e fornecer uma compreensão mais profunda de onde o modelo está errando</p>

Fonte: Elaborado pelos autores a partir de Santos et al., 2019 e Prati et al., 2008.

2.4 Adoção de tecnologias baseadas em inteligência artificial (IA) pelas empresas familiares

Kumar e Ratten (2024) realizaram uma revisão sistemática com o objetivo de sintetizar as pesquisas existentes sobre a adoção e implementação de inteligência artificial (IA) em empresas familiares. A análise revelou-se incipiente, abrangendo apenas 15 artigos publicados em um intervalo de cinco anos, sendo o primeiro trabalho destacado sobre o tema datado de 2020. A revisão enfatiza o potencial significativo da IA para empresas familiares, especialmente na melhoria de operações, na tomada de decisões e no engajamento com os clientes. Entre as oportunidades identificadas estão a análise de dados de clientes, a simplificação de operações e a otimização da experiência do cliente por meio de tecnologias como IA generativa, *chatbots* e processamento de linguagem natural (NLP).

No entanto, o estudo também aponta desafios consideráveis para a adoção da IA em empresas familiares. Esses desafios incluem restrições de recursos, infraestrutura inadequada, baixa personalização das soluções e lacunas de conhecimento técnico relacionadas à IA. Como contribuição, o estudo propõe um roteiro personalizado para a adoção de IA em empresas familiares e sugere direções futuras de pesquisa baseadas em temas emergentes no uso de IA nesse contexto.

Gilani et al. (2023) destacam como a adoção de *machine learning* permite que empresas familiares desenvolvam modelos preditivos para tendências de mercado e comportamento do consumidor, capacitando-as a tomar decisões informadas e a se adaptar às mudanças dinâmicas do mercado. Por sua vez, McCloskey et al. (2024) exploraram a aplicação do processamento de linguagem natural (NLP) como ferramenta para ajudar empresas familiares a avaliar seus pontos fortes e fracos com base em avaliações de clientes. Apesar das limitações nos conjuntos de dados disponíveis, as técnicas de NLP, bem como os *Large Language Models* (LLMs), também conhecidos como IA generativa, mostraram-se promissoras em fornecer insights coerentes e acionáveis. Esses resultados desafiam a noção de que grandes volumes de dados são indispensáveis para análises eficazes, demonstrando o potencial dessas tecnologias em contextos de dados limitados.

Procedimentos metodológicos

Este estudo caracteriza-se como pesquisa-ação, um método de pesquisa qualitativa de caráter prático que busca solucionar problemas reais por meio da intervenção do pesquisador e dos próprios atores envolvidos no contexto estudado (O'Brien, 2001). Tripp (2005) classifica a modalidade utilizada neste estudo como pesquisa-ação prática, em que o pesquisador projeta as mudanças a serem implementadas. Essa modalidade pode ser associada à intervenção profissional descrita por Macke (2006), que se concentra no problema e, em articulação com o grupo, utiliza os recursos disponíveis para propor soluções possíveis. Além disso, por pressupor um processo de intervenção no ambiente, as ações são diagnosticadas de forma participativa, em que o pesquisador atua como participante ativo (Macke, 2006, pp. 223-224).

As técnicas de coleta de dados empregadas neste estudo, alinhadas à metodologia de pesquisa-ação, incluíram entrevistas, observação participante, grupos focais e pesquisa documental (Barbier, 2007; Macke, 2006; Thiollent, 2007). Segundo Thiollent (2007), essas técnicas são aplicadas de acordo com cada fase ou etapa do processo de investigação. Elas incluem ferramentas para coleta e interpretação de dados, resolução de problemas e organização de ações. A observação participante desempenhou um papel predominante neste estudo, uma vez que a participação ativa dos envolvidos no contexto contribuiu para promover um ambiente favorável, minimizar resistências e viabilizar as estratégias de mudança necessárias (Macke, 2006). Nesse contexto, a observação participante adotada enquadra-se no conceito de *observação participante completa*, conforme definido por Barbier (2007), em que “o pesquisador está implicado desde o início, porque já é membro do grupo antes de iniciar a pesquisa ou torna-se membro por conversão”.

De acordo com Macke (2006), o benefício da pesquisa-ação reside “nas observações e na avaliação das ações definidas em conjunto com os participantes, além da identificação dos obstáculos enfrentados”. Ainda segundo o autor, o método constitui “um meio de explicitar o conhecimento tácito construído a partir da reflexão coletiva sobre a ação”.

A análise da literatura revela diferentes abordagens para a implementação de uma pesquisa-ação (Thiollent, 2009; Tripp, 2005; Vergara, 2005). No entanto, todas seguem, de certa forma, um roteiro estruturado em etapas: diagnóstico (identificação do problema), planejamento da ação, implementação e avaliação dos resultados. Seguindo esse roteiro, uma ferramenta foi desenvolvida com base em técnicas de *machine learning*, complementada por insights qualitativos obtidos por meio de uma entrevista com um cientista de dados acerca da elaboração de um modelo preditivo.

Na fase de diagnóstico, foram realizadas reuniões com a equipe administrativa para identificar as principais dificuldades relacionadas ao processo de categorização manual. A etapa de planejamento incluiu o desenho do modelo preditivo, a seleção das ferramentas de *machine learning* mais adequadas e a digitalização das planilhas financeiras. Os dados foram tratados para remover inconsistências, como categorias duplicadas e registros incompletos. As descrições das despesas foram padronizadas por meio de técnicas de processamento de linguagem natural (NLP), e os modelos preditivos foram ajustados para refletir as especificidades da categorização realizada pela equipe da empresa.

Durante a fase de implementação, foram realizados treinamentos com os colaboradores para assegurar o entendimento e a aceitação da solução proposta. Além disso, observações durante o uso da ferramenta permitiram a coleta de dados qualitativos sobre a interação dos

usuários com o sistema. Nas próximas seções, serão apresentados detalhes sobre a empresa estudada, bem como as etapas de diagnóstico (identificação do problema), planejamento da ação e implementação do roteiro da pesquisa-ação.

O caso da varejista de artigos e armarinhos

O caso analisado refere-se a uma empresa familiar localizada em Alagoas, atuante no setor de varejo de artigos e armarinhos. A escolha dessa empresa justifica-se pela necessidade de explorar um cenário real em que a categorização de despesas enfrenta desafios significativos, devido à natureza manual do processo e à falta de padronização nos registros financeiros. A amostra utilizada neste estudo foi composta por dados financeiros fornecidos pela empresa, incluindo registros de pagamentos e suas respectivas categorias de despesas.

Para assegurar a integridade e a privacidade das informações, os dados foram anonimizados e modificados de maneira a preservar a estrutura e os padrões originais, sem expor informações sensíveis. Essa abordagem permite que a análise mantenha sua relevância e aplicabilidade prática, ao mesmo tempo em que protege a confidencialidade dos dados empresariais (Pedregosa et al., 2011). A empresa selecionada é representativa de muitas outras no setor de varejo de artigos e armarinhos, que enfrentam desafios semelhantes relacionados à categorização de despesas e à precisão dos registros financeiros. A análise dos dados dessa empresa pode oferecer insights valiosos que sejam aplicáveis a um contexto mais amplo.

Diagnóstico da situação problema e oportunidades identificadas

A categorização manual de despesas em planilhas financeiras é uma tarefa suscetível a diversos tipos de erros humanos, especialmente em empresas de médio porte que lidam com grandes volumes de transações ao longo do mês. Na planilha de fluxo de caixa da empresa analisada (conforme apresentado na Tabela 3), foi identificado que as categorias de despesas frequentemente são duplicadas, incorretamente categorizadas ou até mesmo deixadas em branco. Esses problemas resultam em análises financeiras imprecisas e ineficientes.

A primeira fonte de erro identificada refere-se à duplicação de categorias, em que uma mesma despesa é registrada sob diferentes códigos ou nomenclaturas. Por exemplo, gastos relacionados a anúncios em redes sociais podem ser classificados tanto como "Publicidade e Propaganda" quanto como "Marketing". Embora essas categorias aparentem ser distintas, na prática, referem-se ao mesmo tipo de despesa. A duplicidade gera inconsistências que

comprometem os relatórios financeiros, dificultando a identificação de áreas de custo e a previsão precisa do fluxo de caixa.

Tabela 3

Amostra da Planilha de Despesas

Data	Descrição	Categoria	Valor
17/03/2024	Produção de material publicitário	Materiais	R\$ 3.779,60
09/03/2024	Campanha do Google Ads	Marketing	R\$ 382,28
21/04/2024	Contabilidade externa		R\$ 3.562,91
10/05/2024	Contabilidade externa	Serviços Profissionais	R\$ 3.077,24
05/03/2024	Anúncios em redes sociais		R\$ 773,62
28/04/2024	Serviços legais	Serviços Profissionais	R\$ 4.071,33
22/04/2024	Serviços legais	Serviços Profissionais	R\$ 1.939,90
29/05/2024	Adiantamento quinzenal	Salários	R\$ 2.411,00
27/04/2024	Folha de pagamento mensal	Adiant. Salari.	R\$ 4.361,36
03/05/2024	Compra de materiais	Estoque	R\$ 1.197,31
03/05/2024	Reparo de mobiliário	Financeiro	R\$ 4.576,42
04/04/2024	Campanha do Google Ads	Marketing	R\$ 4.597,69
16/03/2024	Serviços de limpeza	Manutenção	R\$ 361,01
11/04/2024	Adiantamento quinzenal	Logística	R\$ 927,26
29/04/2024	Telefone e internet	Internet	R\$ 4.870,52
29/05/2024	Telefone e internet	Utilidades	R\$ 2.894,07
27/04/2024	Folha de pagamento mensal	Salários	R\$ 1.791,56
30/01/2024	Serviços de limpeza	Manutenção	R\$ 1.552,05
18/04/2024	Folha de pagamento mensal	Salários	R\$ 4.650,99
26/01/2024	Upgrade de equipamentos de TI	Tecnologia	R\$ 2.949,25
27/02/2024	Pagamento da conta de luz	Utilidades	R\$ 4.154,07
03/02/2024	Anúncios em redes sociais	Publicidade e Propaganda	R\$ 1.202,70
19/04/2024	Serviços legais	Mão-de-obra	R\$ 1.426,90
08/05/2024	Compra de materiais	Estoque	R\$ 3.558,44

Fonte: Elaborado pelos autores.

Outro problema recorrente identificado é a classificação incorreta de despesas, frequentemente resultante da falta de padronização nos critérios utilizados pelos funcionários (Tabela 3). Como o processo de categorização depende do julgamento humano, diferentes funcionários podem classificar transações semelhantes de maneiras distintas. A ausência de um guia claro para categorização, aliada à falta de um sistema de validação automática, contribui para a introdução de erros que, ao longo do tempo, tornam-se acumulativos. Por exemplo, erros de digitação, como registrar "Adiantamento Salarial" como "adiant. salari.", dificultam a categorização correta, uma vez que pequenas variações podem levar à criação de novas categorias desnecessárias.

Esses erros são exacerbados em planilhas de grande escala, onde o volume de transações mensais pode ultrapassar milhares de registros. Revisar manualmente essas planilhas para corrigir erros é um processo extremamente demorado, o que aumenta os custos operacionais e

retarda a tomada de decisões gerenciais. Conforme descrito por Sandhof e Filgueiras (2015), o erro humano em processos manuais de sistemas críticos é amplamente influenciado por fatores como cansaço, distração e pressão por prazos. Esses fatores estão diretamente relacionados ao contexto da categorização de despesas, onde erros de digitação ou julgamentos inconsistentes podem comprometer a precisão de toda a análise financeira da empresa.

Planejamento da ação

Com o objetivo de mitigar esses problemas, a aplicação de técnicas de *machine learning* para automatizar a categorização de despesas apresenta-se como uma solução eficaz. Por meio da utilização de algoritmos como *Random Forest* (RF) e Regressão Logística (RL), o sistema pode aprender a partir de dados históricos e aplicar regras consistentes para classificar novas despesas, eliminando a duplicidade de categorias e padronizando o processo de entrada de dados. Além disso, esses algoritmos são capazes de detectar erros de classificação e corrigir automaticamente categorizações incoerentes, reduzindo a necessidade de intervenção manual e tornando o processo mais eficiente e menos suscetível a erros.

Na prática, a implementação desses modelos contribui para a maior eficiência no processo de categorização de dados, reduzindo a dependência de entradas manuais e minimizando os erros associados. Aprendendo com os dados históricos, os modelos de *machine learning* não apenas classificam despesas com alta precisão, mas também identificam possíveis erros de classificação, fornecendo um mecanismo de correção contínua que melhora a confiabilidade dos relatórios financeiros (Murphy, 2012).

Coleta e pré-processamento de dados

Nesta etapa da intervenção, os dados foram coletados diretamente das planilhas de fluxo de caixa da empresa, que continham informações detalhadas sobre as entradas e saídas de caixa, categorizadas manualmente pelos funcionários. A coleta de dados envolveu a digitalização das planilhas físicas e a conversão dos registros para formato eletrônico (Tabela 3), possibilitando uma manipulação e análise mais eficiente (Géron, 2019). Adicionalmente, foi realizada uma verificação cruzada das informações com outras fontes internas da empresa, como relatórios financeiros mensais e registros de auditoria. Essa triangulação de fontes garantiu maior precisão e integridade aos dados coletados.

O pré-processamento dos dados representou uma etapa crítica, destinada a preparar os dados brutos para análise com *machine learning*. Esse processo incluiu diversas fases, conforme

descrito por Pedregosa et al. (2011), Murphy (2012), Géron (2019) e Jurafsky & Martin (2021): limpeza dos dados, normalização, remoção de stopwords (palavras comuns que não agregam valor significativo à análise, como artigos, preposições e conjunções) e vetorização de texto. A vetorização é o processo de transformar textos em representações numéricas que podem ser interpretadas pelos algoritmos de machine learning.

Implementação prática e ferramentas utilizadas

A implementação dos modelos de *machine learning* foi realizada utilizando bibliotecas amplamente reconhecidas no ecossistema Python, como Scikit-learn, Pandas e NumPy. A escolha dessas ferramentas justifica-se pela sua robustez, facilidade de uso, ampla documentação e suporte da comunidade. A biblioteca **Scikit-learn** foi utilizada para implementar os algoritmos de *machine learning*, incluindo Regressão Logística (RL) e *Random Forest* (RF), além de fornecer ferramentas para validação cruzada e avaliação dos modelos. Essa biblioteca oferece uma ampla gama de algoritmos de aprendizado supervisionado e não supervisionado, bem como métodos avançados para pré-processamento de dados e avaliação de desempenho dos modelos (Pedregosa et al., 2011).

A biblioteca **Pandas** foi empregada para manipulação e análise dos dados, facilitando o pré-processamento e a preparação dos conjuntos de dados para o treinamento dos modelos. Por meio de suas estruturas de dados de alto desempenho, ela permitiu trabalhar de forma eficiente com dados tabulares, sendo essencial para o processamento inicial dos dados coletados (Géron, 2019). Além disso, a biblioteca **NumPy** foi utilizada para operações matemáticas e manipulação de *arrays*, desempenhando um papel fundamental na vetorização dos dados textuais. Sua eficiência no processamento de grandes volumes de dados numéricos tornou-a uma ferramenta indispensável, especialmente em conjunto com Pandas e Scikit-learn (Harris et al., 2020).

Treinamento e avaliação dos modelos

Os algoritmos de RL e RF foram treinados utilizando um conjunto de dados rotulados. O modelo de RL foi treinado com base nos registros de despesas previamente associados a categorias específicas. Os coeficientes do modelo foram estimados pelo método de máxima verossimilhança, e a função logística foi aplicada para calcular a probabilidade de cada despesa pertencer a uma determinada categoria (Menard, 2002). Durante esse processo, ambos os modelos ajustaram seus parâmetros com o objetivo de minimizar os erros, aprendendo a partir dos exemplos rotulados fornecidos (Hosmer et al., 2013).

Os dados foram divididos em dois subconjuntos: 80% para treinamento e 20% para teste. Essa proporção foi escolhida para garantir que os modelos tivessem acesso a uma quantidade suficiente de dados para aprender padrões significativos, enquanto ainda preservava um conjunto de dados separado para avaliar o desempenho dos modelos de forma independente (Géron, 2019). A divisão dos dados foi realizada de forma aleatória para evitar viés na seleção dos registros destinados aos conjuntos de treinamento e teste. Os modelos foram avaliados com base em diversas métricas de desempenho, incluindo acurácia, *recall*, F1-score, matriz de confusão e a área sob a curva ROC (AUC-ROC) (Murphy, 2012). Essas métricas fornecem uma avaliação abrangente da capacidade dos modelos em categorizar corretamente as despesas, permitindo uma análise detalhada de seu desempenho em diferentes aspectos.

Resultados

Desempenho dos modelos de machine learning

Nesta seção, são apresentados os resultados obtidos a partir da aplicação dos modelos de Regressão Logística (RL) e *Random Forest* (RF) na categorização de despesas. A avaliação do desempenho dos modelos foi realizada com base em métricas amplamente reconhecidas, como acurácia, *recall*, F1-score, matriz de confusão e a área sob a curva ROC (AUC-ROC).

Random Forest (RF)

A técnica *Random Forest* (RF) foi empregada para modelar a probabilidade de uma despesa ser atribuída a uma categoria específica. Esse modelo (método *ensemble* baseado em múltiplas árvores de decisão), demonstrou robustez na categorização de dados financeiros. Os resultados obtidos com o modelo de RF são apresentados na Tabela 4.

Tabela 4*Matriz de Confusão e Métricas de Desempenho da RF*

Real / Previsto	Classe 0	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6
Classe 0	13	0	0	0	0	0	0
Classe 1	0	19	0	0	0	0	0
Classe 2	0	0	10	0	0	0	0
Classe 3	0	0	0	12	0	0	0
Classe 4	0	0	0	0	9	0	0
Classe 5	0	0	0	0	0	14	0
Classe 6	0	0	0	0	0	0	17
Métrica				Valor			
Acurácia				0.94			
Recall				0.94			
F1-Score				0.937			
AUC-ROC				0.994			

Fonte: Elaborado pelos autores.

A matriz de confusão apresentada evidencia o desempenho robusto do modelo *Random Forest* (RF), com a maioria das despesas corretamente categorizadas. O modelo RF demonstrou ser mais eficaz para a categorização de despesas, alcançando uma acurácia de 94% e a capacidade de integrar dados textuais e numéricos. Essa robustez foi essencial para capturar interações complexas entre variáveis, eliminar duplicidades e corrigir inconsistências.

Em comparação com a Regressão Logística (RL), o modelo RF apresentou um desempenho ligeiramente superior em todas as métricas avaliadas, especialmente na capacidade de generalizar em cenários. A combinação de características preditivas e a habilidade em lidar com dados heterogêneos posiciona o RF como uma solução ideal para organizações que buscam ferramentas escaláveis e adaptáveis para problemas de categorização financeira. As métricas obtidas reforçam a eficácia do modelo: acurácia de 94%, *recall* de 94% e F1-score de 0,937, indicando que o modelo é eficiente na categorização correta, minimizando tanto falsos positivos quanto falsos negativos. A AUC-ROC de 0,994 reflete uma excelente capacidade discriminativa do modelo entre as diferentes classes, como descrito por Hosmer et al. (2013).

Regressão RL

A Regressão Logística (RL) foi aplicada como uma abordagem alternativa para a categorização de despesas. Os resultados obtidos são apresentados na Tabela 5, fornecendo uma base comparativa para avaliar o desempenho relativo em relação ao modelo RF.

Tabela 5*Matriz de Confusão e Métricas de Desempenho da RL*

Real / Previsto	Classe 0	Classe 1	Classe 2	Classe 3	Classe 4	Classe 5	Classe 6
Classe 0	8	0	0	0	0	0	0
Classe 1	0	13	0	0	7	0	0
Classe 2	0	0	13	0	0	0	0
Classe 3	0	0	0	22	0	0	0
Classe 4	0	0	0	0	15	0	0
Classe 5	0	0	0	0	0	10	0
Classe 6	0	0	0	0	0	0	12
Métrica				Valor			
Acurácia				0.93			
Recall				0.93			
F1-Score				0.929			
AUC-ROC				0.995			

Fonte: Elaborado pelos autores.

A matriz de confusão do modelo de Regressão Logística (RL) evidencia um desempenho sólido, embora com uma leve redução em acurácia e *recall* em comparação ao modelo *Random Forest* (RF). Com acurácia de 93%, *recall* de 93% e F1-score de 0,929, o modelo RL demonstra uma alta capacidade de categorização correta das despesas. Além disso, a AUC-ROC de 0,995 ressalta a eficácia da RL na discriminação entre classes, alinhando-se ao desempenho esperado, conforme descrito por Breiman (2001).

Avaliação dos resultados

Nesta etapa de avaliação dos resultados da intervenção proposta, Tripp (2005) aponta que frequentemente são gerados dados tanto sobre os efeitos das mudanças implementadas durante o processo quanto sobre seu impacto após a conclusão da implementação. Considerando essa perspectiva, nesta seção são apresentados os resultados obtidos com os modelos de RL e RF. A comparação entre ambos é essencial para identificar a abordagem mais eficaz para a categorização de despesas, oferecendo insights sobre suas respectivas vantagens e limitações. Ao final desta seção, também é realizada uma avaliação do processo de implementação, a fim de analisar as implicações práticas e as lições aprendidas. A Tabela 6 resume os principais resultados de desempenho dos dois modelos, destacando as métricas avaliadas e os respectivos desempenhos em relação às demandas do contexto analisado.

Tabela 6*Comparação dos Modelos de Machine Learning*

Métrica	<i>Random Forest (RF)</i>	Regressão Logística (RL)
Acurácia	0.94	0.93
Recall	0.94	0.93
F1-score	0.937	0.929
AUC-ROC	0.994	0.995

Fonte: Elaborado pelos autores.

Os resultados mostram que ambos os modelos apresentaram desempenhos satisfatórios na categorização de despesas. A técnica *Random Forest* (RF) teve uma ligeira vantagem em termos de acurácia, *recall* e F1-score, enquanto a Regressão Logística (RL) apresentou uma AUC-ROC marginalmente maior. Conforme observado por Hosmer et al. (2013), a RL é eficaz em problemas de classificação binária, sendo valorizada por sua simplicidade e interpretabilidade. A acurácia e o *recall* de 93% indicam que a RL é altamente eficiente em prever corretamente as categorias de despesas. Além disso, a alta AUC-ROC de 0,995 reflete uma excelente capacidade de discriminação entre as diferentes categorias, evidenciando a robustez do modelo em cenários de classificação binária.

Por outro lado, Breiman (2001) destaca que a RF é particularmente eficaz em tarefas de classificação complexas, devido à sua capacidade de lidar com interações não lineares entre variáveis. A acurácia e o *recall* de 94%, combinados com um F1-score de 0,937, indicam que a RF é ligeiramente mais precisa que a RL, apresentando um desempenho altamente robusto. A AUC-ROC de 0,994 também confirma a eficácia da RF na discriminação entre classes.

Esses resultados corroboram os achados de outros estudos, como os de Aghware et al. (2024), que demonstraram a eficácia da RF em contextos com alta variabilidade e complexidade de dados. A capacidade da RF de reduzir a variância do modelo, ao combinar múltiplas árvores de decisão, contribui para o desempenho robusto observado, refletido na alta AUC-ROC obtida.

Durante a fase de testes do programa desenvolvido, observou-se que o modelo de *machine learning* apresentou alta acurácia ao classificar as despesas com base na base de dados original utilizada no treinamento. O algoritmo RF demonstrou uma excelente capacidade de aprendizado dos padrões e regularidades presentes nos dados históricos da empresa. No entanto, alguns desafios emergiram quando ocorreram mudanças nos padrões de descrição das despesas.

Embora o modelo tenha sido eficaz na categorização dos dados que seguiam o padrão previamente aprendido, a RF encontrou dificuldades em generalizar corretamente quando as descrições das despesas apresentavam variações ou terminologias não contempladas no conjunto de dados de treinamento. Por exemplo, se um termo utilizado na descrição de uma

despesa não estivesse previamente rotulado, o modelo tinha maior probabilidade de classificá-lo incorretamente ou de não identificar a categoria adequada. Esse comportamento reflete a dependência dos algoritmos de *machine learning* em relação aos padrões de dados fornecidos durante o treinamento. Quando esses padrões mudam significativamente, sem exemplos similares no treinamento, o modelo perde parte de sua capacidade de realizar previsões precisas.

Esse problema é amplamente conhecido como *overfitting* ao conjunto de dados de treinamento. O *overfitting* ocorre quando um modelo aprende excessivamente os detalhes e particularidades dos dados de treinamento, o que resulta em uma baixa capacidade de generalização para novos dados. Conforme descrito por Ying (2019), esse fenômeno acontece quando o modelo se ajusta muito bem aos dados de treinamento, incluindo os ruídos e *outliers*, tornando-se menos eficaz ao lidar com dados novos que apresentam variações sutis. Métodos como regularização, poda de árvores e estratégias como *early stopping* podem ser aplicados para mitigar o *overfitting*, controlando a complexidade do modelo e melhorando sua capacidade de generalização. Um exemplo prático é a variação na nomenclatura de despesas, como "Serviço de limpeza" e "Limpeza de escritórios". Se o modelo foi treinado apenas com o primeiro termo, ele pode não identificar corretamente o segundo, resultando em uma classificação incorreta.

Nesta etapa da implementação, foi observada uma redução significativa nos erros manuais ao comparar as categorias sugeridas pelo modelo com aquelas originalmente atribuídas pelos funcionários. O processo foi iterativo, com ajustes a partir do feedback coletado durante as interações práticas. Essa abordagem assegurou que a solução não apenas corrigisse as inconsistências iniciais, mas também fosse adaptada às particularidades da empresa.

Essa melhoria impactou a consistência e a precisão dos registros financeiros, reduzindo a necessidade de revisões manuais demoradas. A otimização do tempo dedicado à categorização permitiu que os funcionários direcionassem seus esforços para atividades estratégicas de maior valor agregado, gerando um impacto positivo no desempenho geral da organização.

Considerações finais

O presente estudo teve como objetivo principal a aplicação de uma ferramenta desenvolvida com base em técnicas de *machine learning* para solucionar o problema de categorização incorreta de despesas no fluxo de caixa de uma empresa familiar alagoana do setor de varejo de artigos e armarinhos. O método adotado foi a pesquisa-ação, que, no ambiente organizacional, tem como objetivo frequente resolver problemas de ordem técnica.

O foco esteve na identificação do modelo mais eficiente entre a Regressão Logística (RL), amplamente reconhecida por sua eficácia em classificação binária e interpretação de resultados, e o *Random Forest* (RF), aplicado para corrigir a categorização de despesas, anteriormente realizada manualmente e marcada por inconsistências significativas. Após a implementação e avaliação dos modelos, o RF foi escolhido como a abordagem mais eficiente, apresentando um desempenho ligeiramente superior em termos de acurácia, *recall* e F1-score. O RF alcançou uma acurácia de 94%, enquanto a RL apresentou 93%. Embora a diferença seja pequena, a robustez do RF e sua capacidade de lidar com interações complexas entre variáveis destacaram-no como a escolha ideal para este estudo. Apesar de a literatura sobre a adoção de IA em empresas familiares apontar barreiras significativas, como resistência cultural e percepções da IA como uma ameaça (Upadhyay et al., 2023), sua implementação é crucial para a sobrevivência e o sucesso dessas organizações em um cenário de mercado cada vez mais competitivo (Lannon et al., 2023; Ulrich et al., 2023). A adoção do modelo RF trouxe diversos benefícios para a empresa, que podem ser resumidos em três principais aspectos.

Primeiramente, a precisão na categorização de despesas foi significativamente aprimorada, minimizando erros humanos e aumentando a confiabilidade dos registros financeiros. Isso facilitou a análise financeira e proporcionou uma visão mais clara das finanças da empresa. Em segundo lugar, a implementação do sistema automatizado reduziu substancialmente o tempo despendido pelos funcionários na tarefa manual de categorização, aumentando a eficiência e permitindo que os colaboradores se concentrassem em atividades mais estratégicas e de maior valor agregado para a organização. Por fim, o sistema garantiu consistência na categorização das despesas, eliminando variações ocasionadas por interpretações subjetivas dos funcionários. A consistência nos registros financeiros é essencial para a tomada de decisões precisas e bem fundamentadas.

Além disso, a maior precisão na categorização e a implementação do processo automatizado resultaram em uma redução de custos operacionais. A empresa foi capaz de identificar e eliminar despesas redundantes ou desnecessárias, otimizando seu orçamento e aumentando a lucratividade. A flexibilidade do modelo permite futuras expansões e ajustes, sem a necessidade de reformular completamente o sistema. Com registros financeiros mais precisos e consistentes, a gestão da empresa passou a tomar decisões mais informadas e estratégicas, incluindo o planejamento de fluxo de caixa, a identificação de áreas com custos excessivos e a otimização dos recursos financeiros.

A aplicação de técnicas de *machine learning*, com destaque para o modelo RF, demonstrou ser uma solução eficaz para os desafios de categorização de despesas enfrentados

pela empresa. Os benefícios obtidos transcendem a precisão na categorização, abrangendo melhorias significativas na eficiência operacional, na consistência dos registros e na qualidade da tomada de decisões. Este estudo destaca a importância da adoção de tecnologias avançadas, como o *machine learning*, em processos empresariais tradicionais. A implementação dessas tecnologias não apenas resolve problemas imediatos, mas também estabelece uma base sólida para o crescimento e a inovação contínuos. Espera-se que os insights apresentados contribuam para a evolução da gestão financeira em empresas com desafios semelhantes.

Como limitação, destaca-se que a acurácia do modelo foi impactada por descrições menos padronizadas, como descrições vagas ou excessivamente específicas, que dificultam a categorização automática. Esse problema é relevante em empresas onde funcionários utilizam terminologias variadas para descrever as mesmas despesas. Para mitigar essa limitação, futuros aprimoramentos no modelo podem incluir o uso de técnicas mais avançadas de Processamento de Linguagem Natural (NLP), como *word embeddings*. Segundo Chiu et al. (2016), o uso de *word embeddings* permite capturar relações semânticas entre palavras, facilitando o entendimento de descrições semanticamente semelhantes, ainda que textualmente diferentes.

Por fim, o objetivo principal da pesquisa-ação não é a generalização, dado que, conforme Macke (2006, p. 228), as intervenções nesse método “são experiências únicas por conta dos fatores inerentes ao contexto vivido no momento do estudo”. No entanto, é possível alcançar um nível de generalização ao acumular experiências em diferentes pesquisas. Assim, para estudos futuros, sugere-se a realização de novas investigações utilizando o método de pesquisa-ação, testando diferentes algoritmos e expandindo a base de treinamento. A inclusão de mais exemplos de descrições variadas pode aumentar a capacidade de generalização do modelo, contribuindo para uma categorização ainda mais precisa e eficiente.

Referências

- Aghware, F., Ojugo, A. A., Adigwe, W., Odiakaose, C. C., Ojei, E. O., Ashioba, N. C., Okpor, M. D., & Geteloma, V. O. (2024). Enhancing the Random Forest Model via Synthetic Minority Oversampling Technique for Credit-Card Fraud Detection. *Journal of Computing Theories and Applications*. <https://doi.org/10.62411/jcta.10323>
- Altieri, M. P., Apro, P. H., & Basile, A. L. (2021). O uso de machine learning na classificação de textos com ênfase em fake news. *SIBGRAPI*. Disponível em <https://adelpha->

api.mackenzie.br/server/api/core/bitstreams/3bb2fb93-810c-494d-93af-69c90961f0b8/content

Avelar, E. A., Leocádio, V. A., Campos, O. V., Ferreira, P. O., & Orefici, J. B. P. (2022).

Algoritmo *random forest* para previsão de comportamento de preços de ativos. *Revista FSA 19*(10). Disponível em

<http://www4.unifsa.com.br/revista/index.php/fsa/article/view/2592>

Barbier, R. (2007). *A pesquisa-ação*. Brasília: Liber Livros Editora.

Bondielli, A., & Marcelloni, F. (2019). A survey on fake news and rumour detection techniques. *Information Sciences*, 497, 38–55.

<https://doi.org/10.1016/j.ins.2019.05.035>

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

<https://doi.org/10.1023/A:1010933404324>

Chiu, B., Crichton, G., Korhonen, A., & Pyysalo, S. (2016). How to train good word embeddings for biomedical NLP. *Proceedings of the 15th Workshop on Biomedical Natural Language Processing*, 166–174. <https://doi.org/10.18653/v1/W16-2922>

Chopra, A., Prashar, A., & Sain, C. (2013). Natural language processing. *International Journal of Technology Enhancements and Emerging Engineering Research*, 1(4), 131–134. <https://doi.org/10.1.1.407.6907&rep=rep1&type=pdf>

Coelho, F. F., de Lima Amorim, D. P., & de Camargos, M. A. (2021). Analisando métodos de *machine learning* e avaliação do risco de crédito. *Revista Gestão & Tecnologia*, 21(1), 89–116. Disponível em <https://revistagt.fpl.emnuvens.com.br/get/article/view/2089>

Dumitrescu, E., et al. (2022). Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *European Journal of Operational Research*, 297(3), 1178–1192. <https://doi.org/10.1016/j.ejor.2021.06.053>

- Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (3ª ed.). O'Reilly Media. Disponível em https://anayamultimedia.es/primer_capitulo/aprende-machine-learning-con-scikit-learn-keras-y-tensorflow-tercera-edicion.pdf
- Gilani, S. A., Copiaco, A., Gernal, L., Yasin, N., Nair, G., & Anwar, I. (2023). Savior or distraction for survival: Examining the applicability of machine learning for rural family farms in the United Arab Emirates. *Sustainability*, 15(4), 3720. <https://doi.org/10.3390/su15043720>
- Gitman, L. J., & Zutter, C. J. (2017). *Princípios de administração financeira*. São Paulo: Editora Pearson.
- Halteh, K., Kumar, K., & Gepp, A. (2018). Financial distress prediction of Islamic banks using tree-based stochastic techniques. *Managerial Finance*, 44(6), 759–773. <https://doi.org/10.1108/MF-12-2016-0372>
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., & Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Harris, M., Opp, C. & Opp, M. (2020). The Aggregate Demand for Bank Capital (September 2020). NBER Working Paper No. w27858, Available at SSRN: <https://ssrn.com/abstract=3700686>
- Hendriksen, E. S., & Breda, M. F. V. (2014). *Teoria da contabilidade* (5ª ed.). São Paulo: Atlas.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3ª ed.). John Wiley & Sons. Disponível em <https://onlinelibrary.wiley.com/doi/epub/10.1002/9781118548387>

- Jurafsky, D., & Martin, J. H. (2021). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. (3^a ed.). Prentice Hall.
- Koralun-Bereźnicka, J. (2013). The relative importance of corporate working capital determinants: Evidence from Polish listed companies. *Contemporary Economics*, 7(3), 3–24. <https://doi.org/10.5709/ce.1897-9254.87>
- Kumar, M., Charles, V., & Mishra, C. S. (2016). Evaluating the performance of the Indian banking sector using DEA during post-reform and global financial crisis. *Journal of Business Economics and Management*, 17(1), 156–172.
<https://doi.org/10.3846/16111699.2013.809785>
- Kumar, D., & Ratten, V. (2024). Artificial intelligence and family businesses: A systematic literature review. *Journal of Family Business Management*.
<https://doi.org/10.1108/JFBM-08-2024-0160>
- Lannon, F., Lyons, R., & O'Connor, C. (2023). Generation AI and family business: A perspective article. *Journal of Family Business Management*, 14(3), 470–474.
<https://doi.org/10.1108/JFBM-07-2023-0116>
- Lima, M. S. M., & Delen, D. (2020). Predicting and explaining corruption across countries: A machine learning approach. *Government Information Quarterly*, 37(3).
<https://doi.org/10.1016/j.giq.2019.101407>
- Macke, J. (2006). A pesquisa-ação como estratégia de pesquisa participativa. In C. K. Godoi, R. Bandeira-de-Mello, & A. B. Silva (Eds.), *Pesquisa qualitativa em estudos organizacionais: Paradigmas, estratégias e métodos*. São Paulo: Saraiva.
- Maia, N. N., & Salton, G. D. (2022). Utilização de machine learning para classificação de sentimentos no idioma Português-Brasil. *Brazilian Journal of Development*, 8(6), 43568–43580. DOI: <https://doi.org/10.34117/bjdv8n6-070>

- McCloskey, B. J., LaCasse, P. M., & Cox, B. A. (2024). Natural language processing analysis of online reviews for small business: Extracting insight from small corporations. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-023-05816-2>
- Menard, S. (2002). *Applied logistic regression analysis* (2^a ed.). SAGE Publications.
- Montevechi, A. A., de Carvalho Miranda, R., Medeiros, A. L., & Montevechi, J. A. B. (2024). Advancing credit risk modelling with machine learning: A comprehensive review of the state-of-the-art. *Engineering Applications of Artificial Intelligence*, 137, 109082. <https://doi.org/10.1016/j.engappai.2024.109082>
- Moraes, G., & Oliveira, L. (2014). A gestão financeira e sua importância nas empresas. *Revista Brasileira de Administração Científica*, 5(2), 100–112. <https://sustenere.inf.br/index.php/rbadm/issue/view/59>
- Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.
- O'Brien, R. (2001). An overview of the methodological approach of action research. In R. Richardson (Ed.), *Theory and practice of action research*. João Pessoa: Universidade Federal da Paraíba. Disponível em <http://www.web.ca/~robrien/papers/arfinal.html>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>
- Petropoulos, A. V., Stavroulakis, E., & Vlachogiannakis, N. E. (2020). Predicting bank insolvencies using machine learning techniques. *International Journal of Forecasting*, 36(3), 1092–1113. <https://doi.org/10.1016/j.ijforecast.2019.11.005>
- Prati, R. C., Batista, G. E. A. P. A., & Monard, M. C. (2008). Evaluating Classifiers Using ROC Curves. Latin America Transactions, IEEE. *Revista IEEE América Latina*, 6(2), 215–222. <https://doi.org/10.1109/TLA.2008.4609920>

- Russell, S., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4ª ed.). Prentice Hall.
- Sandhof, K., & Filgueiras, L. V. L. (2015). Defeitos de software como erros humanos. *II Workshop Um Olhar Sociotécnico sobre a Engenharia de Software – WOSE*.
- Santos, H. G. D., Nascimento, C. F. D., Izbicki, R., Duarte, Y. A. D. O., Chiavegatto Filho, P., & Dias, A. (2019). Machine learning para análises preditivas em saúde: Exemplo de aplicação para prever óbito em idosos de São Paulo, Brasil. *Cadernos de Saúde Pública*, 35(7), e00050818.
- Silva e Navarro, J. V.M.& Valverde, A. R. (2023). A importância da gestão financeira nas micro e pequenas empresas: O caso de uma empresa de pequeno porte no setor alimentício. *Gestão – Revista Científica*, 5(2).
<https://doi.org/10.5281/zenodo.13772993>
- Thiollent, M. (2009). *Pesquisa-ação nas organizações* (2ª ed.). São Paulo: Atlas.
- Tripp, D. (2005). Pesquisa-ação: Uma introdução metodológica. *Educação e Pesquisa*, 31(3), 443–466.
- Ulrich, P., Frank, V., & Buettner, R. (2023). Artificial intelligence in small and medium-sized family firms: An empirical study on the impact of family influence. *Corporate Governance and Organizational Behavior Review*, 7(1), 72–80.
<https://doi.org/10.22495/cgobrv7i1p7>
- Upadhyay, N., Upadhyay, S., Al-Debei, M. M., Baabdullah, A. M., & Dwivedi, Y. K. (2023). The influence of digital entrepreneurship and entrepreneurial orientation on intention of family businesses to adopt artificial intelligence: Examining the mediating role of business innovativeness. *International Journal of Entrepreneurial Behavior and Research*, 29(1), 80–115. <https://doi.org/10.1108/IJEER-02-2022-0154>

Vergara, S. C. (2005). *Métodos de pesquisa em administração* (1ª ed.). São Paulo: Editora Atlas S.A.

Ying, X. (2019). An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, 1168(2), 022022.

Zöller, M., & Huber, M. F. (2021). Benchmark and survey of automated machine learning frameworks. *Journal of Artificial Intelligence Research*, 70, 409–472. Disponível em <https://arxiv.org/pdf/1904.12054>