Análise Preditiva de Despesas com o Programa de Participação nos Resultados (PPR) a partir de Demonstrativo do Resultado do Exercício (DRE) utilizando modelos de Aprendizado de Máquina (AM)

Welliton Slaviero

[1] Instituto de Tecnologia - Universidade de Passo Fundo (UPF)
BR 285 Km 292,7 – Bairro São José – Passo Fundo, RS, Brasil

178342@upf.br

Abstract. *Companies face significant challenges in managing bonus expenses. Therefore, this study aims to forecast the expenses of the Profit Sharing Program (PSP), using financial data from the Profit and Loss (P&L) Statement of a small company. The techniques adopted, such as SARIMAX, Prophet and Random Forest, were evaluated in terms of accuracy, revealing potential improvements in expense control and operational sustainability. This work contributes to the strategic use of Machine Learning (ML) in business scenarios.*

Resumo. *Empresas enfrentam desafios significativos na gestão de despesas provenientes de bonificações. Portanto, o presente estudo tem como objetivo prever as despesas do Programa de Participação nos Resultados (PPR), utilizando dados financeiros do Demonstrativo do Resultado do Exercício (DRE) de uma empresa de pequeno porte. As técnicas adotadas, como SARIMAX, Prophet e Random Forest, foram avaliadas em termos de precisão, revelando potenciais melhorias no controle de despesas e na sustentabilidade operacional. Este trabalho contribui para o uso estratégico de Aprendizado de Máquina (AM) em cenários empresariais.*

1. Introdução

Em um contexto geral, empresas tendem a enfrentar constantemente desafios quanto a gestão eficaz de seus encargos financeiros, fazendo-se de extrema importância a adoção de planejamentos estratégicos, principalmente em períodos turbulentos de incertezas econômicas [Silva et al. 2007].

Frente a essa realidade, muitas empresas optam pela utilização e manutenção de sistemas de bonificação de seus funcionários pelo desempenho individual em diversos formatos [Salami et al. 2018], como no caso do Programa de Participação nos Resultados (PPR), sendo uma abordagem de interesse das organizações devido aos seus resultados positivos amplamente difusos [Souza et al. 2005]. Frizon e Binsfeld alegam que os resultados de programas como o PPR demonstram não apenas melhorias nos índices de produtividade, mas também um fortalecimento da cultura de comprometimento organizacional [Frizon & Binsfeld 2021].

O acompanhamento periódico dessas despesas e outros indicadores financeiros é essencial para tomada de decisões estratégicas. Neste contexto, o Demonstrativo do

Resultado do Exercício (DRE) ajuda a identificar áreas que necessitam de ajustes ou otimizações [Belisário & Nikolay 2022].

Entretanto, as principais influências, como mudanças econômicas e a volatilidade de indicadores financeiros, podem impactar diretamente as despesas com PPR, levando a desvios superiores a 20% do que é previsto [Leal et al. 2009]. Ademais, variações no desempenho dos colaboradores e alterações nos critérios de avaliação podem comprometer até 30% da previsibilidade, aumentando os riscos ao planejamento estratégico [Zenker 2010], tornando o processo de controlar este gasto muito desafiador.

O uso de modelos de Aprendizado de Máquina (AM) tem-se mostrado muito competente para análise preditiva envolvendo problemas dessa natureza, como pode ser observado nos trabalhos [Elrefaei 2024], [Ghosh 2024] e [Gao & Huang 2021]. Essas tecnologias permitem previsões financeiras mais confiáveis ao incorporar abordagens que lidam com incertezas na análise de grandes volumes de dados [Saha et al. 2023].

No contexto deste estudo, tal abordagem é particularmente relevante para empresas de pequeno porte, como a empresa Soluct, uma software house localizada em uma cidade pequena e que possui entre 10 a 15 funcionários em operação.

A Soluct enfrenta desafios específicos, como a dificuldade em prever despesas com PPR devido à sazonalidade trimestral atualmente adotada. Além disso, a empresa considera a realização das metas gerais do time e uma série de indicadores de desempenho baseados nas atividades realizadas pelos colaboradores. Isso resulta em uma pontuação individual que é usada para determinar a proporção de bonificação que o indivíduo receberá, considerando seu salário. Além de tudo, a empresa trabalha com recursos financeiros limitados sendo um desafio único para este estudo.

Apesar disso, a empresa busca continuamente ajustar seu modelo de PPR para alinhar melhor a valorização do desempenho dos colaboradores com a sustentabilidade financeira. A adoção de um modelo de bonificação trimestral visa simplificar a previsão e distribuição das despesas, mas não elimina os desafios relacionados à sazonalidade e à precisão nas projeções.

Este estudo também explora a possibilidade de estender a análise para sazonalidades anuais, oferecendo insights que podem beneficiar outras empresas que optem por diferentes abordagens de periodicidade em seus programas de PPR.

Com base no exposto, este trabalho tem como objetivo geral desenvolver algoritmos de previsão utilizando modelos de Aprendizado de Máquina (AM), com o diferencial de estimar as despesas sazonais relacionadas ao Programa de Participação nos Resultados (PPR) a partir do Demonstrativo do Resultado do Exercício (DRE), contribuindo para um planejamento financeiro mais eficaz e alinhado à realidade operacional da empresa Soluct e outras semelhantes.

2. Fundamentação Teórica

Nesta seção, apresentam-se os conceitos essenciais para o entendimento e embasamento do trabalho em questão.

2.1. Programa de Participação nos Resultados (PPR)

Modelo de remuneração variável amparado pela Lei nº 10.101/2000 [Brasil 2000], que oferece aos colaboradores uma gratificação baseada no desempenho organizacional e no alcance de objetivos específicos. Essa modalidade tem ganhado destaque por sua capacidade de alinhar os interesses dos funcionários aos da empresa, promovendo maior engajamento e produtividade.

Os PPRs são classificados como despesas operacionais variáveis, influenciando diretamente a estrutura de custos das organizações. A natureza variável do PPR significa que ele está sujeito a oscilações de acordo com a performance organizacional e as metas estabelecidas, impactando, consequentemente, os resultados financeiros.

Empresas que adotam esse modelo precisam equilibrar os benefícios do incentivo ao desempenho com os desafios de previsibilidade e sustentabilidade financeira. De acordo com Motta, ele difere do PLR[1], pois foca em indicadores de desempenho mais amplos e customizáveis para cada organização [Motta 2005].

2.2. Demonstrativo do Resultado do Exercício (DRE)

Consiste em um relatório contábil que resume todas as operações financeiras de uma empresa que ocorreram em períodos predeterminados, frequentemente sendo contabilizados no formato anual [Marion 2019]. Ele segue o regime de competência e é obrigatório conforme a Lei nº 6.404/1976 [Brasil 1976].

As principais variáveis utilizadas neste estudo são oriundas da DRE, haja vista que ela apresenta de forma sucinta e consolidada a situação econômica que a empresa enfrenta, com o intuito de identificar padrões e prever tendências a longo prazo.

2.3. Aprendizado de Máquina (AM)

Subcampo da Inteligência Artificial (IA), o Aprendizado de Máquina (AM) permite que sistemas compreendam e façam previsões com base em grandes volumes de dados, sem a necessidade de programação manual para cada situação. Na otimização das despesas com PPR, o AM auxilia na identificação de padrões, aprimorando o planejamento financeiro [Jung 2022].

2.4. Análise Preditiva

Emprega análise de informações e dados históricos na tentativa de antecipar o futuro, reconhecendo padrões que podem auxiliar na compreensão de tendências [Ogunleye 2014]. No contexto empresarial, ela é amplamente aplicada para prever vendas, demandas de mercado, e despesas, como as associadas ao PPR. A capacidade de prever valores monetários com precisão, é um diferencial crucial para reduzir os riscos financeiros nas empresas.

3. Trabalhos Relacionados

Nesta seção, são brevemente apresentados trabalhos relacionados recentes que abordam análise preditiva com auxílio de Aprendizado de Máquina (AM) em séries temporais e variações relacionadas ao cenário empresarial.

[1]PLR (Participação nos Lucros e Resultados): modelo de incentivo empresarial que visa distribuir parte dos lucros da empresa aos funcionários com base em resultados financeiros globais.

3.1. Time Series Forecasting Using Machine Learning

O trabalho de [Elrefaei 2024] oferece uma análise bem abrangente com uma grande variedade técnicas de Aprendizado de Máquina (AM), destacando modelos clássicos como ARIMA[2] e Seasonal Auto-Regressive Integrated Moving Average with Exogenous Variables (SARIMAX), modelos supervisionados como Random Forest e SVM[3] e técnicas de aprendizado profundo como LSTM[4] e RNN[5]. Destaca o uso de Python, bibliotecas relevantes e uma interface que permite personalizar previsões, discutindo desafios e limitações como dados não estacionários e lacunas em datasets.

Destacando a diversidade de modelos avaliados, a implementação de uma interface para simulações e a análise criteriosa de métricas como RMSE e MAE. Esses elementos reforçam a aplicabilidade prática e oferecem insights valiosos para o uso de AM em contextos empresariais.

Por outro lado, o artigo apresenta limitações ao tratar de modelos multivariados, focando principalmente em séries univariadas. Além disso, depende fortemente de um pré-processamento de dados bem executado, sem detalhar suficientemente este processo. Do mesmo modo, a robustez dos modelos em cenários drasticamente diferentes não é discutida em profundidade.

Este trabalho serve como uma base sólida para a aplicação e desenvolvimento de soluções práticas no uso de Aprendizado de Máquina (AM) para previsão de séries temporais, sendo uma das principais influências pelas escolhas de materiais e métodos do presente estudo.

3.2. Forecasting Firm Decision to Change in Capital Structure

Já [Ghosh 2024] examina o uso de Aprendizado de Máquina (AM), com destaque a modelos como Random Forest, para prever decisões corporativas relacionadas à estrutura de capital, destacando a relevância de dados financeiros detalhados, por utilizar relatórios 10-K[6] de 51 empresas americanas entre 2017 e 2023.

Uma das principais contribuições do trabalho é a habilidade de incorporar diversas variáveis não lineares e dados extensivos em modelos preditivos, algo essencial para cenários complexos e dinâmicos. A estrutura adotada pode ser replicada para modelar o comportamento de despesas, aproveitando padrões históricos para antecipar variações futuras.

No entanto, o artigo também ressalta limitações relacionadas à qualidade dos dados e à complexidade dos algoritmos. A dependência de dados estruturados como XBRL[7]

[2]ARIMA (AutoRegressive Integrated Moving Average): modelo estatístico usado para análise e previsão de séries temporais, combinando componentes autorregressivos, diferenciação e médias móveis.

[3]SVM (Support Vector Machines): algoritmo de aprendizado supervisionado que classifica dados ao encontrar o hiperplano que melhor separa as classes.

[4]LSTM (Long Short-Term Memory): técnica de aprendizado profundo baseada em redes neurais recorrentes, projetada para capturar dependências de longo prazo em sequências.

[5]RNN (Recurrent Neural Network): modelo de aprendizado profundo que processa dados sequenciais, reutilizando informações de etapas anteriores para capturar dependências temporais e contextuais.

[6]10-K: documentos anuais exigidos pela SEC nos EUA, detalhando o desempenho financeiro, riscos e operações.

[7]XBRL (eXtensible Business Reporting Language): linguagem padronizada para a troca de dados financeiros e empresariais.

pode não ser viável em todas as realidades empresariais, especialmente quando os dados disponíveis não seguem padrões internacionais.

O estudo é imporante por abordar o uso de dados financeiros o que corrobora com a utilização da DRE para observação de tendências nas previsões de despesas com PPR por exemplo.

3.3. Forecast of Compensation Amount Based on Big Data Network and Machine Learning Algorithm in Intellectual Property Law

Por fim, [Gao & Huang 2021] discutem a utilização de algoritmos de Aprendizado de Máquina (AM), como K-means[8] e SVMs, para prever compensações financeiras no contexto da lei de propriedade intelectual. Ele aborda como a compensação baseada em desempenho está correlacionada à satisfação no trabalho e eficiência corporativa.

O estudo sugere que essas técnicas podem melhorar a precisão das previsões de compensação e contribuir para a gestão estratégica de recursos humanos. Destaca também o uso de FPGAs[9] para otimizar o processamento de dados em tempo real, proporcionando soluções eficientes para análises complexas.

Os resultados obtidos no estudo indicaram que essas técnicas podem melhorar e muito a precisão das previsões de compensação, tendo como um dos principais pontos fortes do estudo a escolha do FPGAs, que proporcionou maior eficiência nas operações.

Contudo, o artigo apresenta limitações, como a ausência de comparações detalhadas fazendo o uso de outros modelos, o que dificulta a avaliação da robustez do método em outros tipos de cenários. Além disso, sua aplicabilidade a pequenas empresas não foi explorada.

Essa perspectiva reforça a importância de incorporar métodos preditivos no contexto de análise de dados financeiros. Porém devido a complexidade desnecessária e elementos de alto uso de hardware tornou inviável no contexto do presente trabalho.

4. Materiais e Métodos

Nesta seção, são apresentados os dados, ferramentas e métodos utilizados para realizar a previsão proposta no presente trabalho.

4.1. Dataset

O conjunto de dados utilizado neste trabalho foi extraído do DRE da empresa Soluct pela equipe de gestão, que garantiu a confiabilidade das informações. Os dados possuem formato tabular, organizados em uma planilha Excel contendo 22 variáveis numéricas no formato decimal com duas casas decimais, além de uma variável temporal. A tabela inclui 82 registros mensais que abrangem o período de janeiro de 2018 a outubro de 2024, com cada linha representando o último dia do respectivo mês. Os valores foram sintetizados para refletir o somatório dos registros financeiros mensais da empresa.

[8]K-Means: algoritmo de aprendizado não supervisionado que agrupa dados em k clusters, minimizando a variância interna de cada grupo.

[9]FPGAs (Field Programmable Gate Arrays): circuitos integrados reconfiguráveis que permitem implementar hardware personalizado.

Apesar de sua confiabilidade, algumas limitações do dataset devem ser consideradas. As despesas com o PPR apresentam registros apenas nos meses de janeiro, abril, julho e outubro devido à sazonalidade trimestral adotada pela empresa, resultando em lacunas nos dados para os demais meses. Além disso, as variáveis receitas e despesas apresentavam valores que extrapolavam tanto positivamente quanto negativamente, sendo necessário um cuidado adicional no pré-processamento para tratar outliers e garantir a consistência dos dados.

Os valores históricos de despesas com o PPR foram ajustados para refletir a sazonalidade trimestral atual, uma vez que o modelo trimestral implementado pela empresa só passou a valer a partir de janeiro de 2023. Antes disso, as bonificações eram realizadas sem um padrão sazonal estabelecido. Esse ajuste foi feito com base em uma média de crescimento entre 5% e 15%, conforme consultoria com a empresa.

Para permitir análises em uma perspectiva anual, foi criado um novo dataset a partir da cópia do original, onde as lacunas foram preenchidas utilizando interpolação linear, capturando variações mês a mês em vez de apenas trimestrais. O tamanho reduzido do dataset, com apenas 82 registros, também impõe limitações quanto à generalização dos resultados, restringindo sua aplicabilidade para períodos mais extensos ou outros contextos organizacionais.

4.1.1. Variáveis Analisadas

As variáveis analisadas no estudo foram selecionadas a partir do dataset principal da DRE da empresa Soluct. Entre elas, destacam-se:

- **Mês**: representando o último dia de cada mês como indicador temporal essencial para análises sazonais;
- **Receitas**: abrangem todas as entradas financeiras mensais da empresa;
- **Despesas**: indicando as saídas financeiras;
- **Lucro Líquido**: obtido pela diferença entre receitas e despesas mensais;
- **Lucro Líquido Acumulado**: reflete uma soma progressiva do lucro líquido ao longo do tempo e serve como variável explicativa;
- **Despesa com Pessoal PPR**: definida como o custo direto do programa de bonificação trimestral da empresa, considerada variável-alvo.

A análise de correlação entre essas variáveis, apresentada no gráfico 1, revelou fortes associações entre “Despesas”, “Receitas” e o “Lucro Líquido”, com correlações positivas significativas entre receitas e lucro, e negativas entre despesas e lucro. Essa relação evidencia que as variáveis selecionadas são altamente representativas para capturar as dinâmicas financeiras da empresa. A escolha dessas variáveis justifica-se pela relevância do impacto das receitas e despesas no cálculo do lucro líquido, que, por sua vez, influencia diretamente a previsão de despesas com PPR.

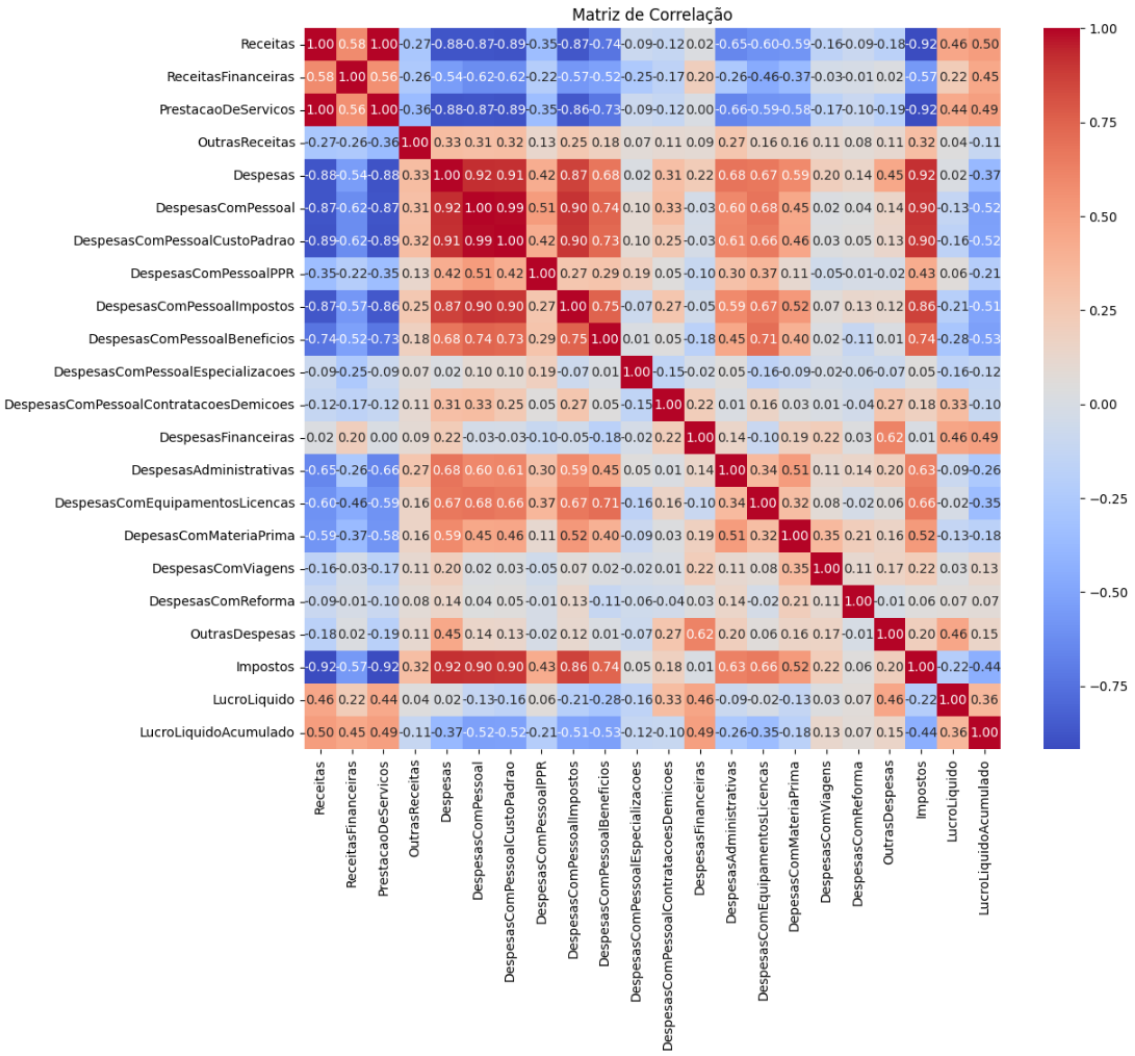

Figura 1. Mapa de calor da matriz de correlação da variáveis.

4.1.2. Sazonalidade

A análise da sazonalidade das despesas com PPR feita com auxílio do gráfico 2 revelou padrões significativos na série temporal. A tendência apresenta um comportamento decrescente ao longo dos anos, refletindo uma redução contínua nas despesas, indicando que a empresa está destinando valores cada vez maiores ao pagamento de bonificações aos funcionários.

O componente sazonal destaca oscilações trimestrais regulares, com picos em janeiro, abril, julho e outubro, alinhados à periodicidade de distribuição do PPR adotada pela empresa a partir de 2023. Este padrão reforça a relevância de considerar a sazonalidade nos modelos preditivos.

Os resíduos, distribuídos aleatoriamente em torno de zero, indicam que os modelos capturaram bem a tendência e a sazonalidade. Contudo, pequenas variações sugerem fatores externos não modelados que podem influenciar as despesas.

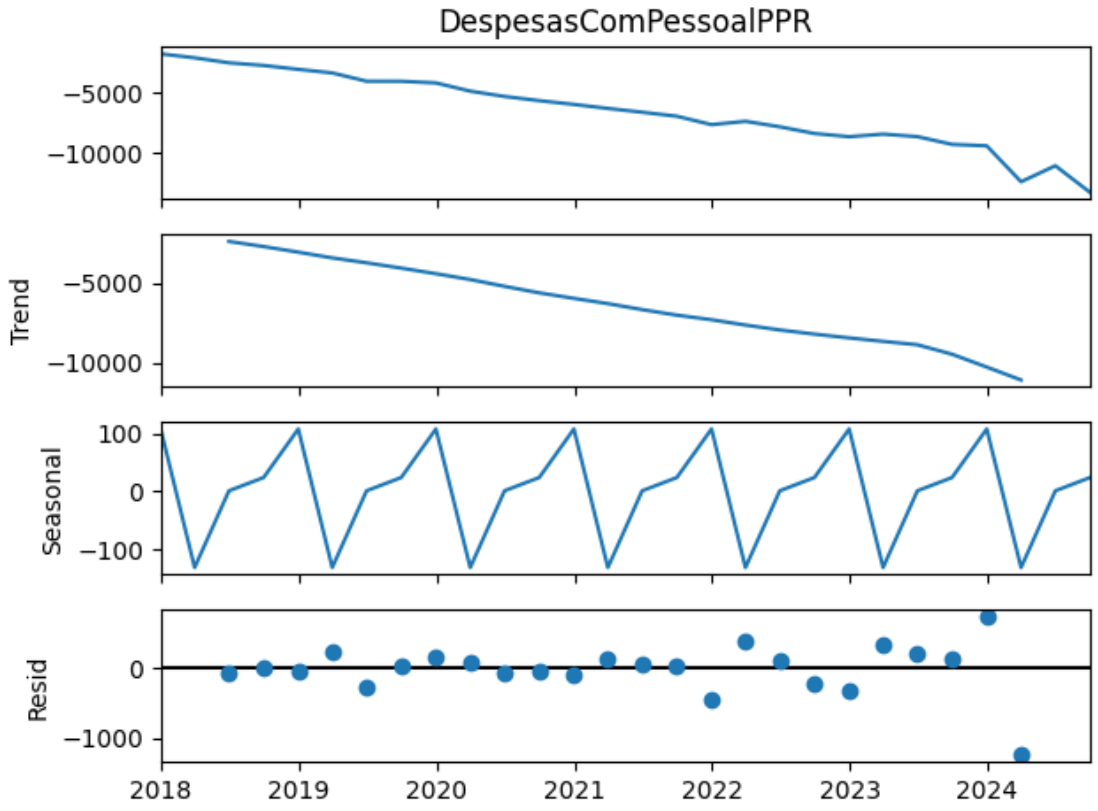

Figura 2. Mapa de calor da matriz de correlação da variáveis.

4.1.3. Processamento de Dados

O pré-processamento escolhido envolveu a transformação dos dados, convertendo as datas para o formato datetime e realizando o preenchimento de valores nulos por 0 para garantir que nenhum erro ocorra nas próximas etapas do pipeline. Quanto ao primeiro algoritmo, responsável pela previsão de lucro, que será detalhado mais adiante no tabalho, optou-se pela remoção de outliers para garantir a qualidade nos dados, utilizando os métodos estatísticos de remoção por Percentis, IQR e Z-Score, abordados brevemente a seguir:

O método de Percentis utiliza valores percentuais da distribuição dos dados para definir limites. Nesse caso, esses limites foram calculados diretamente da distribuição dos dados de cada variável, sendos eles o limite inferior de ($P_{1\%}$) e superior de ($P_{99\%}$).

Qualquer valor fora deste intervalo foi considerado um outlier e removido. Este método é eficaz para eliminar valores extremos sem fazer suposições sobre a forma da distribuição dos dados.

Já o Intervalo Interquartil (IQR) é uma medida baseada na diferença entre o terceiro quartil (Q3) e o primeiro quartil (Q1) da distribuição dos dados. Esses quartis dividem os dados em limites sendo eles o inferior de ($Q1 - 1.5 \times IQR$) e superior de ($Q3 + 1.5 \times IQR$).

Valores fora desse intervalo foram classificados como outliers e removidos. Esse método é amplamente utilizado por ser robusto contra valores extremos em distribuições

assimétricas.

O método de Z-Score mede o número de desvios padrão de um valor em relação à média de cada variável. Ele é calculado pela fórmula: $Z = \frac{x-\mu}{\sigma}$, onde x representa o valor da variável analisada, μ é a média dos valores da variável e σ é o desvio padrão.

Com base nessa métrica, foram considerados outliers os valores com $|Z| > 3$, ou seja, valores que estavam mais de três desvios padrão acima ou abaixo da média. Esses valores foram removidos para evitar que pontos extremos distorcessem as análises.

Para os demais algoritmos que focam na previsão da despesa com PPR, foi descontinuada a remoção de outliers, devido a necessidade de manter a sazonalidade trimestral. Remover qualquer mês nesta etapa, corromperia a análise adicionando complexidade desnecessária no seguimento do pipeline.

Como solução simples e direta para ausência de valores fora da sazonalidade trimestral, foi feito o isolamento dos dados trimestrais e a cópia dos mesmos, aplicando interpolação linear na despesa com PPR, para preencher os dados em todos os meses. Dessa forma, possibilitando a análise tanto trimestral quanto anual.

Vale citar que a interpolação linear, estima os valores ausentes com base em uma progressão linear entre os valores adjacentes, garantindo uma transição suave e consistente.

4.2. Ferramentas Utilizadas

O desenvolvimento deste trabalho foi realizado utilizando a linguagem Python, reconhecida por sua ampla aplicação em ciência de dados e Aprendizado de Máquina (AM) devido à simplicidade sintática e à vasta biblioteca de ferramentas disponíveis. Para a manipulação e organização dos dados, foi utilizada a biblioteca Pandas, que possibilitou o carregamento, limpeza e transformação dos datasets da DRE e do PPR. Essas operações incluíram a padronização das frequências temporais (anual e trimestral), essenciais para a análise de séries temporais.

O suporte para operações matemáticas e manipulação de arrays foi fornecido pela biblioteca NumPy, especialmente importante para o cálculo de métricas de desempenho e preparação dos dados de entrada dos modelos. A visualização dos dados e resultados foi realizada por meio das bibliotecas Matplotlib e Seaborn, que permitiram criar gráficos detalhados, como mapas de calor para análise de correlação entre variáveis, bem como séries temporais para verificar tendências e sazonalidades.

O modelo SARIMAX foi implementado com a biblioteca Statsmodels, que oferece suporte para modelagem de séries temporais com variáveis exógenas, possibilitando incorporar o "Lucro Líquido Acumulado" como um fator externo relevante. Adicionalmente, o Prophet foi utilizado para lidar com séries temporais não lineares, caracterizadas por tendências e sazonalidades complexas. Essa escolha se justifica por sua capacidade de adicionar sazonalidades customizadas, como a sazonalidade trimestral e anual observadas nos dados do PPR.

A biblioteca Scikit-Learn foi empregada para a aplicação do modelo Random Forest Regressor, que é amplamente utilizado para capturar relações não lineares e interações entre variáveis, além de possibilitar a avaliação com métricas como MAE, MAPE, RMSE

e R^2. O Pmdarima foi utilizado para automatizar a busca de parâmetros adequados para o SARIMAX, reduzindo o tempo necessário para encontrar combinações eficientes de ordem ARIMA e sazonalidade. Finalmente, a biblioteca SciPy auxiliou na remoção de outliers dos dados por meio de técnicas como Z-Scores, Percentis e Intervalo Interquartil, aumentando a confiabilidade das previsões.

4.2.1. SARIMAX

O modelo Seasonal Auto-Regressive Integrated Moving Average with Exogenous Variables (SARIMAX) é um modelo poderoso para análise de séries temporais, ideal para capturar padrões complexos que se repetem ao longo do tempo. Com sua capacidade de incluir sazonalidade e variáveis externas, ele não apenas entende os dados históricos, mas também considera fatores externos que influenciam o futuro. É uma ferramenta robusta e confiável para previsões em cenários como economia, vendas e comportamento climático.

Foi implementado tanto na previsão do lucro quanto das despesas com PPR, porém serão abordados os parâmetros e ajustes feitos exclusivamente para despesa com PPR. Sendo assim, o modelo foi utilizado para capturar padrões temporais complexos e modelar a influência do "Lucro Líquido Acumulado" como variável exógena. Este modelo foi escolhido devido à sua robustez em lidar com séries temporais que apresentam tanto componentes sazonais quanto não sazonais.

Para os dados anuais, a ordem ARIMA foi definida como (5, 0, 0), indicando que o modelo inclui cinco termos autorregressivos e nenhuma diferenciação ou média móvel. A sazonalidade anual foi capturada pela ordem sazonal (0, 1, 1, 12), refletindo um padrão recorrente a cada 12 meses. Para os dados trimestrais, foi utilizada a configuração (1, 0, 1) para a ordem ARIMA, enquanto a ordem sazonal foi ajustada para (0, 1, 1, 4), considerando a periodicidade trimestral. Métricas como MAPE, MAE, RMSE e R^2 confirmaram a eficácia do modelo em previsões de despesas com PPR, tornando-o adequado para o cenário analisado.

4.2.2. Prophet

O Prophet, por outro lado, se destaca pela simplicidade e flexibilidade. Desenvolvido pelo Facebook, ele transforma a previsão de séries temporais em um processo descomplicado, capaz de lidar com sazonalidade, tendências imprevisíveis e até mesmo lacunas nos dados.

Foi selecionado devido à sua capacidade de lidar com séries temporais com tendências não lineares e padrões sazonais customizáveis. Além disso, a biblioteca permite adicionar variáveis regressoras, como o "Lucro Líquido Acumulado", para refinar as previsões.

No caso dos dados anuais, foi incorporada uma sazonalidade anual com ordem de Fourier[10] 15, capturando variações significativas ao longo do ano. Para os dados trimestrais, uma sazonalidade trimestral foi ajustada com ordem de Fourier 5, adequada para padrões de curto prazo observados no PPR. A simplicidade de uso do Prophet, aliada à

[10]Ordem Fourier: número de termos usados em uma série de Fourier para aproximar uma função periódica, determinando sua precisão.

sua flexibilidade, garantiu previsões precisas e interpretáveis para diferentes frequências temporais.

4.2.3. Random Forest Regressor

Já o Random Forest é a personificação da eficiência no aprendizado supervisionado. Ao combinar várias árvores de decisão, ele cria uma "floresta" que trabalha em conjunto para entregar previsões robustas. Seja prevendo preços, classificando dados ou analisando padrões complexos, o Random Forest é confiável, versátil e sempre pronto para entregar resultados precisos.

Foi aplicado para explorar padrões não lineares entre as variáveis do PPR. Este modelo foi configurado com lags[11] e sazonalidades como variáveis de entrada, o que permitiu capturar interações temporais e dependências complexas.

Para os dados anuais, o modelo foi configurado com 50 árvores (estimadores), profundidade máxima de 5 e lags de até 9 períodos. Já para os dados trimestrais, o modelo utilizou 200 árvores e profundidade máxima de 10, refletindo a necessidade de maior granularidade nas previsões trimestrais. As métricas de desempenho indicaram que o Random Forest capturou padrões não triviais nos dados, complementando as abordagens baseadas em séries temporais.

5. Experimentos e Resultados

Nesta seção, descrevem-se os experimentos realizados para avaliar a eficiência de diferentes modelos preditivos aplicados na previsão de despesas com o Programa de Participação nos Resultados (PPR). O processo experimental foi organizado em etapas que englobaram desde a previsão de indicadores financeiros gerais da empresa, como receitas, despesas e lucros, até o uso desses resultados como insumos para a previsão específica das despesas com PPR.

5.1. Descrição do Processo Experimental

O experimento foi dividido em duas etapas principais. Na primeira etapa, previram-se os indicadores financeiros gerais da empresa com base no histórico do Demonstrativo do Resultado do Exercício (DRE). Utilizou-se o modelo SARIMAX para capturar padrões sazonais anuais. O pré-processamento incluiu remoção de outliers, e interpolação linear para tratar valores ausentes. As previsões de "Receitas" e "Despesas" serviram de base para calcular o "Lucro Líquido" como a diferença entre receitas e despesas previstas. O "Lucro Líquido Acumulado" foi gerado ao somar os resultados mês a mês, e as informações foram armazenadas em um arquivo Excel para posterior utilização.

Na segunda etapa, realizou-se a previsão específica das despesas com PPR. Para isso, configuraram-se os modelos SARIMAX, Prophet e Random Forest, que utilizaram como variável exógena o "Lucro Líquido Acumulado" previsto na etapa anterior. O treinamento baseou-se nos dados históricos de PPR ajustados ao formato trimestral. Posteriormente, os modelos realizaram previsões futuras considerando as dinâmicas de lucro esperado, o que garantiu maior precisão ao capturar relações diretas entre lucros acumulados e gastos com PPR.

[11] Lags: atrasos aplicados a séries temporais para usar valores passados como variáveis explicativas.

5.2. Métricas de Avaliação

Para avaliar a performance dos modelos utilizados, aplicaram-se métricas amplamente aceitas na literatura. O Erro Absoluto Percentual Médio (MAPE) foi utilizado para medir a precisão percentual das previsões, enquanto o Erro Absoluto Médio (MAE) avaliou a magnitude média dos erros absolutos. A Raiz do Erro Quadrático Médio (RMSE) destacou desvios extremos, e o Coeficiente de Determinação (R^2) mensurou a proporção de variância explicada pelos modelos. Essas métricas permitiram analisar a eficiência dos modelos em granularidades temporais distintas, como dados trimestrais e anuais.

5.3. Resultados dos Modelos

Os resultados das previsões, avaliados com base nas métricas descritas, mostraram desempenho variado entre os modelos testados. Para os dados trimestrais, o modelo Random Forest apresentou os melhores resultados em termos de precisão, com o menor MAPE de 0,95% e MAE de 181,99. O Prophet mostrou flexibilidade em capturar padrões complexos, enquanto o SARIMAX apresentou consistência em dados com sazonalidade bem definida. Para dados anuais, o Random Forest manteve a maior precisão, atingindo um R^2 de 0,998, seguido pelo Prophet, que apresentou bom equilíbrio entre simplicidade e precisão.

As Tabelas 1 e 2 sintetizam os resultados obtidos pelos modelos para granularidades trimestrais e anuais, respectivamente.

Tabela 1. Resultados para as previsões trimestrais.

Modelo	**MAPE (%)**	**MAE**	**RMSE**	R^2
SARIMAX	14.83	587.34	816.65	0.928
Prophet	3.73	327.36	310.61	0.982
Random Forest	0.95	181.99	284.47	0.982

Tabela 2. Resultados para as previsões anuais.

Modelo	**MAPE (%)**	**MAE**	**RMSE**	R^2
SARIMAX	4.94	158.52	332.53	0.908
Prophet	3.34	219.88	295.67	0.989
Random Forest	0.61	48.38	95.77	0.998

Os gráficos 3 e 4 mostram a comparação entre os valores previstos e os valores reais das despesas com PPR, tanto para os dados trimestrais quanto para os anuais. A linha pontilhada nos gráficos representa a previsão perfeita, onde os valores previstos seriam exatamente iguais aos valores observados. A proximidade dos pontos em relação a essa linha reflete o desempenho de cada modelo.

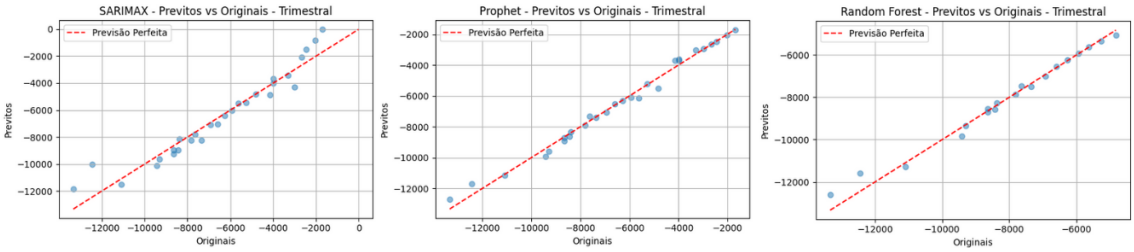

Figura 3. Valores históricos vs. valores previstos para dados trimestrais.

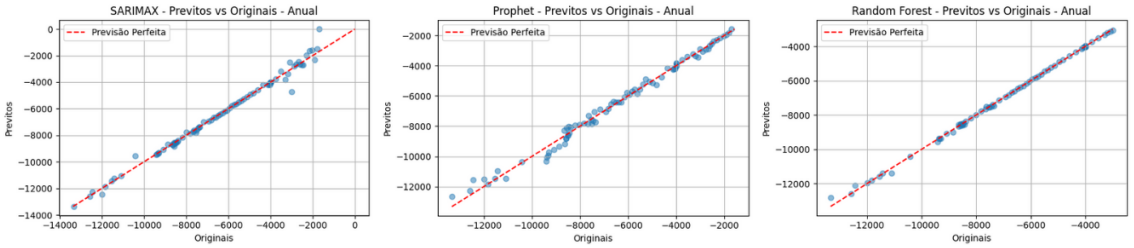

Figura 4. Valores históricos vs. valores previstos para dados anuais.

Para os dados trimestrais, os modelos apresentaram desempenhos variados. O SARIMAX demonstrou uma boa capacidade de previsão, mas com maior dispersão em pontos relacionados a valores extremos. O Prophet destacou-se por sua aderência aos valores reais, com a maioria dos pontos muito próximos à linha de previsão perfeita, indicando menor erro relativo em comparação aos outros modelos. Já o Random Forest apresentou um desempenho consistente, mas exibiu uma leve dispersão em valores mais elevados, sugerindo maior sensibilidade às flutuações dos dados trimestrais.

Nos dados anuais, os resultados foram ainda mais alinhados à previsão perfeita, refletindo a menor variabilidade dos dados agregados. O SARIMAX manteve um bom desempenho, com os pontos bem próximos à linha de previsão perfeita, especialmente para valores médios. O Prophet novamente apresentou alta precisão, com pontos bem distribuídos ao longo da linha, capturando adequadamente padrões sazonais e de tendência. O modelo Random Forest, apresentou melhor desempenho nesta granularidade, com pontos quase perfeitamente alinhados à linha de previsão, destacando sua robustez para previsões em granularidades maiores.

Os gráficos 5 e 6 apresentam a comparação entre previsões e valores reais para dados trimestrais e anuais, respectivamente. Essas visualizações ilustram a aderência dos modelos aos dados observados, destacando as tendências e sazonalidades capturadas.

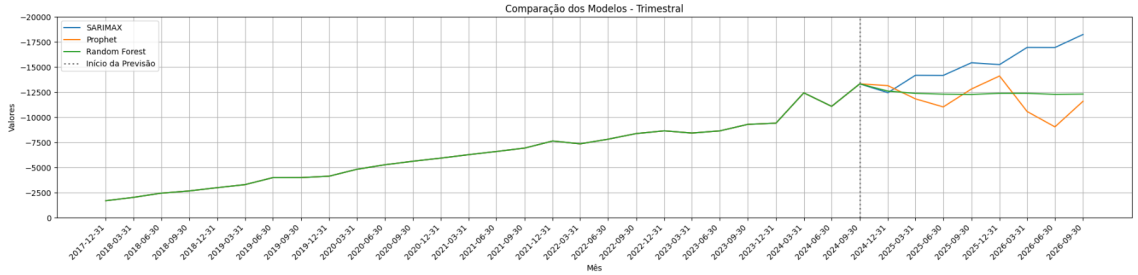

Figura 5. Valores históricos e valores previstos para dados trimestrais.

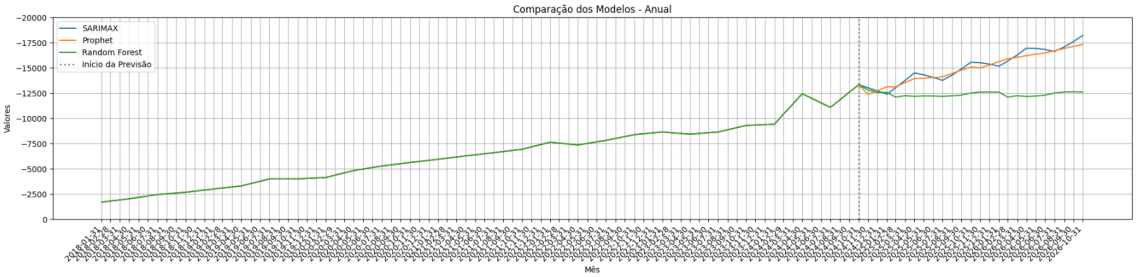

Figura 6. Valores históricos e valores previstos para dados anuais.

5.4. Discussão dos Resultados

Os resultados obtidos demonstram que cada modelo possui características específicas que influenciam seu desempenho em diferentes granularidades temporais. O modelo SARIMAX apresentou resultados consistentes em dados com padrões sazonais claros, mas não foi tão eficiente para capturar variações complexas nos dados trimestrais. O Prophet, por sua vez, mostrou alta flexibilidade, destacando-se principalmente em previsões trimestrais, graças à sua capacidade de ajustar sazonalidades e tendências não lineares.

O modelo Random Forest apresentou melhor desempenho geral, com excelente precisão tanto para dados trimestrais quanto anuais. No entanto, devido a precisão quase perfeita nos resultados de R^2, trouxe questões sobre possível overfitting[12], também, mostrou maior suscetibilidade ao superajuste, especialmente em granularidades menores, o que exigiu ajustes criteriosos de hiperparâmetros. Entre as principais limitações encontradas, destaca-se a volatilidade nos dados trimestrais, que dificultou a modelagem precisa em alguns casos, a dependência de interpolação linear no pré-processamento, que introduziu incertezas em padrões mais complexos e a limitação da quantidade observada de dados trazendo dúvidas sobre a compreensão dos modelos perante a variação dos valores em relação ao tempo.

Comparando os resultados com o estudo de [Ghosh 2024], nota-se que, apesar de o modelo de Ghosh ter apresentado uma precisão superior na previsão geral do mercado, com AUC[13] variando de 70% a 79%, os resultados obtidos no presente estudo mostram que modelos como SARIMAX e Prophet são eficazes no contexto específico de previsão de despesas de PPR, com uma margem de erro média ajustada às necessidades práticas da empresa estudada.

Por outro lado, no trabalho de [Gao & Huang 2021], nota-se que, embora ambos os trabalhos abordem previsões financeiras, o foco de Gao e Huang está em modelos de categorização, enquanto o foco do presente estudo se concentra em previsões temporais. Comparativamente, os resultados alcançados indicam um desempenho competitivo ao prever variações nas despesas com base em tendências históricas, enquanto o trabalho de Gao e Huang priorizou análises qualitativas de categorias de compensação, dificultando comparações diretas de acurácia.

[12]Overfitting: ocorre quando um modelo aprende detalhes excessivos dos dados de treino, prejudicando sua capacidade de generalizar para novos dados.

[13]AUC (Area Under the Curve): mede a capacidade de um modelo de classificação em distinguir entre classes, variando de 0.5 a 1.

Por fim, os resultados do estudo de [Elrefaei 2024] apresentam métricas de erro como RMSE e MAE, similares aos utilzados no presente estudo. Em comparação, os modelos utilizados apresentaram resultados equivalentes em termos de precisão para séries temporais univariadas, destacando a viabilidade da abordagem em cenários práticos de pequenas empresas. A inclusão de variáveis adicionais, como proposto por Elrefaei, poderia ampliar o escopo da análise e aumentar a robustez das previsões.

Essas observações indicam que a combinação de diferentes modelos pode ser uma estratégia promissora para otimizar previsões em cenários similares, aproveitando os pontos fortes de cada abordagem.

6. Conclusão e Trabalhos Futuros

Este estudo explorou a análise preditiva de despesas associadas ao Programa de Participação nos Resultados (PPR), utilizando o Demonstrativo do Resultado do Exercício (DRE) como base para a construção de modelos preditivos de Aprendizado de Máquina (AM). Com foco em uma pequena empresa, a análise abordou os desafios enfrentados por organizações em contextos de sazonalidade financeira e recursos limitados.

A aplicação dos modelos SARIMAX, Prophet e Random Forest revelou que o uso de AM pode fornecer previsões financeiras mais confiáveis e alinhadas às necessidades operacionais. O Random Forest destacou-se pelo desempenho geral superior em granularidades anuais, enquanto o Prophet e o SARIMAX demonstraram potencial em cenários com forte sazonalidade e padrões bem definidos. Esses resultados reforçam a relevância de ferramentas preditivas como suporte à gestão estratégica de empresas de pequeno porte, que buscam equilibrar incentivos ao desempenho com sustentabilidade financeira.

Este trabalho também enfatizou a importância do uso do DRE como instrumento central para análise financeira, corroborando sua aplicação como suporte para ajustes operacionais e estratégicos em períodos de incerteza econômica, como mencionado inicialmente. Os resultados indicam que modelos preditivos podem ser uma solução prática para minimizar desvios significativos nas despesas e aprimorar a previsibilidade em planejamentos de longo prazo.

Outras considerações para trabalhos futuros incluem expandir a análise para incluir dados multivariados que incluam outros indicadores financeiros e operacionais relevantes para a empresa, como produtividade dos funcionários e métrica de pesquisa de satisfação organizacional. O uso de técnicas mais avançadas, como redes neurais recorrentes ou transformadores, pode proporcionar insights adicionais nos padrões complexos observados nos dados. A utilização de datasets com quantidades superiores de registros, fortalecendo a interpretação nas variações. Além disso, o desenvolvimento de um sistema automatizado que integre esses modelos à funcionalidade da empresa pode facilitar a adoção e melhorar a gestão constante de despesas com PPR.

Siglas

R^2 Coeficiente de Determinação. 10, 12, 14

AM Aprendizado de Máquina. 2–5, 9, 15

DRE Demonstrativo do Resultado do Exercício. 1–3, 5, 9, 11, 15

IA Inteligência Artificial. 3

MAE Erro Absoluto Médio. 4, 9, 10, 12, 15
MAPE Erro Absoluto Percentual Médio. 9, 10, 12

PPR Programa de Participação nos Resultados. 1–3, 5–7, 9–12, 14, 15

RMSE Raiz do Erro Quadrático Médio. 4, 9, 10, 12, 15

SARIMAX Seasonal Auto-Regressive Integrated Moving Average with Exogenous Variables. 4, 9–15

Referências

Belisário, S. d. R. & Nikolay, S. (2022). A importância da demonstração de resultado do exercício para a tomada de decisão na gestão empresarial. *Artigo Científico do curso de MBA – Faculdades Integradas de Taquara (Faccat)*. Disponível em: http://www.faccat.br. Acesso em: novembro de 2024.

Brasil (1976). Lei nº 6.404, de 15 de dezembro de 1976. Dispõe sobre as sociedades por ações e dá outras providências. Diário Oficial da União, Brasília, DF, 17 dez. 1976. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l6404consol.htm. Acesso em: outubro de 2024.

Brasil (2000). Lei n.º 10.101, de 19 de dezembro de 2000. Dispõe sobre a participação nos lucros ou resultados e dá outras providências. Diário Oficial da União, Brasília, DF, 20 dez. 2000. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l10101.htm. Acesso em: outubro de 2024.

Elrefaei, I. (2024). Time series forecasting using machine learning. Dissertação de Mestrado, Brno University of Technology, Faculty of Electrical Engineering and Communication, Department of Telecommunications, Brno, República Tcheca. Disponível em: https://www.vut.cz/studenti/zav-prace/detail/161929. Acesso em: setembro de 2024.

Frizon, J. A. & Binsfeld, I. (2021). Programa de participação nos resultados e comprometimento organizacional: um estudo em uma cooperativa de crédito. *Revista de Gestão e Organizações Cooperativas*, 8(15):1–36. Publicado sob licença CC BY-NC-SA 4.0.

Gao, Y. & Huang, X. (2021). Forecast of compensation amount based on big data network and machine learning algorithm in intellectual property law. *Microprocessors and Microsystems*, 80:103623. Disponível em: https://doi.org/10.1016/j.micpro.2020.103623. Acesso em: setembro de 2024.

Ghosh, S. (2024). Forecasting firms' decision to change in capital structure. *SSRN Electronic Journal*. Disponível em: https://ssrn.com/abstract=4836746. Acesso em: setembro de 2024.

Jung, A. (2022). *Machine Learning: The Basics*. Springer. Disponível em: https://link.springer.com/book/10.1007/978-981-16-8193-6. Acesso em: novembro de 2024.

Leal, E. A., Soares, M. A., & Godoi-de Sousa, E. (2009). A influência de fatores contingenciais no sistema orçamentário nas empresas atacadistas: estudo de multi-casos. In *XVI Congresso Brasileiro de Custos*, Fortaleza, Ceará, Brasil.

Marion, J. C. (2019). *Contabilidade Introdutória*. Atlas, São Paulo, Brasil.

Motta, W. H. (2005). Remuneração variável: "stock options", valor econômico agregado, participação nos lucros ou resultados e benefícios flexíveis. In *Congresso Nacional de Excelência em Gestão*, Universidade Federal Fluminense, Niterói, Brasil. Disponível em: http://www.uff.br/congresso. Acesso em: novembro de 2024.

Ogunleye, J. (2014). The concepts of predictive analytics. *International Journal of Knowledge, Innovation and Entrepreneurship*, 2(2):82–90. Disponível em: https://www.ijkie.org/IJKIE_December2014_JAMES%20OGUNLEYE.pdf. Acesso em: novembro de 2024.

Saha, D., Young, T. M., & Thacker, J. (2023). Predicting firm performance and size using machine learning with a bayesian perspective. *Machine Learning with Applications*, 11:100453.

Salami, A. D. G., Silva, E. R., Tisott, P. B., Dias, D. T. A., & Bocchese, P. (2018). Gamificação: a proposição de um modelo para aplicação nas empresas. In *XVIII Mostra de Produção Universitária*, Universidade de Caxias do Sul, Caxias do Sul, Brasil.

Silva, G. M., Borges, R. F., & Moraes, J. P. M. (2007). A importância do planejamento estratégico para pequenas empresas. *Revista Administra-Ação*, 4.

Souza, A. M., Pereira, L. M., & Farias, R. T. (2005). Programa de participação nos lucros e resultados: impactos nas relações de trabalho. *Revista Brasileira de Gestão e Desenvolvimento Regional*, 1(1):71–83.

Zenker, R. (2010). A influência da estratégia na folga orçamentária mediada pela incerteza: um estudo de caso em uma empresa familiar. Master's thesis, Fundação Getulio Vargas, Escola Brasileira de Administração Pública e de Empresas, Rio de Janeiro, Brasil. Dissertação de Mestrado.