



UNIVERSIDADE FEDERAL DE PERNAMBUCO

DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO

CURSO DE GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

MATEUS VIEIRA TORRES

**Modelo de recomendação para alocação de recursos humanos em
projetos de engenharia utilizando Machine Learning**

Recife

2023

MATEUS VIEIRA TORRES

Modelo de recomendação para alocação de recursos humanos em projetos de engenharia utilizando Machine Learning

Trabalho de Conclusão de Curso apresentado à Coordenação do Curso de Graduação em Engenharia de Produção, da Universidade Federal de Pernambuco, como requisito para a obtenção de título de Bacharel em Engenharia de Produção.

Orientadora: Caroline Maria de Miranda Mota.

Recife

2023

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Torres, Mateus Vieira.

Modelo de recomendação para alocação de recursos humanos em projetos de engenharia utilizando machine learning / Mateus Vieira Torres. - Recife, 2023. 58, tab.

Orientador(a): Caroline Maria de Miranda Mota

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, Engenharia de Produção - Bacharelado, 2023.

1. Gestão de Projetos. 2. Alocação de Recursos Humanos. 3. Machine Learning. 4. Engenharia. I. Mota, Caroline Maria de Miranda. (Orientação). II. Título.

620 CDD (22.ed.)

MATEUS VIEIRA TORRES

Modelo de recomendação para alocação de recursos humanos em projetos de engenharia utilizando Machine Learning

Trabalho de Conclusão de Curso apresentado à Coordenação do Curso de Graduação em Engenharia de Produção, da Universidade Federal de Pernambuco, como requisito para a obtenção de título de Bacharel em Engenharia de Produção.

Aprovado em: 11/05/2023

BANCA EXAMINADORA

Profª. Dra. Caroline Maria de Miranda (Orientadora)
Universidade Federal de Pernambuco

Profª. Dra. Luciana Hazin Alencar (Examinador Interno)
Universidade Federal de Pernambuco

Profª. Dra. Ísis Didier Lins (Examinador Externo)
Universidade Federal de Pernambuco

RESUMO

Este trabalho de conclusão de curso se concentra no âmbito da gestão de projetos em empresas de consultoria em engenharia que possuem como centro do modelo de negócio a entrega de valor por meio de serviços ao cliente. O principal problema abordado é a alocação de recursos humanos, que muitas vezes é feita com base apenas no conhecimento empírico do gerente de projetos, aumentando as chances de tomada de decisão errada. Para lidar com esse desafio, este trabalho propõe o uso de modelos de *machine learning* para criar uma ferramenta capaz de auxiliar no processo de alocação de recursos humanos em projetos. Essa abordagem busca encontrar padrões e prever resultados mais precisos e pode ser uma alternativa aos modelos de otimização baseados em equações matemáticas, que exigem a definição antecipada de níveis de recurso por atividade. A utilização de modelos de *machine learning* permite uma abordagem diferente e pode ajudar a reduzir as incertezas do processo de tomada de decisão.

Palavras-Chave: Gestão de Projetos. Engenharia. Alocação de Recursos Humanos. *Machine Learning*.

ABSTRACT

This graduation project focuses on the field of project management in engineering consulting firms whose business model revolves around delivering value through customer services. The main problem addressed is the allocation of human resources, which is often based solely on the project manager's empirical knowledge, increasing the chances of making wrong decisions. To tackle this challenge, this study proposes the use of machine learning models to create a tool capable of assisting in the process of allocating human resources to projects. This approach seeks to find patterns and predict more accurate outcomes, and it can be an alternative to optimization models based on mathematical equations that require the upfront definition of resource levels per activity. The use of machine learning models allows for a different approach and can help reduce uncertainties in the decision-making process.

Keywords: Project Management. Engineering. Human Resource Allocation. Machine Learning

LISTA DE FIGURAS

Imagem 1 - Quantidade de atividades por Classificação

Imagem 2 - Resultados *RF*

Imagem 3 - Importância de variáveis através do *RF*

Imagem 4 - Resultados *Light GBM*

Imagem 5 - Importância de variáveis através do *Light GBM*

Imagem 6 - Resultados *GB*

Imagem 7 - Importância de variáveis através do *GB*

LISTA DE TABELAS

Tabela 1 – Instruções Gerais para a condução e avaliação da *Design Science Research*.

Tabela 2 – Características consideradas no modelo

SUMÁRIO

1. INTRODUÇÃO	9
1.1. DESCRIÇÃO DO CONTEXTO	12
1.2. JUSTIFICATIVA E RELEVÂNCIA	13
1.3. OBJETIVO	15
1.4. METODOLOGIA DE PESQUISA	15
1.5. ESTRUTURAÇÃO	18
2. BASE CONCEITUAL E REVISÃO DE LITERATURA	19
2.1. REFERENCIAL TEÓRICO	19
2.1.1. Gestão de Projetos	19
2.1.2. Gestão de Portfólio	21
2.1.3. Gerenciamento de Recursos	22
2.1.4. Inteligência Artificial	23
2.1.5. Machine Learning	24
- Modelos de Regressão	24
- Random Forest	25
- Light Gradient Boosting Machine (Light GBM)	26
2.2. REVISÃO DA LITERATURA	27
3. MODELO	32
3.1. BASE DE DADOS	32
3.2. IMPLEMENTAÇÃO	36
- Random Forest	41
- Light GBM	43
- Gradient Boosting	44
4. CONCLUSÃO	48
4.1. CONSIDERAÇÕES FINAIS	48
4.2. IMPLICAÇÕES PRÁTICAS	50
4.3. LIMITAÇÕES E FUTUROS TRABALHOS	51
REFERÊNCIAS	53

1. INTRODUÇÃO

Os projetos fazem parte significativa do tempo decorrido em entregas, decisões e resultados nas organizações atuais por exercerem papéis estratégicos na busca pela vantagem competitiva frente ao mercado. Além disso, os mesmos podem surgir por diversos motivos dentro das organizações. Segundo Kieling et al. (2019), projetos são serviços fornecidos para a resolução de uma necessidade específica; projetos de estratégia corporativa; elaboração de soluções mercadológicas inovadoras que permitam o desenvolvimento de um novo produto; ou mesmo de serviço demandado por um cliente específico.

Em concordância com as evoluções tecnológicas, as empresas enfrentam uma crescente demanda de projetos mais complexos com uma variabilidade de contextos, escopos, riscos e objetivos. Em cada um deles, as empresas buscam atingir seus propósitos de maneira a maximizar os benefícios. De acordo com O'brochta e Askenas (2003), em muitos casos, os projetos são grande parte do modelo de negócio da organização e por isso a atenção ao tomar decisões estratégicas tem tomado proporções maiores nos últimos anos.

Nesse contexto, Huselid (1995) aborda o fato de que o principal bem de qualquer empresa é o seu recurso humano, ou seja, a sua equipe de colaboradores. Principalmente em uma organização que tem como cerne do seu modelo de negócio a prestação de serviços por meio da elaboração de projetos. Uma empresa de engenharia consultiva pode ser um exemplo.

Com a finalidade de obter o sucesso organizacional, diversas ferramentas de Gestão de Projetos foram elaboradas ao longo dos anos para conseguir manter um equilíbrio entre os objetivos do projeto e as necessidades de seus stakeholders. Os fatores limitantes como tempo e custo são indispensáveis no estudo de novas soluções que permitam alcançar uma qualidade percebida de alto impacto, ou seja, a forma como um cliente enxerga o produto/serviço deve atingir e/ou superar suas expectativas.

Este estudo pretende aprofundar-se em um dos problemas enfrentados em organizações que possuem como o centro do seu modelo de negócio a entrega de valor por meio de projetos, isto é, empresas de consultoria. Por meio desses projetos, a empresa busca resolver as necessidades do seu cliente, desenvolvendo e/ou implementando soluções inovadoras que tenham alta qualidade.

A problemática abordada por este material enfoca no papel do gerente de projetos, que é responsável por prever quantos recursos humanos serão necessários, designar responsabilidades e gerenciar a alocação de recursos para garantir que o projeto seja concluído dentro do escopo, tempo e qualidade definidos.

No entanto, o principal gargalo do processo de planejamento e gerenciamento de projetos está nas previsões baseadas apenas no conhecimento empírico do gerente, o que aumenta a possibilidade de tomada de decisões erradas. Portanto, para buscar reduzir as incertezas do processo e fornecer melhores meios para a tomada de decisão neste quesito, a abordagem do presente trabalho pretende desenvolver uma ferramenta capaz de auxiliar nesse processo decisório.

Os meios tradicionais de resolução referentes às problemáticas de alocação de recursos humanos em projetos utilizam modelos de otimização baseados em equações matemáticas que buscam encontrar a solução ótima para um determinado objetivo do problema, geralmente objetivando redução de apenas uma perspectiva, custo ou tempo, considerando as restrições impostas.

Esses desafios são frequentemente solucionados através de uma problemática de programação matemática. Segundo Zaraket *et al.* (2013), programação inteira e programação não linear são geralmente construídos usando ferramentas como softwares de modelagem de otimização. Entretanto, para além do desafio inerente à busca pela solução ótima, essa modelagem exige a definição antecipada de níveis de recurso por atividade, tendo em vista a necessidade de pressupor a compreensão das relações entre o tempo de execução e o nível de recurso em cada tarefa do projeto.

Por outro lado, Carvalho *et al.* (2010), aborda aplicações de inteligência artificial com o objetivo de encontrar melhores resultados para determinados problemas a partir do uso de dados massivos. No material são apresentados os modelos de *machine learning* que são baseados em dados históricos e permitem aprender com os mesmos para encontrar padrões e prever o resultado mais provável. Eles são frequentemente usados em problemas de classificação, regressão e agrupamento, onde o objetivo é identificar um padrão ou tendência nos dados que possa ser usado para fazer previsões futuras.

Nesse sentido, a utilização de modelos de *machine learning* permite uma abordagem de maneira diferente das tradicionais pois busca conhecer o nível adequado de curso necessário em cada atividade tornando-a mais flexível e adaptativa, pois o modelo pode aprender a partir dos dados disponíveis, sem a necessidade de especificação prévia do problema. Isso possibilita a aplicação de modelos mais complexos, capazes de lidar com dados de diferentes naturezas e formatos.

Além disso, a utilização de modelos de *machine learning* permite a identificação de padrões e relações não triviais entre as variáveis, que muitas vezes passariam despercebidas em abordagens tradicionais de otimização de recursos. Dessa forma, o modelo pode aprender

o nível adequado de recurso necessário em cada atividade, aumentando a eficiência e eficácia do projeto.

É importante ressaltar que a aplicação de modelos de *machine learning* não substitui completamente os métodos tradicionais de otimização de recursos em projetos. Cada abordagem possui suas vantagens e limitações, e deve ser avaliada de acordo com as especificidades do problema em questão.

No entanto, a pesquisa busca apresentar que a utilização de modelos de *machine learning* representa um avanço na área de gestão de projetos, permitindo a criação de modelos mais flexíveis, precisos e adaptativos, capazes de lidar com dados complexos e oferecer soluções mais eficientes e eficazes para a classe de problema proposta.

Em termos de aplicação para problemas de alocação de recursos humanos em projetos, os modelos tradicionais de otimização podem ser mais apropriados quando o problema tem uma formulação matemática precisa e as restrições são bem definidas. Os modelos de *machine learning*, por outro lado, podem ser mais apropriados quando os dados históricos são ricos e podem ser usados para prever o futuro desempenho dos recursos humanos em diferentes cenários.

A abordagem utilizada neste estudo pode auxiliar no desenvolvimento de modelos de otimização, ao proporcionar maior compreensão do padrão de recursos necessários para conduzir diversas atividades em projetos. Dessa forma, os resultados obtidos podem ser valiosos para aprimorar o planejamento e execução de projetos.

Considerando a problemática em questão, a presente pesquisa objetiva integrar a gestão de projetos com a utilização de dados históricos com o intuito de aprimorar a tomada de decisões por parte dos gestores de projetos durante o processo de planejamento e gerenciamento dos recursos de projetos empresariais. De maneira mais simples, o modelo buscado pretende retornar um valor da alocação do(s) recurso(s), em forma de porcentagem, para a execução de determinada atividade dentro do projeto.

Para tanto, foi empregada a aplicação de modelos de aprendizado de máquina, visando avaliar a existência de padrões relacionados a características intrínsecas ao projeto, atividades e colaboradores, em comparação com alocações que resultaram em bons desempenhos e metas alcançadas.

Em caso de sucesso, seria possível desenvolver, com os resultados dessa pesquisa, um sistema capaz de gerar uma recomendação da alocação dos recursos humanos nas diferentes atividades a serem desenvolvidas em um projeto de pequeno, médio e grande porte.

1.1. DESCRIÇÃO DO CONTEXTO

A organização foco deste trabalho é uma empresa de grande porte atuante na consultoria brasileira para fornecer soluções em engenharia civil, passando pelo desenvolvimento de projetos, prestando consultoria a outras empresas e até realizando o gerenciamento, supervisão e fiscalização de obras de médio/grande porte. A área de atuação é bastante ampla e passa por áreas como saneamento, transportes, recursos hídricos, meio ambiente, mineração, desenvolvimento urbano e edificações.

Ao longo dos seus anos atuando no desenvolvimento da sua capacidade de entregar valor, a organização de engenharia consultiva sempre possuiu como um de seus valores a busca por melhores ferramentas e formas de realizar os serviços e identificar novas oportunidades de negócio.

Dessa forma, as empresas buscam elevar sua capacidade de gerar valor para seus clientes por meio do aprimoramento da resolução de problemas enfrentados nos seus processos. Esses problemas podem surgir tanto externamente, como novas demandas dos clientes, quanto internamente, como a simplificação de processos altamente burocráticos e desnecessários, principalmente em uma organização que tem como cerne do seu modelo de negócio a prestação de serviços por meio da elaboração de projetos. Uma empresa de engenharia consultiva é um exemplo disso.

Dentro desse cenário, a alocação dos recursos humanos realizada de forma equilibrada se torna imprescindível para garantir uma entrega de valor: com custo reduzido, em tempo adequado e com alta qualidade.

Na organização em questão, a área responsável pelo apoio no processo de gestão dos projetos da empresa é o setor da qualidade. Por meio de subáreas como: Inovação, Saúde e Segurança do Trabalho e Engenharia de Produção de Projetos, a equipe dá total suporte aos gerentes por meio da fiscalização, suporte a ferramentas, planejamento, monitoramento e controle dos projetos e trazendo inovações às práticas já existentes.

Diante disso, o presente trabalho de conclusão de curso pretende estudar um dos gargalos identificados no processo de gerenciamento do projeto pela equipe de qualidade em meio a um ambiente competitivo.

A perspectiva abordada nesta pesquisa está na utilização de modelos de *machine learning* para aprimorar o processo de alocação de recursos humanos em projetos. Em primeiro lugar, ao utilizar dados históricos, é possível identificar padrões e características relevantes que não são consideradas pelos métodos mais tradicionais. Isso permite uma

alocação mais precisa dos recursos humanos, considerando as habilidades individuais dos membros da equipe e as interdependências entre diferentes projetos.

Além disso, os modelos de *machine learning* permitem tratar de múltiplos projetos ao mesmo tempo, o que não é possível com os métodos mais tradicionais, que geralmente são aplicados para apenas um projeto complexo. Com isso, é possível obter uma alocação mais eficiente dos recursos humanos, levando em conta as necessidades e restrições de cada projeto.

A partir dessa abordagem é possível aprender com o processo de alocação e identificar as características mais relevantes para uma alocação eficiente dos recursos. Isso permite uma melhoria contínua do processo de alocação, com base em dados e evidências. O modelo desenvolvido neste trabalho utiliza dados históricos de projetos anteriores para identificar padrões e características relevantes dos projetos e dos membros da equipe.

O gerente do projeto é o responsável por realizar: a previsão de quantos homem-hora serão demandados, ou seja, quantas pessoas serão necessárias para o executar o projeto; a designação dos responsáveis por cada atividade; e, por fim, o nivelamento dos recursos deve ser realizado, ou seja, o gerenciamento da alocação realizada para que sejam atingidos os objetivos de escopo, tempo e qualidade, além disso, garantir que nenhum recurso esteja super alocado ou sub alocado.

O principal ponto de ineficiência no processo de planejamento e gerenciamento do projeto está relacionada ao fato de que, inicialmente, as previsões de quantas horas serão necessárias para finalizar as atividades e quem deve fazer cada etapa são feitas apenas utilizando o conhecimento empírico dos seus gerentes. Com isso, a probabilidade de ocorrer uma tomada de decisão errada é elevada, podendo elevar os custos do projeto, atrasar sua entrega acarretando em uma insatisfação do cliente e até gerar uma superalocação na equipe levando a crises na produtividade.

1.2. JUSTIFICATIVA E RELEVÂNCIA

Este trabalho de pesquisa surgiu a partir das demandas encontradas pelo presente pesquisador em um sistema de planejamento, monitoramento e controle dentro de um contexto de uma empresa de engenharia consultiva. Através desse ponto de vista, se viu a necessidade de ampliar o uso de tecnologias de dados nos processos de planejamento para apoiar esse processo de maneira a aperfeiçoar a alocação de recursos e reduzir as situações em que estimativas eram aplicadas utilizando somente a *expertise* dos gerentes, gerando desperdícios e/ou insatisfação do cliente.

Nesse mesmo sentido, o estudo desenvolvido por Pondel (2016) utilizou o princípio de que o uso de dados pode beneficiar significativamente a maneira como é realizado o gerenciamento de projetos nas principais empresas mundiais. Os autores confirmaram através da pesquisa que os projetos tornaram-se um conjunto muito complexo de atividades que exigem ferramentas mais sofisticadas para suportar a eficácia de todas as decisões.

Pondel (2016) aborda o fato de que cada projeto pode gerar uma quantidade enorme de dados, e desse modo, quando uma empresa for capaz de transformar dados em informações relevantes, pode-se gerar vantagem competitiva frente aos concorrentes além de maximizar a possibilidade de sucesso tanto do projeto quanto da organização.

Outro fator que pode justificar a relevância desta pesquisa é corroborado pelo estudo desenvolvido por Kieling *et al.* (2019), em que é abordada a ideia de que os avanços em modelos computacionais elevam a competência das empresas ao gerar melhorias as decisões durante os processos gerenciais a partir do uso de dados, entre eles, o uso de modelos de aprendizado de máquina para prever os valores adequados de alocação de recursos.

Ainda no contexto de aprendizado de máquina, Jianxiong (2009) concluiu através de um trabalho envolvendo um modelo de previsão de demanda que existem limitações na maioria dos modelos propostos na literatura, o que leva a baixa precisão dos métodos tradicionais. Para a solução desse problema, Jianxiong (2009) desenvolveu a ideia de que ao combinar modelos diferentes, mediante a técnica chamada de *ensemble*, pode-se melhorar significativamente a precisão dos resultados. Segundo o autor, métodos de previsão de demanda aplicados em recursos que utilizam apenas modelos isolados levam a resultados com baixa acurácia e ineficácia nas aplicações.

Outro aspecto importante de ressaltar é que o presente trabalho busca ampliar o espectro encontrado na literatura de modelos e ferramentas que apenas buscam a resolução da primeira pergunta: “Quanto cada recurso estará alocado a determinado projeto/atividade?” e não abordam o princípio de: “Quais recursos farão parte do projeto/atividade?”.

Portanto, esta pesquisa tem significância quanto a aplicação tanto à organização em que esta pesquisa está apoiada, quanto ao enriquecimento da literatura em questão do uso de dados em meio à Gestão de Projetos para solucionar questões relevantes de alocação de recursos no contexto de uma gama de projetos presentes em um portfólio.

1.3. OBJETIVO

A gestão de recursos dentro de um projeto tem uma notória influência no sucesso do mesmo, porém nem sempre é simples o processo de planejamento das horas demandadas nas mais diversas atividades de um projeto

Em decorrência disso, esta pesquisa tem como objetivo desenvolver um sistema capaz de avaliar se existe uma previsibilidade sobre o percentual de alocação que traz resultados positivos para recomendar o percentual que devem ser alocadas para cada atividade e separá-las entre os recursos disponíveis ao projeto de maneira a atender os objetivos organizacionais.

Objetivos específicos:

- Determinar quais atributos utilizar no modelo, visando compreender a alocação de recursos humanos com foco no aperfeiçoamento do algoritmo;
- Modelagem da solução por meio de diferentes técnicas de ML para obter o melhor desempenho;
- Encontrar os melhores parâmetros dos modelos para o melhor resultado;
- Avaliar a utilidade do modelo proposto para aplicação prática;

1.4. METODOLOGIA DE PESQUISA

Para construção metodológica desta pesquisa foi realizado um estudo que une características de natureza qualitativas e quantitativas de base documental, bibliográfica sobre métodos e ferramentas utilizados no contexto de Gestão de Projetos com o objetivo de modelar uma forma estruturada e precisa de realizar a alocação de recursos humanos em uma empresa de consultoria.

Para isso, foi adotado o método de pesquisa Dedutivo proposto por Hodson (1982), em que é conceituado de forma a realizar a integração de leis gerais com casos particulares. Ou seja, o método dedutivo se propõe a, de modo simplificado, conduzir um conhecimento de uma teoria ou lei geral a um evento particular em que são deduzidas certas conclusões. Segundo o autor, após a proposição dessas hipóteses é feito o teste experimental para a comprovação ou refutação das mesmas, por meio da comparação dos dados com as conclusões deduzidas anteriormente.

Esse trabalho de pesquisa, segue os conceitos de pesquisa aplicada, descritos por Thiollent (2009). De acordo com o autor, a pesquisa aplicada concentra-se em torno dos problemas presentes nas atividades de organizações, grupos ou atores sociais. Este modelo de pesquisa está empenhado na elaboração de diagnósticos, identificação de problemas e busca

de soluções e respondem a uma demanda formulada por clientes, atores sociais ou instituições.

Para a composição da base de dados da modelagem desta pesquisa, foi realizado um levantamento de informações por meio documentos dos mais variados projetos desenvolvidos pela organização assim como entrevistas com os gerentes de produto a fim de obter a visão dos mesmos sobre quais os fatores que devem ser considerados ao realizar uma boa alocação dos recursos e posteriormente apresentar os resultados da pesquisa para coletar a viabilidade de aplicação do modelo escolhido no dia a dia destes gerentes.

Com o objetivo de estruturar a base teórica e prática deste trabalho, a de técnica de pesquisa utilizada como inspiração foi o *Design Science Research* (DSR) por se tratar de uma técnica focada na idealização de um artefato objetivando a entrega de valor para aprimorar um sistema. Nesse sentido, Çagdas e Stubkjær (2011) afirmaram que a DSR se constitui em um processo rigoroso de projetar artefatos para resolver problemas, avaliar o que foi projetado ou o que está funcionando e comunicar os resultados obtidos.

A DSR é uma metodologia de pesquisa que se baseia em conceitos advindos da *Design Science* (DS), ou Ciência do Projeto descrita por Simon (1996). De acordo com Lacerda *et al* (2013), a DS seria responsável por conceber e validar sistemas que ainda não existem, seja criando, recombinao, alterando produtos/processos/software/métodos para melhorar as situações existentes.

A respeito do nível de conhecimento gerado pela DSR, Van Aken (2004) define que o conhecimento desenvolvido na DS é generalizável quando válido para uma dada classe de casos. No trabalho de Lacerda *et al* (2013) cada classe de casos são entendidos como uma classe de problemas. No caso do presente trabalho a classe de problemas envolvida está no âmbito da gestão de recursos humanos em empresas de consultoria.

Para a resolução dos problemas dentro da classe descrita, foram propostos artefatos que por sua vez foram conceituados por Simon (1996) como objetos artificiais que podem ser caracterizados em termos de objetivos, funções e adaptações. Em síntese, um artefato pode ser caracterizado por possíveis soluções propostas pelo cientista a fim de integrar o propósito, a solução e o contexto em que ele está inserido.

Os artefatos tratados por esta pesquisa podem ser caracterizados como modelos, de acordo com a classificação de March e Smith (1995). Em decorrência do objetivo desta pesquisa de encontrar o algoritmo de ML que recomenda a melhor utilização dos recursos humanos para os gerentes de projetos, os artefatos deste trabalho de conclusão de curso foram os modelos pré selecionados para a aplicação e avaliação da acurácia e usabilidade no

contexto em questão, garantindo a fortalecer do sucesso do projeto e conseqüentemente, da superação das expectativas dos clientes.

De acordo com March e Smith (1995), na DSR, a preocupação é a utilidade de modelos, não a aderência de sua representação à verdade, deste modo, além da precisão e acurácia das recomendações buscadas por esta pesquisa busca-se também um modelo que atenda a padrões de usabilidade pelos interessados. Por outro lado, embora tenda a ser impreciso sobre detalhes, o modelo objetivo deste trabalho precisa sempre capturar a estrutura da realidade para ser uma representação útil ao pesquisador e a quem a solução é fornecida.

De acordo com o critério de avaliação dos artefatos propostos pela pesquisa, Worren, Moore, Elliott (2002) propuseram uma sequência de instruções descritas pela Tabela 1.

Tabela 1 - Instruções Gerais para a condução e avaliação da Design Science Research.

Passo	Item	Descrição
1	<i>Design</i> como artefato	Projetar um artefato viável, seguindo os princípios e fundamentos do DS. Esse artefato deve ser no formato de: constructo, modelo, método e/ou instanciação.
2	Relevância do problema	Tem como objetivo desenvolver uma solução de origem tecnológica para resolver um problema gerencial do sistema em questão.
3	Avaliação do design	Aplicação de métodos de avaliação que demonstrem, de maneira rigorosa, a utilidade, qualidade e eficácia do artefato para a solução do problema.
4	Contribuições do design	Apresentar fundamentação clara, por meio de metodologias de <i>design</i> , que comprovem de forma clara as contribuições e devem ser verificáveis dentro do contexto da classe do problema.
5	Rigor da pesquisa	Aplicação de métodos rigorosos tanto na construção, quanto na avaliação dos artefatos.
6	<i>Design</i> como Processo de Pesquisa	Utilização dos meios disponíveis para alcançar os objetivos esperados ao mesmo tempo que busca um artefato eficaz e

		efetivo e satisfaz a leis e teorias que regem o contexto em questão.
7	Comunicação da Pesquisa	Apresentação da pesquisa tanto para os especialistas, quanto para a área de gestão.

Fonte: Adaptado de Lacerda (2013)

1.5. ESTRUTURAÇÃO

O presente trabalho seguirá uma estrutura de capítulos da seguinte forma: no segundo capítulo será dada toda a base conceitual necessária para alinhar significados e expectativas sobre os termos e conceitos aqui apresentados e trabalhados, assim como uma breve apresentação sobre o que foi encontrado na literatura envolvendo o problema proposto por este trabalho.

Após essa etapa inicial, no capítulo 3 será apresentado o modelo utilizado para a proposta de solução e seus respectivos detalhamentos, como os dados de entrada utilizados nos modelos, as premissas estipuladas pelo modelo, o processo de construção e aprimoramento dos modelos e como foi o processo de avaliação para escolha do melhor modelo capaz de recomendar a solução ao decisor.

Por fim, no último capítulo serão apresentados os resultados obtidos e uma breve discussão sobre como esse resultados podem de fato auxiliar a decisão do gerente e da equipe de planejamento a realizar a alocação estratégica de recursos.

2. BASE CONCEITUAL E REVISÃO DE LITERATURA

Nesta etapa, serão abordados os principais termos utilizados ao longo do trabalho visando nivelar os conceitos. Para tanto, foi realizada uma pesquisa bibliográfica abrangente, a fim de identificar as principais fontes que abordam o tema em questão. A partir dessa pesquisa, foram selecionados textos relevantes e atualizados, que fornecerão a base teórica necessária para a compreensão do assunto. Esses textos foram lidos com atenção, e deles foram extraídos os principais conceitos e definições de forma clara e objetiva no trabalho.

Além disso, a revisão de literatura permitirá uma análise mais detalhada do que já foi escrito sobre o tema em questão. Isso envolverá a identificação das principais teorias e modelos explicativos utilizados na área de estudo, bem como a análise das principais pesquisas empíricas que foram conduzidas sobre o tema.

Dessa forma, a base conceitual e revisão de literatura permitiram que o autor do trabalho tivesse uma compreensão mais ampla e aprofundada do assunto em questão, permitindo uma análise mais precisa dos dados e uma discussão mais fundamentada dos resultados obtidos. Além disso, essa etapa é fundamental para que a pesquisa seja situada no contexto mais amplo da área de estudo, permitindo que sejam identificadas as principais lacunas na pesquisa e oportunidades para futuras investigações.

2.1. REFERENCIAL TEÓRICO

2.1.1. Gestão de Projetos

Segundo *Project Management Institute* (PMI, 2021), um projeto é todo esforço temporário empreendido para criar um produto, serviço ou resultado específico, portanto seus resultados passam pelas mãos de uma equipe qualificada para gerar o valor esperado e dessa forma atingir seu objetivo. Ao atingir o objetivo estipulado, o PMI (2021) estipula que um projeto pode gerar entregas de alguns tipos:

- Desenvolvimento, aprimoramento ou correção de um produto único;
- Um serviço específico;
- Um resultado único
- Uma combinação única de um ou mais produtos, serviços ou resultados.

Durante o processo de desenvolvimento de qualquer projeto, diversas entidades podem ser afetadas diretamente ou indiretamente, essas entidades que estão interessadas no

sucesso do projeto são chamadas de partes interessadas ou *Stakeholders*. Segundo o PMI (2021), um stakeholder é um indivíduo, grupo ou organização que possa afetar, ser afetado, ou sentir-se afetado por uma decisão, atividade ou resultado de um projeto. O desenvolvimento de um projeto pode envolver uma única pessoa ou uma equipe de desenvolvimento, uma única organização ou um conjunto de organizações em prol de uma solução.

De acordo com o conceito de projeto proposto pelo PMI (2021), pode-se perceber um conceito que norteia a vida de qualquer profissional envolvido com projeto, um aspecto indispensável contemplado pela expressão “esforço temporário”. Portanto, o fator limitante do tempo deve ser considerado a todo momento ao realizar o gerenciamento dos projetos. Além desse ponto, aliado ao tempo, recursos financeiros e o escopo do projeto se tornam aspectos indispensáveis em qualquer modelo de gestão de projetos.

Seja em uma consultoria, que tem como principal objetivo propor soluções das necessidades dos seus clientes por meio de projetos, ou em uma empresa qualquer, os projetos são parte fundamental para o sucesso de qualquer organização. Para isso, Gray e Larson (2009) apresentam que a ideia de que a qualidade e o sucesso final de um projeto são avaliados através da verificação do atendimento ou superação das expectativas do cliente e/ou da alta direção em relação a custo (orçamento), tempo (planejamento) e desempenho (escopo) do projeto. Somente pela união desses três pontos é possível adquirir a qualidade de um projeto.

Porém, Gray e Larson (2009) fazem uma ressalva no que diz respeito à inter-relação entre essas três variáveis. Os autores comentam que determinadas situações forçam e/ou demandam uma priorização de um dos pilares em detrimento de outro, causando um conflito de interesses que deve ser gerenciado pelo Gerente do Projeto.

Segundo Gray e Larson (2009), uma das principais atribuições de um gerente ao longo do projeto é a capacidade de administrar o conflito entre as três variáveis.

Segundo o Kieling *et al.* (2019), na busca por melhores soluções ao longo de projetos, os principais estudos na área de gestão de projetos buscam atingir 3 objetivos estratégicos, são eles:

- redução de custos;
- minimização do tempo;
- atingir a qualidade desejada.

A arte de obter o sucesso em um projeto está diretamente relacionada com a capacidade de tomar boas decisões gerenciais, segundo Pondel (2016), essas decisões gerenciais dizem

respeito a: elaboração de cronograma, planejamento e nivelamento de recursos, gerenciamento de riscos, gestão de aquisições, entre outros.

Há um aspecto intrínseco no estudo proposto por Kieling *et al.* (2019) que tem como princípio a redução de custos e a maximização da utilização dos recursos. Uma boa alocação de recursos permite a mitigação das despesas atreladas à mão de obra, pois permite evitar o desgaste dos colaboradores, uma previsão escassa ou excessiva por parte do gerente do projeto.

A arte de alocar os recursos de maneira adequada é chamada de nivelamento de recursos. De acordo com Rozenfeld *et al.* (2006), a relevância do nivelamento de recursos se deve à necessidade de a empresa conseguir, com recursos escassos, sejam eles humanos e/ou tecnológicos, aproveitar as oportunidades de projetos que lhe confirmam maior vantagem, de acordo com a estratégia competitiva adotada pela organização .

2.1.2. Gestão de Portfólio

A obtenção de sucesso nos três objetivos estratégicos ao longo de uma cadeia de projetos dentro de uma organização depende do processo de gestão de portfólio de projetos bem estruturado. Ao pensar em como estruturar as tarefas que alavancarão seus resultados, os maiores especialistas da área empregaram muitas horas de estudo nos últimos anos.

Segundo o PMI (2021), portfólio é o conjunto de projetos, sub portfólios e/ou programas que são operados em conjunto com o foco no alcance do objetivo estratégico da empresa.

De acordo com Pondel (2016), o processo de Gerenciamento de Portfólio é mais do que executar e gerenciar vários projetos em conjunto, é a necessidade de cada portfólio de projetos precisar ser avaliado por seu valor de negócios e sua aderência à estratégia organizacional. O portfólio deve ser projetado para atingir um objetivo ou benefício de negócios bem definido.

Buttrick (2009) fez um resumo sobre a diferença entre a gestão de projetos e a gestão de portfólios. Segundo o mesmo, dirigir um único projeto corretamente irá garantir que se está no caminho certo. Porém, somente ao dirigir “todos os projetos” com sucesso garantirá que se está fazendo o certo a gestão do portfólio.

Segundo Cooper, Edgett e Kleinschmidt (1999), o conceito de gestão de portfólio pode ser sintetizado como um processo de decisão dinâmico, por meio do qual decisões são constantemente tomadas e constantemente atualizadas e revisadas. Segundo esses autores, gestão de portfólio é a manifestação da estratégia do negócio, que estabelece onde e como o

investimento será feito no futuro. Ademais, as múltiplas tomadas de decisão podem ser caracterizadas pela incerteza e mudanças das informações, oportunidades voláteis, múltiplos objetivos e critérios estratégicos, além de contar com a interdependência entre projetos.

De acordo com Heising (2012), o objetivo final da Gestão do Portfólio de Projetos, ou *Portfolio Project Management (PPM)*, é maximizar a contribuição de projetos para o sucesso corporativo. Assim, o PPM pode ser considerado como a gestão simultânea do conjunto de projetos que compõem uma estratégia de investimento de uma empresa .

Segundo Gray e Larson (2009), os benefícios do gerenciamento de portfólio são:

- Disciplina para o processo de seleção de projetos.
- Integra a gestão de projeto aos indicadores estratégicos.
- Priorização das propostas de projeto de acordo com um conjunto de métricas e não com bases empíricas e emocionais..
- Alocação de recursos alinhados com os objetivos estratégicos.
- Equilíbrio dos riscos entre todos os projetos.
- Eliminação de projetos que não maximizem os ganhos da organização.
- Aprimoramento da comunicação
- Suporte ao entendimento das metas do projeto.

2.1.3. Gerenciamento de Recursos

No que se refere à implementação da gestão de portfólio, Brown e Eisenhardt (1998) comentam que a tomada de decisões em âmbitos complexos não é uma tarefa trivial; pois engloba incertezas de mercado e tecnológicas; negociação e alocações de recursos, na maioria das vezes escassos, entre diferentes projetos e há constantes mudanças em decorrência à competitividade do mercado e aos aspectos intrínsecos de cada ambiente.

Gray e Larson (2009), levantam o questionamento e o estudo sobre a importância do gerenciamento de recursos no contexto de projetos. Os autores comentam que em contextos de empresas de projetos há diversas organizações que lidam em ambiente de multiprojetos, em que há a necessidade de analisar a interdependência de projetos e problemas de compartilhamento de recursos.

Desse modo, Gray e Larson (2009) comentam que os problemas de alocação e planejamento de recursos entre projetos e diferentes atividades por meio de projetos crescem exponencialmente quando o seu número aumenta. Os autores ainda enaltecem que, em ambientes de multiprojetos, as apostas, ou seja, o planejamento realizado apenas por expertise e sem o suporte necessário são mais altas e os benefícios ou penalidades relativas ao bom ou

mau planejamento se tornam ainda mais significantes do que em casos de muitos projetos simples.

Gray e Larson (2009) fazem uma relação entre o alto número de projetos e a quantidade de recursos disponíveis, os autores analisam que a sobrecarga da capacidade na maioria das vezes a utilização de recursos de maneira inadequada. A presença de uma distância entre a implementação e a execução, de políticas de hierarquia e de diversas tarefas sendo realizadas ao mesmo tempo é adicionada ao problema sobre quais recursos serão alocados primeiro e qual deve ser essa alocação.

Os autores ainda enfatizam que a moral e a confiança dos funcionários sofrem porque é necessário que em um ambiente multiprojeto os integrantes da equipe enfrentam problemas graves sem um sistema de prioridades que esteja claramente conectado ao plano estratégico organizacional.

2.1.4. Inteligência Artificial

Na busca por alocar os colaboradores de maneira mais eficiente, surgiram modelos computacionais capazes de aprender com alocações bem sucedidas anteriormente dentro da organização, esses modelos utilizam dados passados para auxiliar na tomada de decisão no nivelamento dos recursos. Os modelos estudados utilizam princípios e vertentes da Inteligência Artificial (IA).

Uma das vertentes da IA é o chamado aprendizado de máquina, ou *Machine Learning* (ML), que é caracterizado dessa maneira pois permite que um programa de computador seja capaz de inferir regras que elevam a eficácia das decisões e que possibilitem atingir os objetivos do modelo construído a partir de experiências anteriores em eventos similares. O aprendizado de máquina foi uma mudança de paradigma da "programação tradicional", em que as instruções devem ser explicitamente fornecidas ao computador, para uma "programação indireta", na qual o computador aprende as "regras ocultas" nos dados fornecidos e fornece uma resposta que melhora as métricas utilizadas na construção desse modelo.

De acordo com Pondel (2016), o uso de ferramentas avançadas que suportam o gerenciamento de projetos baseadas em inteligência artificial tem um papel essencial na gestão de um negócio, particularmente na área de uma Gerenciamento de Projetos.

2.1.5. Machine Learning

Aprendizado de Máquina é uma técnica de inteligência artificial que permite que os computadores aprendam dados sem serem programados explicitamente. A sua utilização tem se mostrado extremamente eficiente no contexto de alocação de recursos, trazendo uma série de benefícios para as empresas.

Um dos maiores benefícios é a melhoria de processos. O aprendizado de máquina permite que os sistemas automatizam a análise de dados e identifiquem padrões de consumo de recursos, tornando possível a alocação mais precisa e eficiente dos mesmos. Além disso, a automação dos processos de alocação reduz a margem de erro humano, tornando o processo mais confiável e consistente.

Outro aspecto importante é a capacidade de antecipação de demandas. O aprendizado de máquina permite que os sistemas priorizem as necessidades futuras de recursos com base em dados históricos e outras informações relevantes. Isso torna possível a alocação proativa de recursos, evitando problemas de falta ou excesso de capacidade.

Ainda, o aprendizado de máquina traz uma série de benefícios em termos de custo. A automação dos processos de alocação reduz os gastos com mão de obra, já que muitos desses processos são realizados por profissionais altamente especializados. Além disso, a melhoria dos processos e a antecipação de demandas permitem que as empresas utilizem de forma mais eficiente os seus recursos, evitando desperdícios e reduzindo os custos.

Em resumo, o aprendizado de máquina tem se mostrado uma ferramenta extremamente eficiente no contexto de alocação de recursos, trazendo benefícios como a melhoria de processos, antecipação de demandas e redução de custos. Para se ter uma ideia da importância dessa tecnologia, basta citar algumas das empresas que a utilizam, como a Amazon, Google e Microsoft (Witten et al., 2016).

- Modelos de Regressão

A regressão é uma técnica de aprendizado de máquina que tem como objetivo estimar a relação entre uma variável dependente e uma ou mais variáveis independentes. É um dos modelos mais utilizados em problemas de previsão e modelagem de dados, e sua aplicação é fundamental em diversas áreas do conhecimento, como finanças, economia, ciência de dados, engenharia e medicina.

Na área financeira, por exemplo, a regressão é utilizada para prever o preço de ações e outros ativos financeiros, possibilitando que investidores tomem decisões informadas. Já na

medicina, a regressão é utilizada para prever a probabilidade de uma pessoa desenvolver uma doença com base em suas características pessoais, permitindo que profissionais de saúde identifiquem precocemente pacientes que precisam de cuidados especiais.

Existem diversos tipos de regressão, como a regressão linear simples e múltipla, regressão logística e regressão de Poisson, cada uma com suas particularidades e aplicações específicas. Em geral, o objetivo da regressão é encontrar uma equação matemática que descreva a relação entre as variáveis, permitindo que sejam feitas previsões com base em dados históricos.

Apesar de ser uma técnica poderosa, a regressão tem suas limitações. Em muitos casos, as relações entre as variáveis podem ser não lineares e a regressão não é capaz de capturar essas relações. Além disso, a regressão pode ser sensível a outliers e pode levar a previsões imprecisas se houver muitos outliers nos dados. É importante ressaltar, porém, que essas limitações não invalidam o uso da regressão, mas sim destacam a importância de escolher o método adequado para cada problema e de interpretar corretamente os resultados obtidos.

- ***Random Forest***

O *Random Forest* é um algoritmo de aprendizado de máquina amplamente utilizado em diversas aplicações de classificação e regressão. O algoritmo foi introduzido por Breiman (2001) e consiste em uma coleção de árvores de decisão que são construídas utilizando amostras aleatórias do conjunto de treinamento. Cada árvore na floresta é construída de forma independente, e as amostras são selecionadas aleatoriamente com substituição. Esse processo de amostragem aleatória é chamado de bootstrap (Breiman, 1996).

Uma das principais vantagens do algoritmo *Random Forest* é a aleatoriedade de características que é utilizada durante a construção de cada árvore individual. Isso resulta em uma floresta de árvores não correlacionadas, o que leva a um resultado de classificação melhor do que o obtido por qualquer árvore individual (Breiman, 2001). Para aumentar ainda mais a aleatoriedade do problema, um subconjunto de características é considerado para divisão, e as operações são realizadas em cada nó da árvore.

Durante o processamento, as árvores de decisão individuais podem apresentar alta variância, o que pode levar ao ajuste excessivo do estimador da árvore. Para resolver esse problema, é possível fornecer dois tipos de aleatoriedade. As amostras de subconjunto

aleatório resultam em diferentes erros, mas esses erros são cancelados ao prever a classe com a maioria dos votos (Breiman, 2001).

Além disso, é importante lembrar que, embora as florestas aleatórias possam resultar em um aumento do viés, é a variância que precisa ser abordada primeiro, em vez do viés. Com isso em mente, é possível obter resultados ainda mais precisos e confiáveis com o algoritmo *Random Forest* (Breiman, 2001).

- ***Light Gradient Boosting Machine (Light GBM)***

O algoritmo *Light Gradient Boosting Machine* é uma técnica de aprendizado de máquina que se destaca em diversas aplicações, como classificação, regressão e ranking. Foi proposto pela Microsoft em 2017 como uma alternativa ao *Gradient Boosting Decision Tree* (GBDT), um dos modelos mais utilizados em problemas de aprendizado de máquina.

De acordo com KE et al. (2017), *Light GBM* utiliza uma técnica de gradiente descendente para treinar árvores de decisão. Em vez de construir as árvores de decisão de forma sequencial, como no GBDT, o *Light GBM* utiliza uma estratégia de construção baseada em histogramas. Essa abordagem torna o algoritmo muito rápido e escalável, mesmo em grandes conjuntos de dados

Outra característica apresentada por Wang et al. (2017) deste algoritmo é a utilização de uma estratégia de amostragem por nó, selecionando um subconjunto aleatório de recursos e amostras para cada nó da árvore de decisão. Isso ajuda a reduzir o overfitting e melhora o desempenho do modelo.

Outra vantagem do modelo em questão é a sua capacidade de lidar com dados desbalanceados. O algoritmo possui uma função de perda balanceada que ajusta os pesos das classes para compensar a desigualdade na distribuição dos dados. Isso ajuda a evitar que o modelo se concentre apenas na classe majoritária e melhora o desempenho do modelo para a classe minoritária.

É um algoritmo de aprendizado de máquina relativamente novo, mas já possui uma comunidade de usuários e desenvolvedores ativa. Existem diversas bibliotecas disponíveis em diferentes linguagens de programação, como Python, R e C++. Além disso, essa técnica tem sido utilizado em diversas competições de ciência de dados e aprendizado de máquina, onde obteve excelentes resultados.

2.2. REVISÃO DA LITERATURA

Para identificação das melhores práticas utilizadas na literatura e os principais estudos associados ao problema em questão, foi realizada uma busca através do Portal de periódicos da CAPES e por meio do *Web Of Science* foram realizadas as pesquisas.

Durante o processo de estudo inicial, procurou-se procurar por quaisquer tipos de métodos e ferramentas para alocação de recursos humanos em projetos com foco na identificação das principais lacunas desses meios a fim de solucionar os problemas. Nessa etapa as palavras-chave são: "Métodos/Ferramentas", "Alocação de recursos", "Alocação de recursos humanos", "Alocação de recursos humanos em portfólio de projetos".

Em seguida, buscou-se entender a quantidade e qualidade dos estudos que aplicam *machine learning* a resolução desta problemática para aprofundar aspectos positivos e melhorar a abrangência desses estudos nos aspectos não trabalhados mais a fundo. A partir disso, as principais palavras-chave foram: "*Machine Learning* aplicado à gestão de projetos", "Alocação de recursos com *Machine Learning*", "Gestão de recursos humanos utilizando *machine learning*"

A tomada de decisão em ambientes de alta complexidade e variabilidade, faz com que muitas das vezes as decisões dos Gerentes de Projetos das grandes organizações estejam em constante evidência e tenham impacto significativo no desenrolar dos projetos. Na empresa em questão, há uma busca por como resolver os problemas latentes na gestão dos projetos de maneira estruturada e que permitisse sua melhoria contínua.

Um aspecto relevante a ser ressaltado é que, em sua maioria, os meios mais estudados e apresentados na literatura para lidar com o dilema da alocação de recursos em projetos não contemplam o emprego de modelos de Aprendizado de Máquina. De maneira geral, as pesquisas realizadas objetivam solucionar unicamente a primeira categoria de problema, ou seja, identificar se um determinado recurso será ou não alocado ao projeto, sem abordar o cerne da presente pesquisa, que consiste em desenvolver um modelo capaz de estimar o percentual do tempo que um recurso específico será alocado para uma atividade específica do cronograma.

Ao examinar minuciosamente os estudos mais relevantes, é possível constatar, por exemplo, que técnicas de programação linear são amplamente empregadas para solucionar apenas o primeiro tipo de problema. Zaraket *et al.* (2013) criaram um modelo baseado em programação linear, focado na recomendação para direcionar decisões do primeiro problema e objetivando a redução de custos.

De acordo com Barbosa e Carvalho (2017) outras ferramentas utilizadas se destacam são: a matriz de alocação, o modelo de capacidade, a gestão de competências, o modelo de risco e a ferramenta de gestão de tempo. A matriz de alocação ajuda a visualizar a disponibilidade de recursos e alocá-los em diferentes projetos, evitando conflitos.

O modelo de capacidade permite gerenciar a alocação de recursos humanos em projetos, determinando a capacidade total disponível e distribuindo-os conforme necessário. A gestão de competências concentra-se no desenvolvimento e gerenciamento das habilidades e conhecimentos dos colaboradores. O modelo de risco considera os possíveis riscos da alocação de recursos em projetos e fornece um plano de contingência. A ferramenta de gestão de tempo monitora o tempo gasto pelos colaboradores em diferentes tarefas e projetos, permitindo que os gerentes de projeto façam ajustes se necessário.

O artigo elaborado por Kieling *et al.* (2019), tem como principal objetivo desenvolver um modelo computacional que seja capaz de elevar a capacidade que os gestores têm de tomar decisões positivas e que sigam princípios estabelecidos por eles na execução do projeto. Com isso, o modelo proposto tem duas contribuições científicas principais, baseadas no estudo de trabalhos relacionados: permitir a escolha do objetivo estratégico da organização e usar as alocações históricas de projetos anteriores.

No estudo desenvolvido por Kieling *et al.* (2019), é abordada a ideia de que os avanços em modelos computacionais propiciam a capacidade de melhorar as decisões durante os processos gerenciais a partir do uso de dados. Dessa forma, a construção do modelo *Smartalloc* dá a oportunidade de o tomador de decisão escolher qual das métricas deseja priorizar (custo, tempo ou qualidade) na definição da estratégia de alocação de recursos e receber uma sugestão de decisão para o seu planejamento com base no aprendizado de máquina. A ferramenta utiliza conceitos de Big Data para manter os dados históricos de projetos anteriores armazenados em um banco de dados e aplica um algoritmo de recomendação capaz de realizar a predição das alocações dos recursos humanos em um novo projeto.

O principal elemento do modelo de recomendação proposto por Kieling *et al.* (2019), é o algoritmo de classificação denominado k-NN. Este algoritmo é o responsável por realizar a predição de alocações de recursos humanos para as tarefas do novo projeto.

Kieling *et al.* (2019) também conduziu uma análise qualitativa, no formato de entrevistas, sobre o perfil das empresas com relação à rotina dos gestores de projetos durante o processo de gerenciamento dos recursos que confirmou a dificuldade prática de fazer alocações de recursos humanos. O gestor sofre pressões sobre o controle de custos e não

existem meios bem difundidos para se realizar a alocação de maneira sistematizada, além de precisar lidar com a pressão do gerenciamento de diversos projetos simultaneamente.

Como resultado da aplicação do modelo, o trabalho de Kieling *et al.* (2019) apresentou níveis de acurácia superiores a 70%, o que foi julgado pelos gerentes de projetos como de grande valia na etapa de planejamento.

De acordo com as entrevistas, foi possível concluir que há evidências suficientes de que as equipes de planejamento e, conseqüentemente, os gestores de projetos estão lidando diariamente com a pressão sobre custos, tempo e qualidade e estão sem subsídios que os apoiem nesse aspecto. Ainda como resultado das entrevistas, a pesquisa qualitativa concluiu que os gestores pesquisados trabalham tendo como o principal o objetivo estratégico o aperfeiçoamento do custo de seus recursos.

O estudo desenvolvido por Jolanta Pondel e Maciej Pondel, em 2016, tem como premissa que o crescimento do uso de dados pode beneficiar significativamente a maneira como é realizado o gerenciamento de projetos nas principais empresas mundiais. De acordo com Pondel (2016), a eficácia da decisão do projeto não é resultante apenas da quantidade de dados coletados, mas também depende da capacidade de escolha das fontes de informação e como elas serão aplicadas.

Segundo os autores, os projetos tornaram-se um conjunto muito complexo de atividades que exigem ferramentas de Tecnologia da Informação (TI) sofisticadas para suportar a eficiência de todas as ações. Para os mesmos, cada projeto tem a capacidade de gerar uma quantidade enorme de dados, ou seja, se a empresa for capaz de transformar dados coletados ao longo da sua vida executando os projetos em informações relevantes, pode-se maximizar a possibilidade de sucesso tanto do projeto quanto da organização.

Em um processo de gerenciamento de projetos e portfólio, a tomada de decisão gerencial é uma atividade inerente aos gerentes, proprietários de projetos, partes interessadas e às vezes também à equipe de execução de projetos. Por esse motivo, Pondel (2016) concluiu que as ferramentas analíticas que fazem uso de dados podem ter um significado crucial na tomada de decisão sobre o projeto e o portfólio.

Em concordância com o estudo de Pondel (2016), técnicas modernas de coleta e processamento de dados podem beneficiar a tomada de decisão durante o projeto especialmente em áreas de identificação de riscos, melhores estimativas de carga de trabalho dos recursos, planejamento, informações mais adequadas e compartilhamento de conhecimento.

Os processos de Gerenciamento de Projetos necessitam do uso de softwares para atuar como um plataforma de perfeito suporte para todas as ações que podem ser realizadas. É essencial que essas aplicações se comuniquem e que tragam valor para os usuários finais e partes interessadas do projeto. De acordo com Pondel (2016), os principais atributos beneficiados pelo uso de ferramentas de dados nos processos de gerenciamento de projetos são:

- processo de avaliação do projeto
- medição de custos e benefícios,
- relatórios de progresso,
- comunicação dos principais dados do projeto,
- planejamento de recursos e capacidade,
- Acompanhamento de custos e benefícios.

Segundo Xingguang (2022), o desenvolvimento e progresso de um país e da sociedade só pode ser possível com um grande número de talentos e com a elaboração de práticas que fomentem os recursos humanos como pontos estratégicos. A equipe de talentos abordada pelo autor necessita de uma organização com uma razoável estrutura e alta qualidade para desenvolver uma nova cultura em que os talentos são produzidos em grande número e melhores utilizados.

Com o foco na disseminação de melhores práticas de gestão de recursos, Xingguang (2022) faz uma comparação entre uma empresa e uma pequena sociedade, analisando o fato de que as organizações contam com seus funcionários como um recurso essencial. As pessoas, assim como na sociedade, exercem um papel importante nos negócios, por esse motivo, se uma empresa quer ser bem sucedida, deve prestar atenção à gestão de recursos humanos, dessa forma é especialmente importante que as empresas antecipem o desenvolvimento e o uso dos recursos humanos no futuro.

O estudo desenvolvido por Xingguang (2022), faz um aprofundamento sobre o impacto de aspectos essenciais na gestão de recursos de uma organização, como a escala de crescimento dos negócios da empresa e o aumento da produtividade com foco nas necessidades dos recursos humanos com o objetivo de melhorar a estrutura da equipe, aprimorar a alocação de pessoal, controlar os custos trabalhistas e melhorar a eficiência e os benefícios. No artigo em questão, foi desenvolvido um modelo de previsão da demanda dos recursos humanos baseado nas estratégias de negócios e nos benefícios financeiros orientado pelo desenvolvimento de recursos humanos.

Xingguang (2022) faz a implementação de uma rede neural em BP aprimorada para construir um sistema de previsão da demanda dos recursos humanos, resultando em um novo método de previsão quantitativa para previsão de demanda de recursos humanos com forte relevância teórica. O modelo proposto analisa e prevê o emprego total de recursos humanos da organização, a gestão de pessoal e a estrutura de qualidade no projeto.

Jianxiong (2009) realizou um estudo visando a previsão de demanda de recursos humanos em uma rede hoteleira na China utilizando uma combinação entre dois modelos diferentes. Segundo o autor, métodos de previsão de demanda aplicados em recursos que utilizam apenas modelos isolados levam a resultados com baixa acurácia e ineficácia nas aplicações.

Através de seu estudo, Jianxiong (2009) concluiu que existem limitações nos modelos propostos na literatura o que leva a baixa precisão dos métodos tradicionais. Como solução, Jianxiong (2009) levantou a hipótese de que uma combinação de modelos de previsão pode melhorar a precisão dos resultados e passar a ser um bom método preditivo para melhorar a capacidade preditiva da alocação de recursos no ramo da hotelaria chinesa e mundial.

O método proposto por Jianxiong (2009), utilizou conceitos de séries temporais, com a utilização do “*Grey Theory Model Forecast*” em conjunto com modelos de Artificial Neural Networks (ANN), sendo ele ANN com BP. A partir desse estudo, conclui-se que a utilização de métodos tradicionais e de apenas uma abordagem não garantem os melhores resultados, portanto a utilização da combinação de métodos pode ser benéfica para a problemática em questão, como a aplicação de modelos ensemble.

3. MODELO

Neste capítulo, serão apresentados os modelos propostos, os quais foram implementados como parte deste trabalho de conclusão de curso. O objetivo principal é explorar o potencial dos modelos de *machine learning* para auxiliar no processo de alocação de recursos humanos, visando melhorar a tomada de decisão nesse contexto específico.

Inicialmente, será apresentada a base de dados utilizada para treinamento e validação dos modelos. Essa base de dados consiste em informações detalhadas sobre projetos anteriores. A qualidade e a abrangência dessa base de dados são fundamentais para o desenvolvimento e a avaliação adequada dos modelos de recomendação.

Em seguida, serão descritos os modelos de *machine learning* utilizados, bem como suas respectivas técnicas e estampas aplicadas. Serão explorados modelos de aprendizado supervisionado, como regressão através de modelos como *Random Forest*, *Light GBM*, e *Gradient Boosting*. Cada modelo será aplicado à base de dados, levando em consideração métricas de avaliação e validação para verificar sua eficácia na recomendação de alocação de recursos humanos.

3.1. BASE DE DADOS

A base de dados utilizada para o treinamento do modelo foi consolidada a partir de um conjunto de cronogramas utilizados em alguns produtos com o consentimento dos gerentes de produto para a sua utilização e com as suas devidas adaptações para a implementação desta pesquisa.

Os dados de entrada buscados para esta pesquisa foram pensados para englobar dois principais fatores que impactam na decisão da alocação dos recursos, são eles:

- Fatores relacionados ao projeto: importância do projeto para a vida da sociedade impactada, meio ambiente, grau de importância estratégica e comercial, nível de criticidade de segurança, tecnologias necessárias,
- Fatores relacionados ao produto a ser entregue: nível de experiência necessário para realização da atividade, datas de início e término planejado e executado, duração planejada e executada, área de atuação.

A organização do cronograma segue uma estrutura livre de muitas recomendações com relação ao detalhamento, ou seja, fica a critério do gerente detalhar o quanto for

interessante para ele. Em decorrência disso, os cronogramas utilizados possuem características distintas do nível de detalhe. Por exemplo, há cronogramas detalhados até o nível de atividade e há cronograma detalhados até o nível de pacote de trabalho.

A tabela 2 abaixo sintetiza os dados considerados no modelo e apresenta suas características:

Tabela 2 – Características consideradas no modelo

#	Título	Descrição	Tipo da Variável	Valores
1	Alocação %	Variável alvo que representa o percentual de alocação demandado pela atividade	float	Valores de 0 a 1
2	Classificação	Representa o tipo de execução demandada pela atividade de acordo com os escopos de atuação da empresa.	texto	Os gerentes recebem uma lista de possíveis valores
3	Revisão	Indica que a atividade representa uma revisão de um trabalho já elaborado anteriormente.	booleano	0 ou 1
4	Duração	Representa a diferença entre as durações realizadas	float	-
5	Variância	Representa a diferença entre as durações planejadas	float	-
6	Impacto	Avalia a criticidade do impacto do produto na vida das pessoas envolvidas, no meio ambiente, na saúde e segurança ocupacional e na integridade.	ordinal	1: Baixo impacto. 2: Médio impacto. 3: Alto impacto.

7	Comercial	Avalia o impacto (positivo ou negativo) que o produto terá na imagem da organização somado com a análise do setor de atuação do produto	ordinal	1: Baixo impacto. 2: Médio impacto. 3: Alto impacto.
8	Equipe	Avalia o impacto que os colaboradores, as terceirizações e as equipes dos consórcios terão no desenvolvimento do produto.	ordinal	1: Baixo impacto. 2: Médio impacto. 3: Alto impacto.
9	Custo	Avalia o impacto que o custo do produto (receita, faturamento, custos diretos e indiretos e despesas corporativas) terá no resultado econômico da empresa.	ordinal	1: Baixo impacto. 2: Médio impacto. 3: Alto impacto.
10	Número de colaboradores	Quantidade de colaboradores envolvidos na execução da atividade	inteiro	-
11	Experiência	Nível de experiência necessária para executar a atividade	ordinal	Estagiário, Júnior, Pleno, Sênior, Especialista, Gerente e Diretor
12	Nota	Avaliação da alocação após a finalização da atividade descrita	float	Valores de 0 a 10

Fonte: Elaborado pelo autor (2023)

Por fim, após a entrega do componente das atividades, o gerente responsável dá uma nota, de 0 a 10, para a alocação realizada, essa nota servirá para ajustar, proporcionalmente à

nota, os valores da alocação no dataset de treino caso a nota seja inferior a 8. Basicamente, o valor da alocação é ajustado da seguinte forma:

$$\text{Nova Alocação} = \frac{10 * \text{Alocação Inferior a 8}}{\text{Nota}}$$

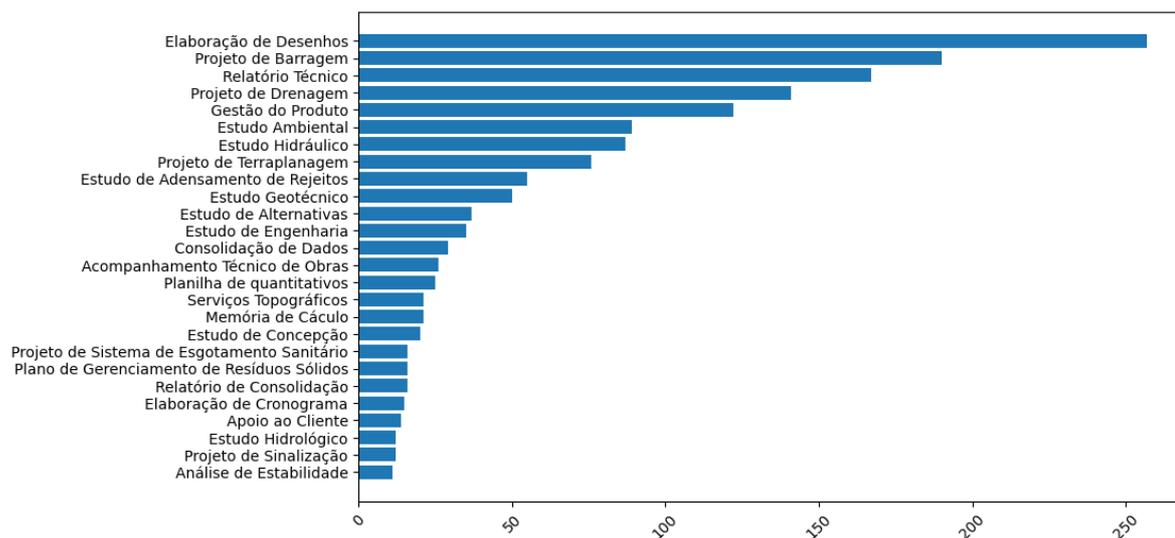
A partir disso, foram sendo ajustados os valores indesejados ao modelo por terem performado mal na execução das atividades. É importante destacar que essa nota se refere apenas ao desempenho do percentual de tempo despendido por essa atividade para a execução com qualidade da mesma. Ou seja, é importante para o decisor, no caso do gerente, avaliar corretamente o quão positiva foi a alocação para a entrega.

Para fins desta pesquisa, buscou-se adquirir informações de 6 tipos de escopos de serviços diferentes fornecidos pela empresa para que fosse possível englobar uma quantidade razoável de cenários possíveis e testar a possível generalização do modelo. Para isso, foram selecionados os seguintes escopos:

- Projeto de Mineração
- Projeto de Rodovias
- Projeto de Estação de Tratamento de Óleo
- Projeto de Sistema de Esgotamento Sanitário
- Projeto de Estação de Tratamento de Água
- Projeto de Adensamento de Rejeitos

Com relação a essas especialidades, a imagem abaixo retrata detalhadamente quais as principais classificações presentes no banco de dados utilizado.

Imagem 1 - Quantidade de atividades por Classificação



Fonte: Elaborado pelo autor (2023)

Entretanto, ainda assim, cerca de 1050 entradas (na aplicação inicial) continuam sem uma classificação preenchida pelo gerente, isso pode acontecer por se tratar de uma atividade ampla e sem uma característica única que a defina. Porém, para retratar a realidade do problema esses casos foram tratados como “Sem Classificação” e foram incluídos no modelo.

3.2. IMPLEMENTAÇÃO

O modelo em questão pretende aprofundar a integração de gestão de projetos com a utilização de dados históricos, visando simplificar o processo de alocação de recursos e identificação de padrões. Com a aplicação de modelos de aprendizado de máquina, os gestores de projetos poderão tomar decisões mais precisas durante o planejamento e gerenciamento dos recursos em projetos empresariais de diferentes tamanhos.

Ao avaliar a existência de padrões relacionados às características intrínsecas do projeto, atividades e colaboradores em comparação com alocações que resultaram em bons desempenhos e metas alcançadas, a pesquisa tem como objetivo gerar um sistema capaz de recomendar a alocação dos recursos humanos nas diferentes atividades de um projeto. Com isso, a tomada de decisões será mais ágil e assertiva, permitindo que os gestores possam se concentrar em outras áreas críticas do projeto. Em suma, a pesquisa tem o potencial de transformar a forma como os projetos são gerenciados e aperfeiçoar o uso dos recursos empresariais.

O modelo em questão é uma ferramenta de análise de dados que utiliza modelos de regressão para identificar padrões presentes nos cronogramas de projetos elaborados pela equipe de planejamento. Ao coletar os dados fornecidos e aplicá-los à fase de treinamento do

algoritmo, o modelo utiliza análises e aplicações estatísticas para aprender e compreender os padrões presentes nos dados, com o objetivo de replicá-los no futuro.

O processo de treinamento é fundamental para garantir a eficácia do modelo, pois é durante essa fase que os dados históricos são utilizados para identificar correlações e padrões que podem ser aplicados em projetos futuros.

Para a aplicação em questão, pretendeu-se testar três diferentes modelos de regressão que possuem características de ensemble para aprimoramento dos resultados. Dessa forma seria possível identificar o algoritmo de melhor desempenho com relação à assertividade das predições e de performance de tempo.

Para o desenvolvimento da proposta de solução, a presente pesquisa utilizou a linguagem de programação *Python*, que se destaca por sua simplicidade. Sua sintaxe é fácil de entender, tornando-a uma excelente opção. Além disso, *Python* possui uma grande biblioteca de módulos e pacotes disponíveis, o que permite economizar tempo e esforço na criação de novas funcionalidades, em vez de ter que desenvolvê-las do zero. *Python* também é conhecido por sua eficiência e velocidade de execução, o que o torna uma ótima escolha para aplicativos que requerem processamento rápido de dados. A linguagem também é compatível com outras linguagens, permitindo que os desenvolvedores integrem facilmente o código *Python* com outros sistemas.

O passo a passo da construção do código segue os princípios do *Data Pipeline* o qual visa construir uma sequência de etapas lógicas que levam desde a adquirir os dados da sua fonte e levam ao produto final da proposta. Dessa forma, o Pipeline de dados estrutura as etapas do processo com um objetivo final.

Para o presente trabalho, o *Pipeline* do código desenvolvido seguiu as seguintes etapas:

1. Instalação de Pacotes
2. Criação de funções de suporte
3. Importar pacotes
4. Carregamento de dados
5. Tratamento de dados
6. Separação dos dados de treino, teste e validação
7. Criação dos modelos
8. Treinamento dos modelos
9. Avaliação dos modelos

As três primeiras etapas visam basicamente apoiar a implementação do modelo através das bibliotecas, pacotes e funções que economizam tempo no decorrer do código e simplificam a visualização. É importante lembrar que uma implementação eficiente do modelo é crucial para o sucesso do projeto. Além disso, a simplificação do código torna o processo de manutenção mais fácil e menos propenso a erros. Isso é especialmente importante em projetos de longo prazo, nos quais a manutenção pode se tornar uma tarefa difícil e consome tempo. Ao utilizar bibliotecas, pacotes e funções, o processo de implementação pode ser concluído de forma mais rápida e eficiente, permitindo mais tempo para outras tarefas importantes, como análise de resultados e planejamento de próximos passos.

A fase de carregamento de dados contou com a utilização de uma API de uma plataforma de Gestão de Projetos utilizada pela empresa que tem como foco a criação e manutenção de cronogramas, o Smartsheet. Para isso, utilizou-se uma biblioteca, disponível em Python, específica para essa ferramenta.

A etapa seguinte, Tratamento de Dados, é primordial para atingir os objetivos do trabalho. Esta etapa consiste em um conjunto de procedimentos que preparam os dados coletados para que fosse possível utilizar os modelos e testá-los. Durante o tratamento de dados, os dados brutos foram limpos, organizados e transformados em um formato adequado para análise. Isso envolve a remoção de dados ausentes ou inconsistentes, a normalização de dados para que estejam na mesma escala, a criação de novas variáveis compostas a partir de variáveis existentes e a seleção de variáveis relevantes para a análise. Essa etapa é importante para garantir que os dados estejam em condições ideais para serem analisados pelos modelos, e que os resultados obtidos sejam precisos e confiáveis.

Com os dados prontos para a implementação, a última etapa antes da criação do modelo é realizar a separação entre conjunto de dados de treino, validação e de teste. Essa etapa permite avaliar o desempenho do modelo em dados que não foram usados para treinamento, ajudando a detectar problemas como *overfitting* e *underfitting* e aprimorando a capacidade de generalização do modelo.

Ao separar o dataset em três conjuntos, o resultado é um conjunto de treino, que é usado para ajustar os pesos do modelo aos dados, um conjunto de validação, usado para avaliar a performance do modelo durante o treinamento e realizar ajustes nos hiperparâmetros do modelo, e um conjunto de teste, usado para avaliar a performance final do modelo em dados que ele nunca viu antes e medir sua capacidade de generalização.

Finalizado essa etapa, os modelos propostos inicialmente foram construídos a partir de pacotes da biblioteca *scikit-learn*. Os modelos testados foram:

- *Random Forest Regressor (RF)*
- *Light GBM (Light GBM)*
- *Gradient Boosting Regressor (GB)*

Para melhor seleção dos hiperparâmetros, foi realizado *GridSearch* em conjunto com validação cruzada. Essa combinação foi utilizada por ser uma técnica importante e eficaz para a seleção de hiperparâmetros em modelos de aprendizado de máquina. A utilização do número adequado de *folds* na validação cruzada permite avaliar o desempenho do modelo de forma estável e eficiente. A escolha do número de *folds* é importante, pois pode afetar a estabilidade dos resultados. O valor $cv=5$ é frequentemente utilizado, pois fornece um bom equilíbrio entre a variação dos resultados e a eficiência computacional. Com $cv=5$, os dados são divididos em 5 partes iguais, e o modelo é treinado e validado 5 vezes, utilizando cada parte como conjunto de validação uma vez.

A aplicação dessa técnica pode melhorar significativamente o desempenho do modelo, contribuindo para resultados mais precisos e confiáveis em diversas áreas de aplicação.

Os modelos foram sendo rodados consecutivamente a fim de testar diferentes tipos de hiperparâmetros até chegar a um certo subconjunto de parâmetros que foram aprovados experimentalmente a partir das melhores pontuações de saída.

Como medida de avaliação dos modelos, foi utilizado o R^2 , também conhecido como coeficiente de determinação, é uma métrica que indica a proporção da variação total da variável dependente que é explicada pelas variáveis independentes incluídas no modelo.

O material apresentado por Zou e Hastie (2005) fornece uma justificativa sólida para o uso do R^2 como métrica de avaliação em modelos de regressão. O artigo argumenta que é uma medida útil para avaliar a adequação do modelo aos dados e para comparar diferentes modelos de regressão.

O R^2 é definido como a proporção da variação na variável dependente que é explicada pelo modelo de regressão. Em outras palavras, quanto maior o valor do R^2 , melhor é o ajuste do modelo aos dados. Isso é importante porque o objetivo de um modelo de regressão é explicar a variação na variável dependente com base nas variáveis independentes.

Além disso, o artigo de Zou e Hastie destaca que o coeficiente de determinação é uma medida padronizada, o que significa que é independente da escala dos dados. Isso é importante porque diferentes conjuntos de dados podem ter unidades de medida diferentes e escalas diferentes, o que pode afetar outras medidas de avaliação, como o erro médio absoluto (MAE) ou o erro quadrático médio (MSE).

Ainda segundo o artigo, o R^2 também é uma medida fácil de interpretar e comunicar, o que é importante para uma ampla audiência. Um valor de R^2 próximo a 1 indica que o modelo explica a maior parte da variação nos dados, enquanto um valor próximo a 0 indica que o modelo não é capaz de explicar a variação nos dados.

Entretanto, o uso exclusivo do R^2 na avaliação de modelos pode levar a conclusões enganosas, isso ocorre porque o critério não leva em consideração a complexidade do modelo, o que significa que um modelo mais complexo pode ter um valor mais alto do que um modelo mais simples, mesmo que o último seja mais adequado. Além disso, não fornece informações sobre a magnitude ou direção dos erros de previsão, o que pode ser importante em muitas aplicações.

Portanto, embora o R^2 seja uma métrica importante na avaliação de modelos de regressão e amplamente utilizada na prática, é crucial lembrar que não se trata a única métrica a ser considerada. É necessário levar em consideração outros fatores. Em decorrência disso, foi levado em consideração também o tempo de processamento como uma medida simples para avaliar se a complexidade do modelo e de seus resultados são positivos para os objetivos estratégicos da organização.

Outras medidas também foram utilizadas para avaliação a nível de comparação entre os três modelos testados para o problema, entre eles:

- *Explained Variance*
- *Mean absolute error (MAE)*
- *Mean squared error (MSE)*

Ao iniciar o processo de aplicação e testes dos modelos, foi definida uma grade de valores para cada hiperparâmetro que seria avaliado para os respectivos modelos foram testados a fim de potencializar os resultados dos testes. Essa grade foi criada a partir de uma análise prévia dos dados, considerando as características do conjunto de treinamento e validação, bem como a literatura especializada na área.

A partir da definição da grade de valores, foram realizadas diversas iterações do processo de treinamento e validação do modelo, testando cada combinação possível de valores dos hiperparâmetros. Em cada iteração, foram avaliados os resultados de desempenho do modelo, as métricas utilizadas serão citadas na seção a seguir.

Ao final do processo de *GridSearch (GS)*, foi selecionada a combinação de hiperparâmetros que apresentou o melhor desempenho, com base nas métricas escolhidas. Essa combinação foi então utilizada para gerar as previsões do modelo em um conjunto de

teste independente, a fim de avaliar o desempenho real do modelo em um cenário de produção.

Em um primeiro momento, os testes foram aplicados apenas para os 6 cronogramas listados e todos os modelos foram testados. Para fins de experimentação, a quantidade de dados foi sendo acrescida a partir da simples duplicação dos cronogramas consecutivamente e os testes foram sendo aplicados a fim de verificar se o incremento de dados seria capaz de aumentar a capacidade de previsão do modelo.

Para a avaliação da importância das características, foi utilizada a função *Feature Importances*, presente na biblioteca *scikit-learn*. Essas métricas são calculadas considerando a redução média na medida de impureza (Gini ou entropia) ou erro quadrático quando um determinado recurso é usado para dividir os dados em um nó da árvore. Quanto maior a diminuição na impureza ou erro quadrático, maior é a importância do recurso. Para regressão com árvores de decisão e florestas aleatórias, a métrica usada é a importância média da diminuição do erro quadrático (*mean decrease accuracy*).

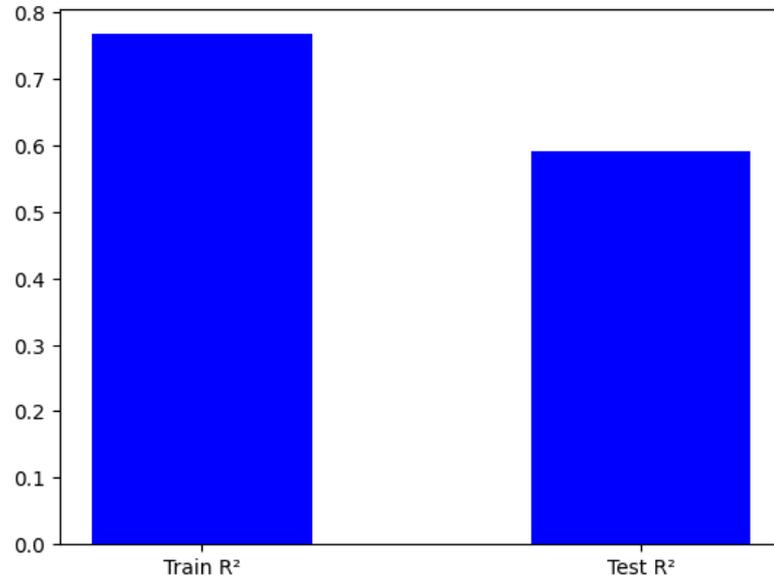
Foram avaliados três modelos de regressão utilizando *machine learning*: *Random Forest*, *Light GBM* e *Gradient Boosting*. Cada algoritmo foi treinado e validado com os dados de um conjunto de dados específico e seus resultados foram analisados em termos de desempenho de previsão e tempo de processamento.

- *Random Forest*

Após o treinamento e predição através dos aprendizados do algoritmo de *Random Forest*, os resultados apresentados foram os seguintes:

- *R² on training set: 0.767*
- *R² on validation set: 0.591*
- *Explained variance score: 0.591*
- *Mean absolute error (MAE): 10.644*
- *Mean squared error (MSE): 322.063*
- *Processing Time: 2189.489*

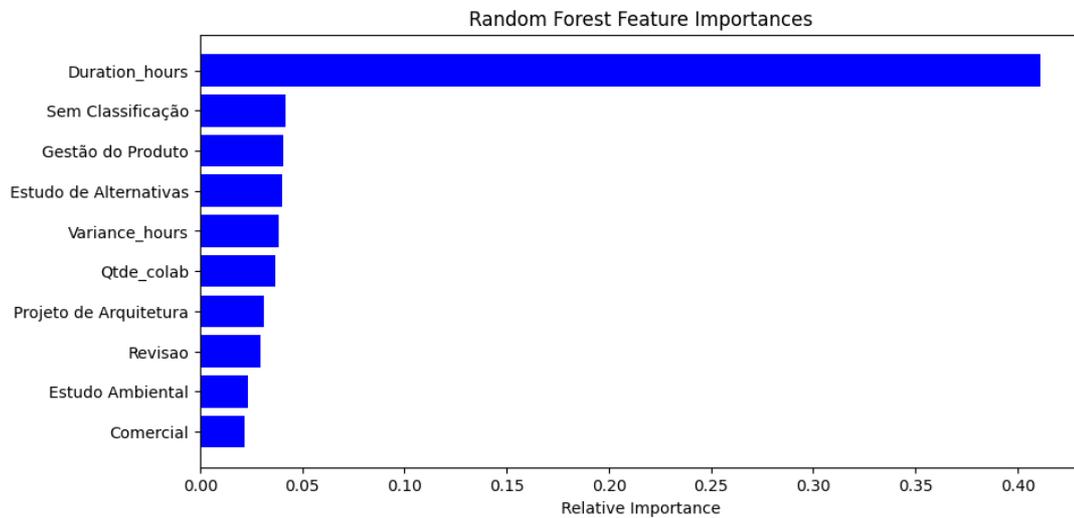
Imagem 2 - Resultados *RF*



Fonte: Elaborado pelo autor (2023)

As 10 principais características que mais impactam a predição do percentual de horas pode ser observada através dos resultados do gráfico abaixo:

Imagem 3 - Importância de variáveis através do *RF*



Fonte: Elaborado pelo autor (2023)

O *RF* apresentou um bom desempenho de previsão, com um R^2 de 0,767 no conjunto de treinamento e um R^2 de 0,591 no conjunto de validação. O tempo de processamento foi o mais longo, com 2189.489 segundos. Esses resultados indicam que o *Random Forest* conseguiu aprender bem com o conjunto de treinamento e generalizar para o conjunto de

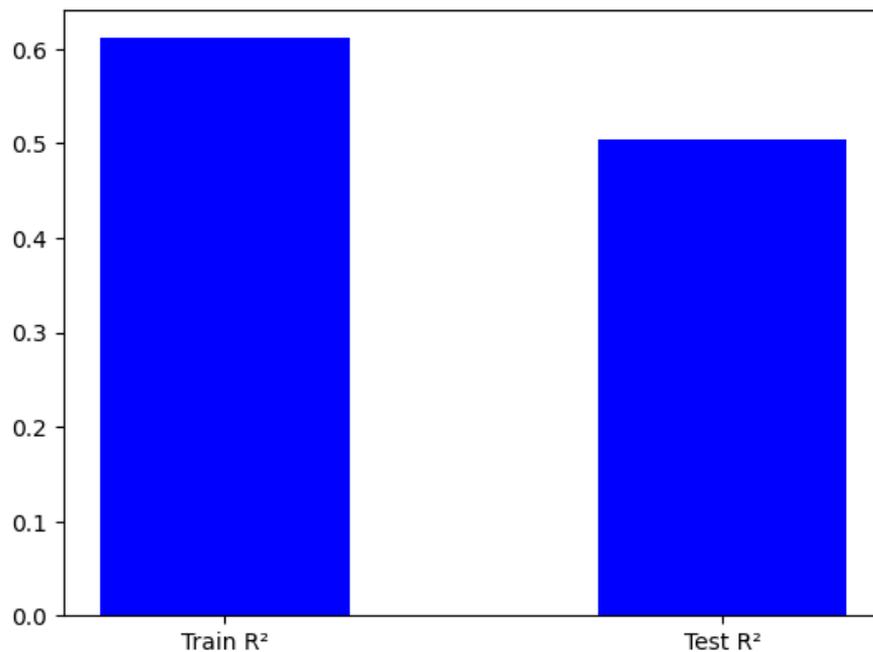
validação, mas também pode ter ocorrido uma ligeira sobreajuste (*overfitting*) durante o treinamento.

- **Light GBM**

Após o treinamento e predição através dos aprendizados do algoritmo de *Light GBM*, os resultados apresentados foram os seguintes:

- *R² on training set: 0.611*
- *R² on validation set: 0.504*
- *Explained variance score: 0.506*
- *Mean absolute error (MAE): 0.070*
- *Mean squared error (MSE): 0.034*
- *Processing Time: 498.602*

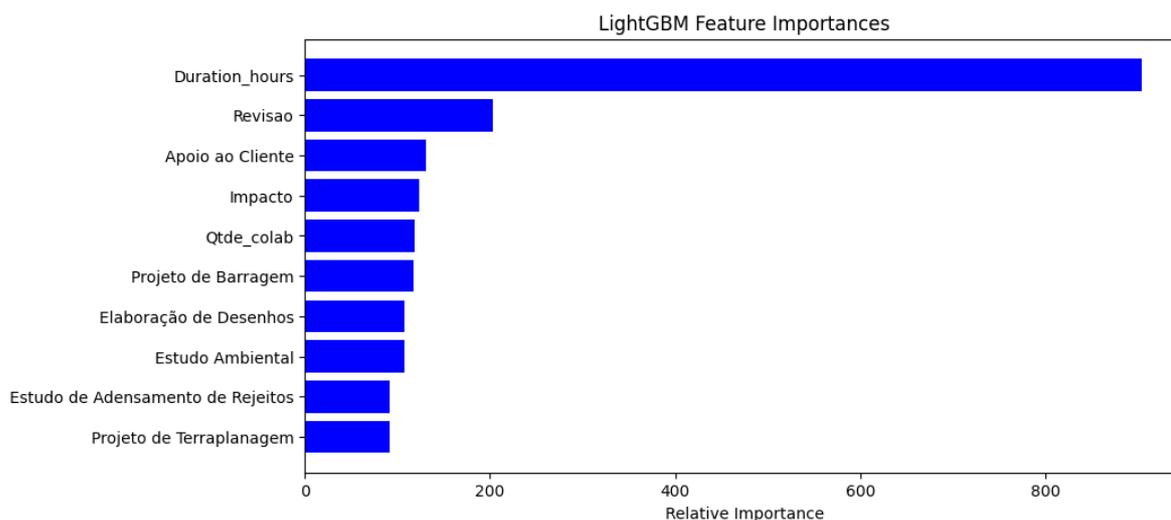
Imagem 4 - Importância de variáveis através do *Light GBM*



Fonte: Elaborado pelo autor (2023)

As 10 principais características que mais impactam a predição do percentual de horas pode ser observada através dos resultados do gráfico abaixo:

Imagem 5 - Importância de variáveis através do *Light GBM*



Fonte: Elaborado pelo autor (2023)

O modelo *Light GBM* apresentou um R^2 de 0,611 no conjunto de treinamento e um R^2 de 0,504 no conjunto de validação, mostrando um desempenho inferior ao modelo *RF*. Porém, é importante ressaltar que o tempo de processamento do *Light GBM* foi significativamente menor, levando apenas 498,6 segundos para treinar o modelo.

Embora o *Light GBM* não tenha aprendido tão bem quanto o *RF*, os resultados indicam que ele ainda possui uma boa capacidade de generalização, ou seja, o modelo conseguiu aplicar o aprendizado dos padrões de forma semelhante nos conjuntos de treinamento e teste. Isso torna o *Light GBM* uma boa opção para treinos futuros, uma vez que o modelo possui potencial para melhorar seu desempenho com ajustes adequados em seus parâmetros.

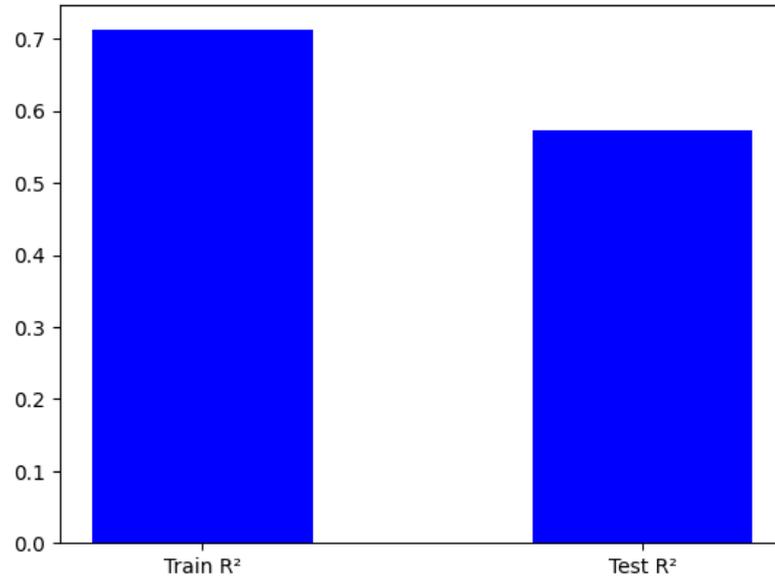
- **Gradient Boosting**

Após o treinamento e predição através dos aprendizados do algoritmo de *Gradient Boosting*, os resultados apresentados foram os seguintes:

- *R² on training set: 0.712*
- *R² on validation set: 0.572*
- *Explained variance score: 0.572*

- *Mean absolute error (MAE): 12.428*
- *Mean squared error (MSE): 336.627*
- *Processing Time: 151.762*

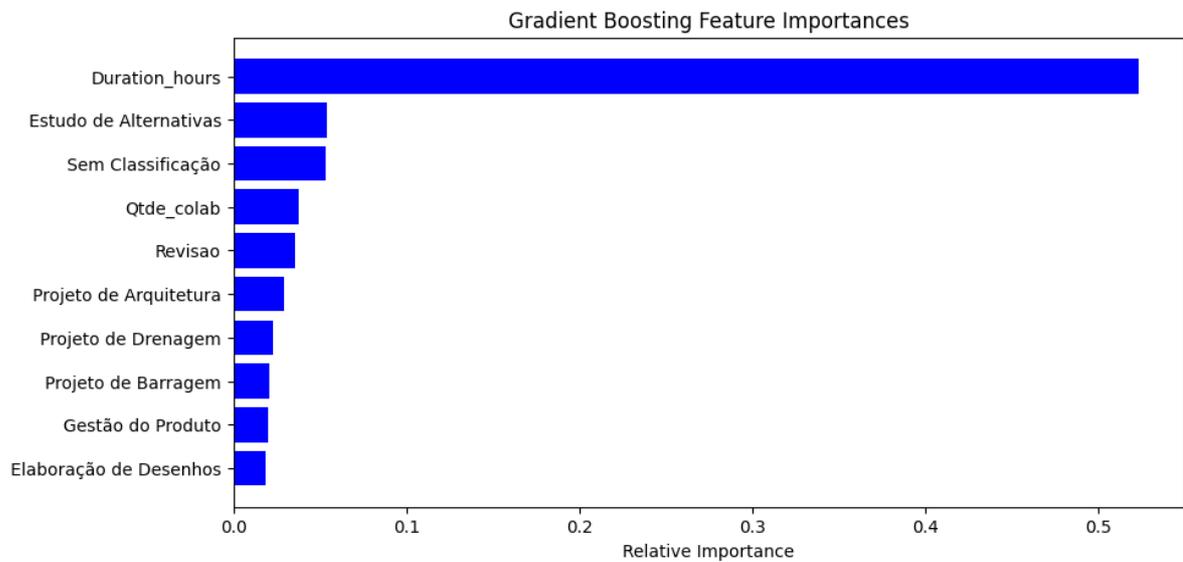
Imagem 6 - Importância de variáveis através do GB



Fonte: Elaborado pelo autor (2023)

As 10 principais características que mais impactam a predição do percentual de horas pode ser observada através dos resultados do gráfico abaixo:

Imagem 7 - Importância de variáveis através do GB



Fonte: Elaborado pelo autor (2023)

O *GB* apresentou um desempenho intermediário em relação aos outros dois modelos, com um R^2 de 0,712 no conjunto de treinamento e um R^2 de 0,572 no conjunto de validação. O tempo de processamento foi moderado, com 151.762 segundos. Esses resultados indicam que o *GB* conseguiu aprender bem com o conjunto de treinamento e generalizar para o conjunto de validação, mas com um desempenho inferior em relação ao *Random Forest*.

Para a interpretação dos resultados com base nos resultados apresentados, pode-se inferir que os três modelos foram capazes de aprender com os dados e realizar previsões relativamente precisas em relação à alocação de recursos humanos em projetos. No entanto, o desempenho do RF foi o melhor em termos de R^2 no conjunto de validação, indicando que é o algoritmo mais adequado para prever esses valores percentuais com mais precisão. No entanto, o *Light GBM* apresentou um tempo de processamento muito menor e um desempenho ainda razoável em relação aos outros dois modelos, sendo uma boa opção quando a velocidade de processamento é um fator crítico.

O MAE e MSE apresentados por cada algoritmo mostraram que, em média, os erros de previsão foram relativamente baixos em todos os casos, o que é um bom sinal para o uso prático desses modelos. A variação explicada também foi relativamente alta, o que sugere que os modelos foram capazes de explicar bem a variação nos dados e fornecer previsões confiáveis.

Entretanto, por questões de tempo e recursos esta pesquisa aplicou a um certo conjunto de dados limitados, porém foi possível observar, após o incremento de dados citados anteriormente na modelagem do algoritmo, que o desempenho do modelo aumentava consideravelmente a partir do incremento de dados.

Com relação a identificação das principais características, pode-se observar que em todos os modelos a Duração da atividade tem grande impacto na alocação sendo fortemente destacada com relação às demais. As demais características se diferem em sua importância dependendo do modelo apesar de ser possível identificar a presença constante dos atributos “Estudo de Alternativas” e Quantidade de Colaboradores como elemento importante na identificação dessa alocação.

A pesquisa teve como objetivo desenvolver um sistema capaz de avaliar a existência de previsibilidade sobre o percentual de alocação de recursos humanos em projetos que traz resultados positivos, e recomendar o percentual que deve ser alocado para cada atividade, de maneira a atender aos objetivos organizacionais. Para alcançar esse objetivo, a pesquisa

buscou determinar quais atributos seriam utilizados como foco da melhora do algoritmo, modelar a solução para obter o melhor desempenho, encontrar melhores parâmetros dos modelos que geram melhores resultados, avaliar a utilidade do modelo proposto para aplicação prática e propor um processo adequado de interação dos gerentes com o modelo.

Com base nos resultados apresentados, pode-se afirmar que a pesquisa atingiu seus objetivos, uma vez que foram desenvolvidos e avaliados três modelos de regressão de *machine learning* capazes de prever o percentual de alocação de recursos humanos em projetos com bom desempenho e precisão. Além disso, os atributos utilizados para aprimorar os modelos foram determinados, e os parâmetros dos modelos foram aprimorados para melhorar o desempenho.

A utilidade do modelo proposto para a aplicação prática foi avaliada, e os resultados indicaram que as previsões geradas pelos modelos são satisfatórias e podem ser usadas como orientação na alocação de recursos humanos em projetos.

Em suma, a pesquisa atingiu seus objetivos e desenvolveu modelos capazes de prever a alocação de recursos humanos em projetos de maneira a se tornar útil para a tomada de decisões organizacionais. Os resultados da análise dos três modelos de regressão mostraram que cada um apresentou pontos fortes e fracos em termos de desempenho de previsão e tempo de processamento, podendo ser aprimorados a partir do incremento de uma quantidade maior de dados quando para cada tipo de classificação realizada pelos gestores durante o preenchimento do cronograma.

4. CONCLUSÃO

Neste capítulo, serão apresentados os resultados da pesquisa que teve como objetivo desenvolver um sistema capaz de avaliar a previsibilidade do percentual de alocação de recursos humanos em projetos, recomendando o percentual que deve ser alocado para cada atividade e separando-os entre os recursos disponíveis de forma a atender aos objetivos organizacionais.

Foram descritos os modelos de regressão de *machine learning* utilizados, seus parâmetros e desempenho, bem como a avaliação da utilidade do modelo proposto para aplicação prática. Também foram apresentados os atributos utilizados para aperfeiçoar os modelos. A interpretação dos resultados reflete a capacidade de o modelo retornar um resultado adequado a partir da entrada dos dados de cada atividade, ou seja, para receber a recomendação da alocação, o gerente deve inserir as características das atividades citadas anteriormente e o algoritmo retorna um valor percentual a partir do aprendizado realizado.

É importante ressaltar que a interpretação dos valores previstos pelos modelos deve ser feita com cuidado, uma vez que a alocação de recursos humanos em projetos é influenciada por uma série de fatores complexos e interconectados, que podem não ser totalmente capturados pelos dados utilizados no treinamento.

Por esse motivo, foram realizados estudos sobre a importância das características sobre a maneira em que os gerentes realizaram as principais alocações que geraram resultados positivos nos projetos. Na apresentação das métricas, há a presença de gráficos intitulados de *Feature Importances*, que fornece uma maneira de avaliar a importância relativa dos recursos em relação aos outros recursos do modelo e ajuda a selecionar os recursos mais informativos para a modelagem.

A recomendação principal é que as previsões geradas pelos modelos sejam utilizadas como uma orientação inicial, e que outras informações e considerações relevantes também sejam levadas em conta na tomada de decisões em relação à alocação de recursos humanos em projetos.

4.1. CONSIDERAÇÕES FINAIS

Dentro do cenário previsto por essa pesquisa e a partir das análises apresentadas por outros autores, pode-se concluir que uma alocação dos recursos humanos realizada de forma assertiva e equilibrada se torna imprescindível para garantir um custo adequado, uma entrega

de valor em tempo esperado e alta qualidade no projeto, além de permitir a identificação de possíveis melhorias.

Além dos aspectos citados anteriormente, uma boa alocação dos recursos permite uma melhor precificação dos seus projetos, evita a ocorrência de sobrecarga de um colaborador, um controle mais adequado do nível de produtividade da equipe, a previsão da demanda de homem-hora em novas tarefas/projetos e reduz drasticamente o tempo gasto do Gerente do Projeto no planejamento das horas necessárias para a execução de cada tarefa que compõe o escopo do projeto.

Para o desenvolvimento da proposta de solução, pode-se perceber que não existem muitas aplicações de modelos de recomendação aplicados a gestão de projetos que utilizem uma combinação de dois ou mais modelos para a proposição da solução e conseqüentemente alavancam os resultados do modelo. Deste modo, esta pesquisa identificou um potencial de agregar conhecimento tanto ao mercado quanto para empresas de consultoria ou qualquer empresa que tenha um interesse em uma melhoria contínua de suas utilizações de recursos.

Além dos pontos citados acima, pode-se adicionar benefícios econômico-financeiros e para RH ao alocar recursos de forma mais adequada em um portfólio de projetos. Em primeiro lugar, pode levar a um aumento significativo na eficiência e produtividade da empresa. Isso se deve ao fato de que, com as ferramentas certas, a empresa pode priorizar projetos de alto valor estratégico e financeiro e recursos mais relevantes podem ser alocados de forma mais eficaz. A eficiência aumentada também pode ser vista na redução de retrabalho e no aumento da qualidade do produto final por meio do aumento de entregas de alta qualidade desenvolvidas no prazo correto.

Como consequência, a adoção de uma ferramenta mais adequada para realizar a alocação de recursos pode levar a empresa a tomar decisões mais informadas e estratégicas. As ferramentas podem fornecer informações valiosas sobre o desempenho passado do projeto e ajudar a empresa a prever o sucesso futuro. Com base nesses dados, a empresa pode tomar decisões mais informadas sobre quais projetos alocar os seus membros mais .

Uma melhor alocação de recursos humanos pode ter um impacto positivo no pessoal de RH da empresa. Isso pode incluir melhorias na capacidade da empresa de reter e atrair talentos, bem como redução no turnover de funcionários, auxiliando a equipe de RH a criar um ambiente de trabalho mais equilibrado, permitindo que os funcionários se concentrem em projetos mais valiosos e gratificantes, em vez de se preocuparem com projetos menos significativos.

No caso do trabalho desenvolvido, a utilização de *Machine Learning* apresentou algumas vantagens. Uma delas é a capacidade de processar grandes quantidades de dados em um curto período de tempo, o que permite uma análise mais completa e precisa. Além disso, o modelo de *Machine Learning* foi capaz de identificar padrões e correlações entre variáveis que seriam difíceis ou impossíveis de detectar manualmente.

Outra vantagem da utilização de *Machine Learning* neste trabalho foi a capacidade de prever os resultados futuros. Com base em dados históricos, o modelo de *Machine Learning* foi capaz de prever tendências e comportamentos futuros, o que permitiu aos pesquisadores tomar decisões mais informadas e baseadas em dados.

Por fim, a utilização de *Machine Learning* neste trabalho permitiu que os pesquisadores vislumbrassem a possibilidade de automatização dos processos que sempre foram realizados manualmente, economizando tempo e recursos valiosos.

Os resultados deste estudo permitem concluir que é viável identificar padrões na estratégia de alocação de recursos em um projeto, utilizando somente algumas características do próprio projeto, bem como das atividades a serem executadas e dos colaboradores envolvidos. A análise do perfil dos projetos investigados e dos padrões identificados na alocação de recursos pelos gerentes de projetos destaca a complexidade prática dessa tarefa. Os gestores de projetos frequentemente enfrentam pressões para controlar os custos, mas, muitas vezes, não dispõem de ferramentas específicas para alocar os recursos de maneira eficiente. Além disso, eles precisam lidar com a carga de trabalho de vários projetos simultâneos.

Nesse contexto, os resultados deste estudo fornecem evidências que podem ajudar os gerentes de projetos a gerir de forma eficaz a alocação de recursos, considerando as pressões inerentes à tríade custo, tempo e qualidade. Adicionalmente, a identificação de padrões pode contribuir para uma gestão mais assertiva e eficiente de recursos humanos, o que pode ter um impacto significativo no sucesso do projeto. É importante ressaltar, no entanto, que a alocação de recursos é apenas um aspecto da gestão de projetos e que outras dimensões, como a comunicação, o monitoramento e a mitigação de riscos, também são críticas para o sucesso do projeto.

4.2. IMPLICAÇÕES PRÁTICAS

A avaliação do perfil dos projetos investigados e dos padrões encontrados na maneira de alocar recursos dos gerentes de projetos corrobora a complexidade prática de efetuar a distribuição de recursos humanos. O gerente depara-se com demandas de controle de custos,

sem a disponibilidade de ferramentas específicas para tal alocação, e a sobrecarga de projetos simultâneos. Nesse sentido, este estudo apresenta evidências que podem auxiliar o gerente de projetos a lidar com as pressões inerentes à tríade custo, tempo e qualidade.

Portanto, de maneira prática a pesquisa traz resultados positivos nesse quesito ao permitir o avanço desses estudos na empresa proposta. A aplicação do modelo permite facilitar o trabalho dos gerentes de produto ao realizar o planejamento da equipe pois apresenta uma série de valores que podem ser conferidos e validados pela equipe de planejamento economizando o tempo dos envolvidos além de permitir que erros sejam evitados ao utilizar modelos estatísticos para analisar os dados históricos e escolher as melhores saídas para a valor de alocação desejada.

4.3. LIMITAÇÕES E FUTUROS TRABALHOS

O presente estudo contém limitações que podem estar atreladas tanto a sua premissas adotadas, quanto aos meios buscados para a solução do problema, ou seja, os modelos de aprendizado de máquina.

Em primeiro lugar, é importante considerar que a equipe responsável pela execução da tarefa deve ser previamente definida, levando em conta não apenas o nível de experiência necessário para a realização do trabalho, mas também as habilidades específicas exigidas para a conclusão bem-sucedida da tarefa. Essa seleção deve ser feita com base em critérios objetivos e transparentes, garantindo a eficiência e eficácia do projeto.

No entanto, é necessário salientar que a alocação de colaboradores em projetos deve levar em conta não apenas o perfil técnico de cada indivíduo, mas também sua disponibilidade para a realização da tarefa. Nesse sentido, é importante destacar que a representação da resposta retorna a alocação necessária para a finalização adequada da tarefa, sem considerar o percentual de alocação em que um colaborador já está associado. Caberá ao decisor entender qual indivíduo será destinado para a atividade, levando em conta não apenas suas habilidades técnicas, mas também sua disponibilidade e carga de trabalho atual.

Outra limitação importante a ser considerada é a quantidade de dados disponíveis sobre determinadas classificações de atividades. Uma base de dados reduzida sobre determinada classificação de atividade pode levar a resultados inconsistentes para essa classificação. Nesse sentido, é importante destacar a necessidade de investimentos em pesquisas e coleta de dados relevantes para garantir a eficácia do modelo de recomendação proposto.

Por fim, é importante destacar que a base de dados utilizada no modelo de recomendação contém cronogramas de apenas 6 tipos de projetos realizados pela empresa, não englobando a totalidade de serviços fornecidos pela organização. Essa limitação pode levar a resultados imprecisos para projetos de natureza distinta, sendo necessário o aprimoramento constante da base de dados para aprimorar a eficácia do modelo.

Há diversas oportunidades de futuros trabalhos que podem contribuir para aprimorar o modelo de recomendação proposto para alocação de recursos humanos em projetos de engenharia utilizando *Machine Learning*. Algumas possibilidades incluem:

- Realizar um retreinamento do modelo a fim de englobar novos tipos de escopos de projetos, de forma a ampliar a capacidade do modelo em lidar com projetos de natureza distintas;
- Aumentar o conjunto de dados utilizados para treinamento do modelo, incluindo informações sobre projetos concluídos anteriormente, bem como dados sobre a alocação de recursos humanos em cada projeto. Isso pode permitir uma análise mais precisa das necessidades de recursos humanos para cada tipo de projeto;
- Testar diferentes modelos dos utilizados na presente pesquisa, buscando avaliar sua eficácia em diferentes cenários e tipos de projetos;
- Desenvolver uma maneira de facilitar a implementação e visualização dos resultados por parte dos gerentes, podendo ser através de um simples web app ou uma maneira mais robusta atrelada ao próprio sistema da empresa para realização das predições de forma automática. Isso pode permitir uma maior agilidade na tomada de decisão e garantir a utilização efetiva do modelo pelos responsáveis pela alocação de recursos humanos na empresa.

Além disso, é importante avaliar a possibilidade de integração do modelo de recomendação com outras ferramentas de gestão de projetos, a fim de aumentar a eficiência e eficácia das operações na empresa. Essas iniciativas podem permitir aprimorar a capacidade da empresa em lidar com projetos complexos e aumentar sua competitividade no mercado.

REFERÊNCIAS

BARBOSA, E. F.; CARVALHO, M. M. **Alocação de recursos humanos em projetos**. São Paulo: Atlas, 2017.

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, p. 5-32, 2001. Disponível em: <https://link.springer.com/article/10.1023/A:1010933404324>

BROWN, S; EISENHARDT, K. **Competing on the Edge: Strategy as Structured Chaos**. Editora Harvard Business. Vol. 31, No. 5, p. 786 a 789, 1998.

BUTTRICK, R. **Project Workout: The Ultimate Handbook of Project & Programme Management**. Editora Trans-Atlantic Publications 4ed. 2009. Acessado em 01 de set. de 2022.

ÇAĞDAŞ, V.; STUBKJÆR, E. Design research for cadastral systems. **Computers, Environment and Urban Systems**, v. 35, p. 77-87, 2011. Disponível em: <http://dx.doi.org/10.1016/j.compenvurbsys.2010.07.003>

CARVALHO, André. **Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina**. LTC, Rio de Janeiro, 2ª edição, 2010.

CHANG, N. **The Application of Neural Network to the Allocation of Enterprise Human Resources**. 2nd International Conference on E-business and Information System Security, 2010, p. 1-4.

COOPER, R.G.; EDGETT, S.J. KLEINSCHMIDT, E.J. New Product Portfolio Management: Practices and Performance. **Journal of Product Innovation Management**. 1999. Acessado em: 01 de setembro de 2022.

COVER, T. **Nearest Neighbor Pattern Classification**. IEEE Transactions on Information Theory. Vol. IT-13, No. 1, 1967.

GRAY, C. ; LARSON, E. **Gerenciamento de Projetos: o processo gerencial**. AMGH Editora Ltda, 4ed. 2009.

HAIR JR, J. F. *et al.* **Análise multivariada de dados**. Porto Alegre: Bookman Editora, 2009.

HODSON, D. **Existe um método científico?**. São Paulo: Universidade de São Paulo. Traduzido e adaptado de: "Is there a scientific method? Education in Chemistry, v. 19, 1982. p.112-116; Disponível em: <https://propi.ifto.edu.br/ocs/index.php/connepi/vii/paper/viewFile/1016/1076>

HUSELID, M. A. **The impact of human resource management practices on turnover, productivity, and corporate financial performance**. Academy of Management Journal, v. 38, n. 3, p. 635-672, 1995. Disponível em: <https://journals.aom.org/doi/abs/10.5465/256741>

JHA, Shashi Bhushan; PANDEY, Vijay; JHA, Rajesh Kumar; BABICEANU, Radu. **Machine Learning Approaches to Real Estate Market Prediction Problem: A Case Study**. 2020. Disponível em: https://www.researchgate.net/publication/343849620_Machine_Learning_Approaches_to_Real_Estate_Market_Prediction_Problem_A_Case_Study

JIANXIONG, T.; XIN, T. Demand Forecast of Human Resources in Starred Hotel Based on Grey Theory and BP Neural Network Combined Model. **Proceedings of the 7th International Conference on Innovation & Management**. Changsha, P.R. China. Acessado em: 25 de ago. 2022.

KE, G. et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. **Advances in Neural Information Processing Systems**, v. 30, 2017. p. 3146-3154. Disponível em: <https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html>

KIELING, E. *et al.* Smartalloc: A model based on Machine Learning for human resource allocation in projects. **WebMedia '19: Proceedings of the 25th Brazillian Symposium on Multimedia and the WebOctober**, 2019, p. 365–368. Disponível em: <https://doi.org/10.1145/3323503.3360643>. Acessado em: 07 de jul. 2022.

LI, J. *et al.* **Brief Introduction of Back Propagation (BP) Neural Network Algorithm and Its Improvement.** Disponível em: https://www.researchgate.net/publication/278653176_Brief_Introduction_of_Back_Propagation_BP_Neural_Network_Algorithm_and_Its_Improvement. Acessado em: 07 de setembro de 2022.

MARCH, S. T.; SMITH, G. F. **Design and natural science research in Information Technology.** Decision Support Systems, v. 15, p. 251-266, 1995. Disponível em: [http://dx.doi.org/10.1016/0167-9236\(94\)00041-2](http://dx.doi.org/10.1016/0167-9236(94)00041-2). Acessado em 07 de setembro de 2022.

MICROSOFT. **LightGBM.** 2019. Disponível em: <https://lightgbm.readthedocs.io/en/latest/>. Acesso em: 03 abr. 2023.

MINISTÉRIO PÚBLICO DO ESTADO DE SÃO PAULO. **Manual de Gestão de Projetos.** Subprocuradoria-Geral de Justiça de Planejamento Institucional. São Paulo, SP: 2017 Disponível em: http://www.mpsp.mp.br/portal/page/portal/Centro_de_Gestao_Estrategica/ManualGestaoProjetos.pdf. Acessado em: 14 de jul. 2022.

O'BROCHTA, M., ASKENAS, R.. **The impact of project management on business results.** Project Management Journal, vol. 34, nº 4, pág. 37-46. Disponível em: <https://journals.sagepub.com/doi/10.1177/875697280303400404>. Acessado em 17/04/2023

ROZENFELD, H. *et al.* **Gestão de desenvolvimento de produto: uma referência para a melhoria do processo.** São Paulo: Saraiva, 2006. Parte I - Cap 1 e 2.

POPPER, K. **A lógica da pesquisa científica.** São Paulo: Ed. Cultrix; 1975. Disponível em: <https://ocondedemontecristo.files.wordpress.com/2011/05/popper-karl-a-logica-da-pesquisa-cientifica.pdf>. Acessado em: 07 de setembro de 2022.

PONDEL, J.; PONDEL, M. **The Concept of Project Management Platform using BI and Big Data Technology.** Wroclaw, Poland. In Proceedings of the 18th International Conference

on Enterprise Information Systems (ICEIS 2016) - Volume 1, 2016, p. 166-173. Acessado em: 08 de ago. 2022.

PROJECT MANAGEMENT INSTITUTE. **Guia PMBOK®**, 7ed. 2013.

SHALEV-SHWARTZ, S.; BEN-DAVID, S. **Understanding Machine Learning: From Theories to Algorithms**. Cambridge University Press, Estados Unidos. 2014. Acessado em: 22 de ago. 2022.

THIOLLENT, M. **Metodologia da Pesquisa-Ação**. Editora Cortez, 2009. Acessado em: 07 de setembro de 2022.

VAN AKEN, J. E. **Management Research Based on the Paradigm of the Design Sciences: The Quest for Field-Tested and Grounded Technological Rules**. Journal of Management Studies, v. 41, n. 2, p. 219-246, 2004. Disponível em: <http://dx.doi.org/10.1111/j.1467-6486.2004.00430.x> Acessado em: 07 de set. de 2023.

WANG, L. et al. On the Sampled Newton Method for Training L1-Regularized Log-Linear Models. **Proceedings of the 32nd International Conference on Machine Learning**. 2015. p. 1418-1426.

WITTEN, I. et al. **Data mining: practical machine learning tools and techniques**. Morgan Kaufmann Publishers, 3ª ed.

WORREN, N.; MOORE, K.; ELLIOTT, R. **When Theories become tools: Toward a Framework for Pragmatic Validity**. Human Relations, v. 55, p. 1227-1250, 2002. Disponível em: <http://dx.doi.org/10.1177/0018726702055010082>. Acessado em: 08 de set. de 2023.

XINGGUANG, L. **A Human Resource Demand Forecasting Method Based on Improved BP Algorithm**. Henan Polytechnic University, Jiaozuo 454003, China. 2022. Disponível em: https://www.researchgate.net/publication/359596507_A_Human_Resource_Demand_Forecasting_Method_Based_on_Improved_BP_Algorithm. Acessado em: 08 de ago. 2022.

ZARAKET, Fadi A.; OLLEIK, Majd; YASSINE, Ali A. **Skill-based framework for optimal software project selection and resource allocation.** *European Journal of Operational Research*, v. 234, p. 308-318, 2013. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S037722171300790X>

ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 67, n. 2, p. 301-320, 2005. Disponível em: <https://rss.onlinelibrary.wiley.com/journal/14679868>