

Área Temática: Tecnologia, Inteligência Artificial e Transformação Digital em Administração

O USO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA NA EFICIÊNCIA FINANCEIRA

RESUMO: A implementação de algoritmos de aprendizado de máquina está revolucionando a gestão financeira das empresas, permitindo a otimização de processos e a maximização da eficiência operacional. O propósito desta pesquisa é aplicar um algoritmo de aprendizado supervisionado para previsão de indicadores financeiros. Para alcançar esta meta, selecionou-se uma empresa no segmento de soluções em recursos humanos. Ao analisar o uso desta tecnologia nas operações financeiras da organização, este trabalho busca por meio do modelo utilizado, compreender a adoção de algoritmos de aprendizado de máquina no contexto atual. A metodologia adotada envolve a coleta de dados financeiros, e a aplicação do modelo de análise preditiva por *gradient boosting regression*.

PALAVRAS-CHAVE: Algoritmos de Aprendizado de Máquina; Inteligência Artificial; Eficiência Financeira; Análise e Planejamento Financeiro; Gestão Financeira.

ABSTRACT: The implementation of machine learning algorithms is revolutionizing the financial management of companies, enabling process optimization and maximizing operational efficiency. The purpose of this research is to apply a supervised learning algorithm to predict financial indicators. To achieve this goal, a company in the human resources solutions sector was selected. By analyzing the use of this technology in the organization's financial operations, this work aims, through the applied model, to understand the adoption of machine learning algorithms in the current context. The methodology involves the collection of financial data and the application of the predictive analysis model using gradient boosting regression.

KEYWORDS: Machine Learning Algorithms; Artificial Intelligence; Financial Efficiency; Financial Analysis and Planning; Financial Management.

1. INTRODUÇÃO

A precisão no planejamento e na previsão financeira é fundamental no atual cenário de mercado para garantir uma alocação de recursos eficiente e eficaz (SEBRAE, 2022). Como resultado, a maioria das empresas conta com equipes especializadas em Planejamento e Análise Financeira integradas à sua estrutura financeira. Com o advento de novas metodologias de análise, a ampliação do acesso ao big data e o acirramento da competição que exige decisões embasadas em dados, as equipes de Financial Planning and Analysis (FP&A) não são mais apenas responsáveis por relatórios, mas sim se tornaram parceiras estratégicas do negócio (JURAFSKY; MARTIN, 2022).

Na contemporaneidade, as técnicas avançadas de Analytics, como Inteligência Artificial, Aprendizado de Máquina (AM) e Processamento de Linguagem Natural, têm sido amplamente implementadas em Planejamento Financeiro. Elas surgiram como um conjunto de técnicas adaptadas a este contexto em evolução, utilizando o uso de dados, algoritmos estatísticos e técnicas analíticas avançadas para identificar a probabilidade de resultados futuros com base em dados históricos e outras variáveis micro/macro relevantes. Essas técnicas permitem que as equipes de FP&A tomem decisões informadas e baseadas em dados em um ambiente de negócio (GANGULI, 2020; KULKARNI, 2023).

Para Russell e Norvig (2020), em convergência com a visão de Surahyo (2023), ao aproveitar o poder da análise de dados, os profissionais de FP&A podem obter conhecimentos mais profundos sobre o desempenho empresarial, identificar os principais impulsionadores do sucesso financeiro e comunicar de forma eficaz aos stakeholders do negócio. Espera-se agora que os profissionais de FP&A forneçam recomendações práticas, apoiem a formulação de estratégias de negócios e impulsionem iniciativas de melhoria de desempenho. Esta mudança de responsabilidades elevou a importância da FP&A como parceiro estratégico da gestão sênior, permitindo o alinhamento das metas financeiras com os objetivos do negócio.

Entre as tecnologias que têm ganhado notoriedade na atualidade por seu uso em diferentes setores, destacam-se os modelos de aprendizado por agrupamento ou *ensemble learning* e a *gradient boosting regression* (RAO *et al.*, 2019). Os métodos de *ensemble learning*, como o *bagging* e o *boosting*, combinam múltiplos modelos base para melhorar a precisão e a robustez das previsões (REN *et al.*, 2016). No caso específico do *gradient boosting regression*, a técnica constrói modelos sequenciais, onde cada novo modelo tenta corrigir os erros dos anteriores, essa abordagem é especialmente eficaz em problemas de previsão e tem a capacidade de gerar previsões significativas (MOHAMMED; KORA, 2023). Nessa linha, ressalta-se que devido a flexibilidade e precisão, ambos tem sido amplamente aplicados em áreas como cadeia de suprimentos, indústrias, diagnóstico médico, finanças e seguros destacando-se como um conjunto de técnicas importantes para a análise de dados complexos (LIN; HUANG, 2022).

De acordo com o estudo da Organização para a Cooperação e o Desenvolvimento Econômico (OCDE, 2021), a adoção de inteligência artificial (IA) é impulsionada pela grande e crescente disponibilidade de dados e pela vantagem competitiva esperada que a IA/AM pode fornecer às empresas. O crescimento no volume de dados e análises, aliada à capacidade de computação mais acessível, como pelos recursos de computação em nuvem, pode ser analisada por modelos de AM para identificar sinais e capturar relacionamentos subjacentes nos dados de uma maneira que está além da capacidade dos humanos. A implantação de IA/AM por

empresas nas áreas financeiras espera-se que impulse cada vez mais a vantagem competitiva das empresas. Além disso, a pesquisa anual FP&A Trends Survey 2023 demonstra uma projeção de que 54% das companhias planejam utilizar análises preditivas e de aprendizado de máquina no próximo ano em suas atividades de gestão financeira.

Este artigo apresenta o referencial teórico no próximo capítulo, seguido pelos procedimentos metodológicos no terceiro capítulo. Em sequência, analise-se os resultados os resultados obtidos no quarto capítulo e as considerações finais no último capítulo. O presente trabalho pretende aplicar métodos de aprendizado de máquina para predição de indicadores financeiros, e compreender a adoção de algoritmos de aprendizado de máquina no contexto atual.

2. REFERENCIAL TEÓRICO

2.1 Algoritmos de Aprendizado de Máquina

Como campo de estudo, a aprendizagem de máquina pode ser conceituada de forma ampla como um campo interdisciplinar que reúne características de ciência da computação, estatística e uma variedade de outras disciplinas, e se dedica ao desenvolvimento, análise e aplicação de técnicas para identificação automática de padrões em conjuntos de dados (JORDAN; MITCHELL, 2015).

O aprendizado de máquina depende de diferentes algoritmos para resolver problemas de dados, sendo enquadrados em sua maioria entre os tipos de aprendizado supervisionado ou não supervisionado. Sendo o aprendizado não supervisionado caracterizado por aprender características dos dados, conforme novos dados são introduzidos, ele usa as características previamente aprendidas para reconhecer a classe dos dados (MAHESH, 2020).

Segundo LeCun, Bengio e Hinton (2015), a forma mais comum de aprendizado de máquina é o aprendizado supervisionado, em que o algoritmo é treinado em um conjunto de dados que inclui pares de entrada e saída esperados. O objetivo é aprender uma função que possa prever a saída correspondente para novas entradas não vistas. Exemplos comuns de modelos supervisionados incluem regressão linear, regressão logística, árvores de decisão, entre outros.

Historicamente, um dos primeiros usos da regressão era o de verificar e desvendar o relacionamento linear entre as variáveis preditoras e uma variável resultante. A regressão linear simples descreve a relação entre uma variável e outra, mostrando como a magnitude de uma variável está relacionada à magnitude de uma segunda variável (BRUCE, P.; BRUCE, A., 2019).

Na atualidade, alguns métodos como o *Gradient Boosting Regression* apresentado por Belyadi e Haghghat (2021), caracterizado por ser um modelo de aprendizado por agrupamento, cuja técnica combina múltiplos modelos de aprendizado de máquina para melhorar a precisão na previsão dos valores e no desempenho das métricas de qualidade.

A principal vantagem dos modelos de aprendizado por agrupamento que utilizam o *boosting* é que ele reduz tanto a variância quanto o viés nas previsões, esta técnica é capaz de detectar padrões complexos em dados financeiros que podem ser ignorados por modelos mais simples, tendo em vista que os modelos subsequentes focam nos erros cometidos pelos modelos anteriores (LÓPEZ DE PRADO, 2018).

Para Joseph (2022), a divisão de dados é uma abordagem comumente utilizada para validação de modelos, onde dividimos um conjunto de dados dado em dois

conjuntos distintos: treinamento e teste. Os modelos estatísticos e de aprendizado de máquina são então ajustados no conjunto de treinamento e validados usando o conjunto de teste.

Bodner, Fortin e Molnár (2020) destacam que três estágios principais delineiam o processo de modelagem preditiva: 1) Formulação da Questão, 2) Construção e Teste do Modelo e 3) Avaliação da Incerteza. A Formulação da Questão é a etapa em que o contexto e o escopo da questão são formados, na etapa de Construção e Teste do Modelo os tipos de modelo são escolhidos, formas funcionais e variáveis são selecionadas, parâmetros são estimados e modelos são avaliados e na Avaliação da Incerteza a incerteza é identificada, considerada e reduzida quando viável.

Segundo Seber e Lee (2012), os modelos estatísticos objetivam compreender a relação entre a variável resposta ou variável prevista, e sua dependência em relação ao conjunto de variáveis explicativas, e para isso se faz necessário durante o desenvolvimento do modelo e criação dos algoritmos ajustes para estabelecer além da previsão e da explicação, a possibilidade de testar hipóteses e se aprofundar no relacionamento entre as variáveis utilizadas.

2.2 O Uso de Aprendizado de Máquina na Eficiência Financeira

De acordo com o relatório *Future of Finance 2023* elaborado pela PwC, as empresas dependem cada vez mais das equipes de análise e planejamento financeiro para modelar cenários de negócios complexos, oferecer previsões financeiras alinhadas às operações e tomar decisões inteligentes em tempo real. O uso de inteligência artificial e aprendizado de máquina é essencial para otimizar as previsões, eliminar processos manuais e melhorar a precisão das análises. Plataformas modernas de planejamento, integrando AI e AM, permitem análises preditivas, identificação de anomalias e comparações em tempo real, impulsionando o crescimento e a eficiência operacional das organizações.

A aplicação empírica de aprendizado de máquina na área financeira vem se consolidando no mundo, a multinacional americana Amazon desde 2020, conta com um modelo de aprendizado de máquina para previsão de receitas explorando o potencial de ferramentas como a linguagem de programação Python e a biblioteca de software XGBoost (JASIMUDDIN; ADAMS, 2020).

A linguagem de programação Python se caracteriza por seu uso geral, e se destaca como uma importante linguagem para computação científica. A codificação em uma linguagem de alto nível, a capacidade de interagir com uma ampla variedade de softwares e o seu vasto número de módulos de bibliotecas científicas, são características que popularizaram o seu uso para o desenvolvimento de algoritmos e análise exploratória de dados (OLIPHANT, 2007). Em consonância com essas ideias Rayhan e Gross (2023), complementam que apesar de ter propósito geral, ela está sendo cada vez mais usada em ambientes acadêmicos e dentro de diferentes campos presentes na indústria.

O Scikit-learn tem sido a principal biblioteca de aprendizado de máquina para programadores de linguagem Python, fornecendo uma extensa variedade de algoritmos e por sua interface de fácil uso e integrada que permite com que seja feita uma programação de análise de dados e criação de modelos produtiva (MCKINNEY, 2022). Para Pfister *et al.* (2021) essas características atendem à crescente necessidade de análise estatística de dados por não especialistas ao proporcionar facilidade na elaboração de algoritmos e modelos supervisionados.

Segundo Chen e Guestrin (2016), o XGBoost (*Extreme Gradient Boosting*) é um algoritmo de aprendizado de máquina de código aberto que foi desenvolvido principalmente por Tianqi Chen, amplamente utilizado em ciência de dados e com uma variedade de aplicações práticas em empresas, o XGBoost é uma implementação eficiente de algoritmos de boosting de árvore de decisão, projetada para ser altamente escalável. Sua capacidade de combinar árvores de decisão fracas torna-o ideal para modelos de regressão robustos e escaláveis. No contexto de regressão, o XGBoost oferece precisão e flexibilidade, suportando uma variedade de objetivos de aprendizado, sua velocidade e eficácia o tornam uma ferramenta essencial em aplicações práticas, desde previsão de séries temporais até análise de dados complexos.

Para Gökçe e Duman (2022), a técnica GridSearchCV, implementada pelo uso do scikit-learn, efetua uma busca abrangente para identificar os melhores hiperparâmetros de um modelo de aprendizado de máquina. Esse método aplica a validação cruzada para avaliar sistematicamente diversas combinações de parâmetros, assegurando resultados robustos ao automatizar a seleção de hiperparâmetros, a GridSearchCV economiza tempo e esforço manual e facilita a comparação organizada entre diferentes configurações, permitindo a determinação da configuração mais eficaz para o modelo. A adoção desta técnica busca minimizar para Bashir *et al.* (2020), que tenhamos risco de *overfitting* que consiste no ajuste excessivo do modelo de aprendizado de máquina criando previsões precisas para dados de treinamento e se tornando incapaz de fazer previsões precisas em dados não vistos anteriormente.

O estudo elaborado pelo OCDE (2021), reforça a importância da criação de modelos de aprendizado de máquina coerentes na área financeira que adotem métricas de qualidade, visando mitigar correlações tratadas como causalidade, desvios de modelos e outras inconsistências e riscos durante o processo de desenvolvimento, implementação e publicação dos resultados. Em conformidade, com a percepção do estudo Botchkarev (2018), fornece em sua pesquisa um amplo conjunto de técnicas para avaliar a qualidade de modelos de aprendizado de máquina, essas métricas de desempenho como o *Mean Absolute Error* (MAE), que define o erro absoluto médio dos erros entre valores observados e previsões, calculando a média das diferenças absolutas e o *Root Mean Square Error* (RMSE), se caracteriza por calcular a raiz quadrada da média dos quadrados dos erros, e o Coeficiente de determinação (R-Quadrado), que é definido por expressar a proporção de variabilidade explicada pela preditora, ambos os indicadores representam um modelo com menor discrepância entre os dados previstos e observados com valores menores de RMSE, MAE e R-Quadrado indicando um modelo preciso.

De acordo com Stobierski (2020), os indicadores de desempenho financeiro são métricas que as organizações usam para rastrear, medir e analisar a saúde financeira da empresa. Esses indicadores se enquadram em diversas categorias, incluindo lucratividade, liquidez, solvência, eficiência e avaliação.

Quadro 1: Definição dos indicadores de desempenho financeiro.

Indicadores de Desempenho Financeiro	Definição
<i>Gross Profit - GP</i>	Segundo Stobierski (2020), o lucro bruto é

	o valor que resta da receita após subtrair o custo direto dos produtos ou serviços vendidos. Esses custos variáveis incluem materiais e mão de obra diretamente envolvidos na produção.
<i>EBITA</i>	Na visão de Ross, Westerfield e Jordan (2013), o indicador representa o lucro operacional antes de juros, impostos e amortização, demonstrando a rentabilidade e eficiência da companhia.
<i>EBITA Margin %</i>	De acordo com o Karnov Group (2024), a medida representa o EBITA como porcentagem das vendas líquidas, e indica a rentabilidade ao longo do tempo para o negócio.
<i>Personal Expenses - PE</i>	Para o Statistics Finland (2024), a medida de despesas com pessoal compreendem salários e vencimentos sujeitos à retenção de imposto e despesas comparáveis, bem como despesas determinadas diretamente com base em salário ou vencimento.
<i>GP/PE</i>	Para o Swiss Steel Group (2023), a margem bruta sobre despesas de pessoal, é uma medida de desempenho financeiro que indica a produtividade e eficiência operacional de uma empresa em relação aos seus gastos com pessoal.

Fonte: Elaborado pelo autor.

Os indicadores de desempenho financeiro podem ser divididos em 10 categorias, para o estudo foram selecionados os indicadores das categorias de lucratividade, fluxo de caixa, qualidade de lucro e capacidade operacional com o objetivo de estabelecer previsões e análises dos dados obtidos da companhia adotada como objeto do estudo (Delen *et al.*, 2013; Hightower, 2008; Zhao *et al.* 2023).

3. PROCEDIMENTOS METODOLÓGICOS

O objetivo deste estudo consistiu em uma pesquisa aplicada de caráter qualitativo e quantitativo, em relação aos propósitos da pesquisa, esta se configura como exploratória e descritiva. No que diz respeito aos aspectos técnicos neste trabalho é adotado um modelo supervisionado com a abordagem da *gradient boosting regression* para predição de indicadores de desempenho financeiro de *Gross Profit*, *EBITA*, *EBITA Margin %*, *PE*, *GP/PE*.

Na visão de Flick (2009), os aspectos essenciais da pesquisa qualitativa consistem em sua natureza exploratória de descobrir novas percepções e gerar hipóteses a partir dos dados coletados. No que se refere a pesquisa quantitativa na visão de Creswell (2021), se caracteriza por relacionar e identificar as variáveis a serem utilizadas e empregar procedimentos estatísticos na análise dos dados.

De acordo com Gil (2017), as pesquisas descritivas têm como objetivo a descrição das características de determinada população ou fenômeno. Sendo elaboradas também com a finalidade de identificar possíveis relações entre variáveis. Enquanto a pesquisa exploratória enfatiza na investigação do problema ou estudo com o objetivo de proporcionar conhecimento e compreensão no desenvolvimento do assunto de estudo (YIN, 2017).

Nesse estudo, utiliza-se como unidade de análise os indicadores de desempenho financeiro históricos de uma empresa multinacional do setor de Recursos Humanos que se destaca como um dos principais competidores no setor de consultoria de recursos humanos globalmente.

A análise do estudo foi conduzida utilizando a implementação do modelo proposto por Belyadi e Haghghat (2021), e o conjunto de técnicas de métricas de desempenho propostas por Botchkarev (2018), com o objetivo de determinar a previsão dos indicadores de desempenho de *Gross Profit*, *EBITA*, *EBITA Margin %*, *PE*, *GP/PE* para o ano de 2023.

Para Kumar e Chong (2018), a análise de correlação é uma técnica que identifica relações entre variáveis nos dados, auxiliando na compreensão da importância dos atributos em relação à classe alvo. Neste estudo, foram investigadas técnicas de correlação e abordagens de aprendizado de máquina com o objetivo de identificar os atributos mais relevantes no conjunto de dados.

As previsões foram realizadas a partir do algoritmo de aprendizagem supervisionada denominado XGBoost, utilizando a técnica de *gradient boosting regression* e a linguagem de programação Python no ambiente de desenvolvimento integrado Google Colab. O período selecionado para análise histórica foi de Janeiro de 2018 até Agosto de 2023.

4. APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

No quadro 2, as medidas descritivas das variáveis utilizadas para criação do modelo de predição foram analisadas para iniciar o processo de análise exploratória dos dados que para Lopes *et al.* (2019), consiste no processo de descobrir informações ocultas e desconhecidas nos dados, visando possibilitar ao analista do estudo uma interpretação confiável de similaridades, diferenças, tendências, clusters e correlações.

Quadro 2: Média, Mediana, Máximo, Mínimo, Desvio Padrão, Intervalo Interquartil e Coeficiente de Variação das variáveis estudadas.

	Média	Desvio Padrão	Mínimo	Percentil 25
Gross Profit	26277.333333	13165.969462	13106.00	16173.5000
EBITA	8557.166667	6240.589921	2869.00	3777.5000
EBITA Margin %	29.481667	8.112172	21.80	22.4475
PE	16700.500000	7217.448240	8893.00	11309.2500
GP/PE	1.541667	0.124967	1.42	1.4550

	Mediana	Percentil 75	Máximo	Coeficiente de Variação (%)
Gross Profit	23373.500	33582.5000	47052.00	50.103902
EBITA	6345.500	12833.7500	17865.00	72.928227
EBITA Margin %	27.665	36.2875	39.89	27.515989
PE	15923.000	20002.0000	28331.00	43.216959
GP/PE	1.500	1.6200	1.73	8.105946

Fonte: Elaborado pelo autor.

Observa-se que a média de *Gross Profit* do período de Janeiro 2018 até Agosto de 2023 foi de 26.277,33 milhões de reais e sua mediana ficou em 23.373,50 milhões de reais, enquanto seu coeficiente de variação foi de 50,1% indicando uma alta variabilidade, contribuindo com essa hipótese temos seu desvio padrão de 13.165,96 que demonstra uma maior dispersão entre os dados do indicador com o valor mínimo de 13.106,00 milhões de reais e valor máximo de 47.052,00 milhões de reais, e obtendo os intervalos interquartil de 25% com o valor de 16.173,50, com 75% de 33.582,50.

O indicador de *EBITA* expressou em sua média o valor de 8.557,16 e obteve como mediana 6.345,50, em contrapartida seu desvio padrão foi de 6.240,58 revelando uma alta dispersão e teve um coeficiente de variação de 72% indicando variabilidade muito alta, com valor máximo de 17.865,00 e valor mínimo de 2.869,00, para o quartil de 25% o resultado foi de 3.777,50 e com 75% tivemos o resultado de 12.833,75 demonstrando um maior grau de dispersão em relação a medida central dos dados.

Para o *EBITA Margin %*, a média obtida foi de 29,48% e a mediana de 27,66%, o indicador teve o baixo desvio padrão de 8,11 e um coeficiente de variação médio de 27,51%, com valor máximo de 39,89% e valor mínimo de 21,80%, e um intervalo interquartil de 25% de 22,44% e de 75% de 36,28%.

O indicador de *PE* retornou a mediana de 15.923,00 e a média de 16.700,50, e denota um desvio padrão de 7.217,44 e coeficiente de variação de 43,21% demonstrando uma alta dispersão e variabilidade, além de possuir valor mínimo de 8.893,00 e valor máximo de 28.331,00, com quartil 25% de 11.309,25 e quartil de 75% de 20.002,00.

Por fim, na análise do *GP/PE* foi obtido a média de 1,54 e a mediana de 1,50, com um baixo desvio padrão de 0,12 sendo o menor entre os indicadores do estudo e um baixo coeficiente de variação de 8,10%, com valor mínimo de 1,42 e valor máximo de 1,73, além de indicar 1,45 no quartil de 25% e para o quartil de 75% o valor de 1,62.

A pesquisa prosseguiu com a análise da correlação linear de Pearson dos indicadores de desempenho financeiro selecionados, considerando o nível de significância de 5%, e excluindo o relacionamento entre *PE* e *GP/PE* que ficou acima do nível de 5%, os resultados obtidos evidenciam que os valores da correlação

possuem uma forte relação e apresentaram resultados estatisticamente significativos tendo os valores de $p < 0,05$, conforme resultados apresentados no quadro 3.

Quadro 3: Matriz de correlação de Pearson entre os indicadores de desempenho financeiro.

Variável 1	Variável 2	Correlação de Pearson	p-valor
Gross Profit	EBITA	0.989333	0.000170
Gross Profit	EBITA Margin %	0.907170	0.012526
Gross Profit	PE	0.993601	0.000061
Gross Profit	GP/PE	0.830840	0.040502
EBITA	EBITA Margin %	0.950322	0.003641
EBITA	PE	0.968760	0.001449
EBITA	GP/PE	0.890348	0.017376
EBITA Margin %	PE	0.873172	0.023108
EBITA Margin %	GP/PE	0.916257	0.010226
PE	GP/PE	0.767005	0.075106

Fonte: Elaborado pelo autor.

De modo geral, os dados demonstraram uma tendência a correlações fortes entre os indicadores do estudo, o coeficiente de correlação forte entre *Gross Profit* e *EBITA* e de *Gross Profit* e *PE*, indica ao longo dos dados históricos uma manutenção de custos operacionais proporcionais e pouca variação em depreciação e amortização, além de uma estrutura de custos eficiente e estável, refletindo em uma gestão financeira e operacional eficaz.

O coeficiente de correlação forte entre *EBITA Margin* e *Gross Profit* representa o relacionamento entre os indicadores das categorias de lucratividade, e reflete a eficiência da empresa do estudo em gerar lucro operacional em relação à sua receita, já os resultados apresentados entre *GP/PE* e *Gross Profit* e *PE* mostram correlações forte entre os indicadores, apesar de serem os menores resultados obtidos, entre possíveis aspectos complementares podemos inferir a falta da análise de aspectos como sazonalidade e ciclos econômicos que podem ocasionar na diminuição da correlação no período de análise do estudo.

As correlações de *EBITA* com os indicadores de *PE*, *GP/PE* e *EBITA Margin* evidenciam relações fortes que entre as explicações com base nos dados históricos podemos elencar uma boa relação entre os níveis de lucro bruto e os gastos com despesas pessoais, tendo como resultado uma razão *GP/PE* elevada, essa eficiência nos custos contribui para um maior *EBITA*.

Enquanto, as correlações fortes entre *EBITA Margin*, *GP/PE* e *PE* demonstram uma eficiência na utilização dos recursos e que a empresa do estudo está convertendo eficientemente seu lucro bruto em lucro operacional, já o *EBITA Margin* e *PE* sugere que o aumento nas despesas pessoais está acompanhado de um aumento na margem de *EBITA*, o que possibilita inferir que a empresa está investindo de forma estratégica, resultando em maior produtividade e lucratividade operacional, além de uma eficiência na gestão de pessoas, tendo em vista que as despesas pessoais não estão exclusivamente sendo compensadas, mas também contribuindo para um aumento na margem de *EBITA*, o que sugere que a gestão de recursos humanos está alinhada com os objetivos de lucratividade da empresa.

As companhias que possuem a característica de possuir uma correlação positiva entre *GP/PE* e *PE* demonstram uma gestão estratégica eficaz dos custos operacionais. Entre as práticas comumente usadas estão a eficiência na contratação, treinamento adequado e desenvolvimento de funcionários para maximizar a produtividade e reduzir desperdícios. Em geral, os resultados obtidos no quadro 3 indicam que caso um dos indicadores de desempenho aumentem os demais indicadores tendem a aumentar, indicando uma relação linear forte e positiva entre eles.

Com base na teoria de *Gradient Boosting Regression* apresentada por Belyadi e Haghighat (2021), o processo de criação do modelo de aprendizado de máquina supervisionado de *gradient boosting* foi desenvolvido no Google Colab com os dados históricos dos indicadores de desempenho financeiro selecionados da empresa objeto de estudo, foram usadas as bibliotecas da linguagem de programação Python como Pandas e NumPy para análise e manipulação do conjunto de dados, sendo complementados com o uso do Scikit-learn e do XGBoost para o desenvolvimento do modelo preditivo.

Além disso, foi dividido o conjunto dos dados entre treino e teste e para buscar melhorias nos hiperparâmetros foi usado o GridSearchCV do Scikit-learn, após feitas as previsões em relação aos valores reais foram obtidos os resultados previstos para o acumulado do ano de 2023 representados no quadro 4.

Quadro 4: Previsão dos indicadores de desempenho financeiro.

Indicadores de Desempenho Financeiro	Valores Previstos Acumulados do Ano de 2023
<i>Gross Profit - GP</i>	52.582
<i>EBITA</i>	17.864
<i>EBITA Margin %</i>	29,4%
<i>Personal Expenses - PE</i>	19.052
<i>GP/PE</i>	1.57

Fonte: Elaborado pelo autor.

Para definir a previsão dos indicadores de desempenho financeiro, no quadro 4, iniciou-se com a importação das bibliotecas para análise de dados como NumPy e Pandas, além da criação e avaliação pelo Scikit-learn e implementação do modelo XGBoost, foi criado um DataFrame que organiza os dados em uma tabela bidimensional, promovendo a substituição do y como variável dependente pelos valores do indicador financeiro selecionado e seus dados históricos e a alteração do x pelo ano na equação.

A partir disso, a previsão realizou-se pelo XGBoost e o aprimoramento de hiperparâmetros pelo GridSearchCV, e para complementar a criação do modelo é acrescido a significância de 5%, um intervalo de confiança de 95% é calculado gerando múltiplas previsões e a qualidade do modelo é avaliada usando R-Quadrado, RMSE e MAE.

Em relação aos resultados previstos alcançados podemos analisar um aumento do *Gross Profit* em relação ao ano anterior em cerca de 11,1%, seguindo para o indicador *EBITA Margin*, o modelo de predição indicou uma diminuição representativa de 25,4% em relação ao ano de 2022.

Já nas métricas de desempenho financeiro de *GP/PE* e *PE* foram obtidos os resultados de diminuição em 5,57% do *GP/PE* e de um aumento do *PE* em cerca de 6,43% em relação ao acumulado do ano de 2022, e por fim a previsão do indicador de *EBITA* demonstrou uma baixa variação percentual em relação aos dados acumulados do ano anterior.

As técnicas de qualidades utilizadas nos modelos de predição de cada indicador mantiveram por padrão um R-Quadrado próximo a 1, além de resultados de MAE e RMSE baixos em relação à escala do conjunto de dados. Fundamentado nos resultados obtidos podemos observar um cenário de menor crescimento em indicadores de qualidade do lucro como o *GP/PE* e *EBITA Margin*, tendo como contrapartida um aumento nos custos operacionais e apesar da previsão de aumento no lucro bruto, não se refletiu em aumento significativo no *EBITA* em virtude de perda de rentabilidade operacional.

5. CONSIDERAÇÕES FINAIS

O presente estudo buscou demonstrar e aprofundar sobre o uso de algoritmos de aprendizado de máquina supervisionados para a eficiência e predição de indicadores de desempenho financeiro, bem como retratar para as companhias, gestores e profissionais dos setores financeiros a relevância da utilização desses algoritmos e seus benefícios.

O uso de técnicas de aprendizado de máquina propicia maior precisão na identificação de padrões complexos e não lineares no conjunto de dados, além de possuir um potencial de automatização e de processamento de grandes volumes de dados de forma automatizada, reduzindo o tempo gasto em análises manuais por equipes de FP&A e permitindo aos membros se concentrarem em decisões estratégicas.

A adaptabilidade dos modelos também contribui para sua difusão no atual ambiente corporativo caracterizado por suas constantes mudanças e inovações, ao possibilitar incorporar dados diariamente e ajustar previsões de acordo com novos cenários.

Com a elaboração do modelo apresentado no estudo, e os dados de predição encontrados compreende-se que a adoção de modelos de aprendizado de máquina contribuem para geração de ideias, previsibilidade e gestão estratégica de custos para o ambiente corporativo. Concluiu-se com a aplicação do método de previsão de *gradient boosting regression* que por meio das projeções futuras torna-se possível para a empresa planejar adequadamente suas ações financeiras. Ao ser fornecido pelo modelo um conjunto de subsídios para a tomada de decisão referente aos resultados esperados para cada indicador de desempenho financeiro. No entanto, se faz necessário a manutenção dos modelos para assegurar resultados e previsões consistentes com a realidade dinâmica das companhias.

Para o ambiente corporativo, a importância deste estudo reside em fornecer uma perspectiva sobre a aplicação de modelos de aprendizado de máquina em uma companhia do setor privado, e apresentar a análise das previsões.

Contudo, o modelo desenvolvido requer aprimoramentos para alcançar maior precisão e robustez nas previsões, seria de grande relevância para o modelo serem

incorporadas informações de mercado e uma análise de sazonalidade perante os dados históricos, a inclusão dessas ações permitirá ao modelo considerar as flutuações e tendências do mercado, bem como os padrões sazonais que podem impactar o desempenho financeiro, essas adições são fundamentais para ajustar o modelo às realidades dinâmicas do mercado, melhorando assim a sua capacidade preditiva.

É importante ressaltar que a utilização de dados textuais, tais como notícias, relatórios públicos de empresas, observações e comentários de especialistas e relatórios de auditores, além da análise de sentimentos e a extração de informações relevantes desses textos podem oferecer contribuições valiosas que complementam os dados quantitativos e proporciona uma visão mais holística do desempenho financeiro.

Em suma, o presente estudo confirma o potencial significativo dos modelos de aprendizado de máquina na previsão de indicadores financeiros, e destaca a necessidade de contínuos desenvolvimentos e aprimoramentos. Com a implementação das melhorias sugeridas, espera-se que esses modelos possam oferecer previsões ainda mais precisas e úteis, proporcionando uma vantagem competitiva substancial em um ambiente financeiro cada vez mais complexo e dinâmico.

REFERÊNCIAS

- BASHIR, D.; MONTAÑEZ, G. D.; SEHRA, S.; SEGURA, P. S.; LAUW, J. **An information-theoretic perspective on overfitting and underfitting**. Springer International Publishing. AI 2020: Advances in Artificial Intelligence: 33rd Australasian Joint Conference, Proceedings 33 (pp. 347-358). Australia, November, 2020.
- BELYADI, H.; HAGHIGHAT, A. **Machine Learning Guide for Oil and Gas Using Python: A Step-by-Step Breakdown with Data, Algorithms, Codes, and Applications**. Gulf Professional Publishing, April 2021.
- BODNER, K.; FORTIN M.-J.; MOLNÁR P.K. **Making Predictive modelling ART: accurate, reliable, and transparent**, Ecosphere 11(6):e03160. 10.1002/ecs2.3160, 2020.
- BOTCHKAREV, A. **Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology**. arXiv, 2018.
- BREALE R.; MYERS S.; ALLEN F. **Princípios de Finanças Corporativas**. 12a. Ed., 2018.
- BRUCE, P.; BRUCE, A. **Estatística prática para cientistas de dados: 50 conceitos essenciais**. Rio de Janeiro: Alta Books, 2019.
- CHEN, T.; GUESTRIN, C. **Xgboost: A scalable tree boosting system**. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). New York, NY, USA: ACM, 2016.
- CRESWELL, J. W. **Projeto de pesquisa: métodos qualitativo, quantitativo e misto**. 5 ed. Porto Alegre: Penso, 2021.
- DELEN, D.; KUZEY, C.; UYAR, A. **Measuring firm performance using financial ratios: A decision tree approach**. Expert Systems with Applications, Volume 40, Issue 10, 2013.
- FLICK, U. **Introdução à pesquisa qualitativa**. Tradução Joice Elias Costa. 3.ed. Porto Alegre: Artmed, 2009.

FP&A TRENDS GROUP. **FP&A Trends Survey**. June, 2023.

GANGULI, R. **Future Direction of Financial Planning and analysis**, Mangalore University, India, December 2020.

GIL, A. C. **Como elaborar projetos de pesquisa**. 6. ed. São Paulo: Atlas, 2017.

GÖKÇE, M. M.; DUMAN, E. **Performance Comparison of Simple Regression, Random Forest and XGBoost Algorithms for Forecasting Electricity Demand**. 2022 3rd International Informatics and Software Engineering Conference (IISEC), Ankara, Turkey, 2022.

HIGHTOWER, R. **Accounting and finance policies and procedures**, Hoboken, N.J., John Wiley & Sons, 2008.

JASIMUDDIN, A.; ADAMS, L. **AI and ML in FP&A – Two Case Studies**, Digital FP&A Events Insights, 2020.

JORDAN, M.I.; MITCHELL, T.M. **Machine Learning: Trends, Perspectives, and Prospects**. Science, 2015.

JOSEPH, V.R. **Optimal ratio for data splitting**, Wiley, Statistical Analysis and Data Mining: The ASA Data Science Journal, N° 4, p. 531-538, 2022.

JURAFSKY D.; MARTIN, J. H. **Speech and Language Processing**, (3rd edition) Pearson, 2022.

KARNOV GROUP. **Financial definitions**, Disponível em: <<https://www.karnovgroup.com/en/financial-definitions/>>. Acesso em: 12 Abril 2024.

KULKARNI, P. **Advanced Analytics Driven Financial Management: An Innovative Approach to Financial Planning & Analysis**, International Journal of Computer Trends and Technology, 2023.

KUMAR, S.; CHONG, I. **Correlation Analysis to Identify the Effective Data in Machine Learning: Prediction of Depressive Disorder and Emotion States**. International Journal of Environmental Research and Public Health, 2018.

LECUN, Y.; BENGIO, Y.; HINTON, G. **Deep learning**. 521, 436–444, Nature, 2015.

LIN, K. Y.; HUANG, C. **Ensemble learning applications in multiple industries: A review**. Inf. Dyn. Appl., vol. 1, no. 1, pp. 44-58, 2022.

LOPES, G. R.; ALMEIDA, A. W. S.; DELBEM, A. C. B.; TOLEDO, C. F. M. **Introdução à Análise Exploratória de Dados com Python**. Escola Regional de Computação Aplicada à Saúde (ERCAS), 2019.

LÓPEZ DE PRADO, M. **Advances in Financial Machine Learning**. John Wiley & Sons, 2018.

MAHESH, B. **Machine Learning Algorithms - A Review**. International Journal of Science and Research, 9, 381-386, 2020.

MCKINNEY, W. **Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter**. 3rd Edition, 2022.

MOHAMMED, A.; KORA, R. **A comprehensive review on ensemble deep learning: Opportunities and challenges**. Journal of King Saud University - Computer and Information Sciences Volume 35, Issue 2, Pages 757-774, February 2023.

OECD – ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges, and Implications for Policy Makers**, 2021.

OLIPHANT, T. **Python for scientific computing**. Computing in Science & Engineering, vol. 9, no. 3, pp. 10-20, May-June 2007.

PFISTER, H.; WATTENBERG, M.; BEYER, J.; NOBRE, C. **Exploring the Gap Between Informal Mental and Formal Statistical Models**. Harvard Data Science Review, 2021.

PwC Consulting. **Future of Finance - Today's CFO - shaping tomorrow, today**. July, 2023.

RAO *et al.* **Feature selection based on artificial bee colony and gradient boosting decision tree**. Applied Soft Computing, Volume 74, Pages 634-642, 2019.

RAYHAN, A.; GROSS, D. **The rise of Python: A survey of recent research**. 2023.

REN, Y.; ZHANG, L.; SUGANTHAN, P. N. **Ensemble Classification and Regression-Recent Developments, Applications and Future Directions**. IEEE Computational Intelligence Magazine, vol. 11, no. 1, pp. 41-53, 2016.

ROSS, S. A.; WESTERFIELD R. W.; JORDAN, B. D. **Fundamentos de Administração Financeira**. 9. Ed., 2013.

RUSSELL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach**, 4th edition, Pearson, 2020.

SEBER, G. A. F.; LEE, A. J. **Linear Regression Analysis**. 2nd Edition. ISBN: 978-1-118-27442-2, January 2012.

SEBRAE. **A importância da gestão financeira empresarial**, 2022. Disponível em: <<https://sebrae.com.br/sites/PortalSebrae/artigos/a-importancia-da-gestao-financeira-empresarial,624d36b750c32810VgnVCM100000d701210aRCRD>>. Acesso em: 17 Março 2024.

STATISTICS FINLAND. **Personnel expenses**, Disponível em: <https://www.stat.fi/meta/kas/henkku_en.html>. Acesso em: 01 Maio 2024.

STOBIERSKI, T. **CASH FLOW VS. PROFIT: WHAT'S THE DIFFERENCE?**. Harvard Business School Online's Business Insights Blog, April, 2020.

SURAHYO, B. **The Present And Future Of Financial Planning And Analysis (FP&A)**, Forbes, 2023.

SWISS STEEL GROUP. **Annual Report**. 2023.

YIN, R. K. **Case Study Research and Applications: Design and Methods**. Washington DC: Sage Publications, 2017.

ZHAO, C.; YUAN, X.; LONG, J.; JIN, L.; GUAN, B. **Financial Indicators Analysis by Machine Learning: Evidence from Chinese Stock Market**. July 2023.