

Aplicação de Técnicas e Mineração de Dados e Machine Learning com Python para Análises Preditivas sobre o Turnover de Funcionários: Estudo de Caso numa empresa multinacional de Óleo e Gás

Danielle Oliveira da Luz
danimarcio.luz@gmail.com
UVA

Alfredo Nazareno Pereira Boente
professor@boente.eti.br
UFRJ

Vinicius Marques da Silva Ferreira
profvmarques@gmail.com
UFRJ

Ricardo Marciano dos Santos
r221070@gmail.com
UFRJ

Resumo: A retenção de profissionais qualificados é um desafio crescente para as empresas, devido à intensa competição por talentos. Por isso, as organizações estão em busca de soluções eficientes que permitam identificar de forma preditiva a redução do turnover, uma vez que isso impacta diretamente nos custos e no alcance da estratégia empresarial. Nesse contexto, os custos associados à rotatividade de funcionários são decorrentes da reposição de funcionários qualificados, principalmente no que tange a despesas com recrutamento, treinamento e desenvolvimento. Assim, a pesquisa teve como objetivo identificar fatores preditivos em relação às principais causas de turnover. Isto posto, foi possível desenvolver planos de ação preventivos para reduzir pedidos de desligamento e, conseqüentemente, reduzir despesas da empresa relacionada à contratação de novos funcionários. Ademais será possível reduzir os riscos de comprometimento da empresa para alcance de seus objetivos estratégicos. Para tanto, foi utilizado o Python como linguagem de programação orientada a objetos para conectar dados de diversas bases de dados estruturados e métodos mineração de dados para preparar a estrutura, tratamento e armazenagem de dados relevantes, necessários para os testes de hipóteses que puderam auxiliar na identificação de fatores que influenciam na

rotatividade de profissionais.

Palavras Chave: Mineração de dados - Aprendizado de Máqui - Python - API - Turnover

1. INTRODUÇÃO

Com o aumento da competitividade por profissionais qualificados, empresas têm buscado formas mais eficientes para aumentar sua capacidade de retenção de funcionários. Nesse cenário buscar soluções que identifiquem de forma preditiva a possibilidade de redução de *turnover* é de extremo interesse para os executivos, pois tem impacto direto nos custos e alcance dos objetivos de negócios das empresas.

De fato, o índice de rotatividade de profissionais acarreta aumento de custo da gestão salarial das empresas, que é normalmente uma das maiores despesas para a organização. Enquanto os profissionais são beneficiados com o melhor reconhecimento, oportunidades de desenvolvimento e remuneração, as empresas investem tempo na busca de profissionais qualificados, desenvolvimento, adaptação à cultura organizacional e na manutenção do patrimônio técnico e intelectual.

Neste cenário, é fundamental desenvolver estratégias eficazes para a retenção de talentos, visando minimizar os custos relacionados à rotatividade, potencializando assim o capital humano dentro das organizações.

O principal objetivo deste artigo é a geração de dados preditivos que possam auxiliar a empresa na identificação de fatores relacionados ao *turnover* de modo a permitir que sejam desenvolvidos planos de ação preventivos para reduzir os pedidos de desligamento e, conseqüentemente, minimizar as despesas da empresa relacionada à contratação de novos funcionários.

2. REFERENCIAL TEÓRICO

2.1. RETENÇÃO DE TALENTOS

Hoje, as empresas vêm intensificando as ações para aumentar a retenção dos seus funcionários, devido um cenário cada vez mais competitivo, onde elas têm dificuldade de atrair, engajar e reter os profissionais adequados para alcançar seus objetivos.

Um dos maiores desafios é a retenção de talentos, a redução da rotatividade de profissionais que tem intenção voluntária de se desligar da empresa e a minimização de custos administrativos referentes a estes processos. Segundo Chiavenato (2010), a rotatividade de pessoal influencia significativamente os processos e resultados da empresa, pois é um aspecto importante na dinâmica organizacional.

Normalmente a rotatividade é representada através do percentual obtido como resultado do cálculo que considera a quantidade de profissionais admitidos e demitidos em relação ao quadro de funcionários totais, conforme descrito abaixo:

$$\frac{[(\text{Número de funcionários Admitidos} + \text{Número funcionários Demitidos})]}{2} \div \text{Número Total de funcionários} \times 100$$

Alguns estudos apontavam que um percentual de *turnover* ideal seja inferior a 10% de rotatividade, considerável saudável para uma empresa, em virtude da renovação de seu quadro de funcionário, de maneira tal que não impactasse o alcance dos objetivos. Existem algumas classificações que ajudam a identificar os diferentes tipos de *turnover*. São eles:

1. Involuntário: Empresa decide realizar o desligamento do funcionário;
2. Voluntário: Funcionário decide tomar a iniciativa de solicitar o desligamento;
3. Funcional: Funcionário com baixa performance solicita o desligamento;

4. Disfuncional: Funcionário com alta performance solicita o desligamento;
5. Inevitável: O pedido de desligamento é impulsionado por razões alheias a ações da empresa, como exemplo, o funcionário deseja realizar a mudança para outro país;
6. Evitável: O pedido é impulsionado pelo fato de o funcionário ter recebido oportunidades de trabalho em outra empresa.

Empresas não dedicam atenção suficiente para calcular os custos indiretos em decorrência da substituição de profissionais, tais como horas necessárias para o recrutamento e seleção, ocasionando, portanto, um mal planejamento para o processo.

2.2. COMPROMETIMENTO DE FUNCIONÁRIOS

Um estudo realizado pela empresa *Willis Towers Watson*, em setembro de 2022, conforme ilustra a Figura 1, mostra alguns indicadores que apresentam o comprometimento de funcionários junto a empresa, o que caracteriza, a redução da probabilidade de ocorrer o *turnover* voluntário disfuncional evitável.

Os empregados que consideram que o seu empregador deu suporte ao bem-estar durante a pandemia estão mais comprometidos e “fidelizados”

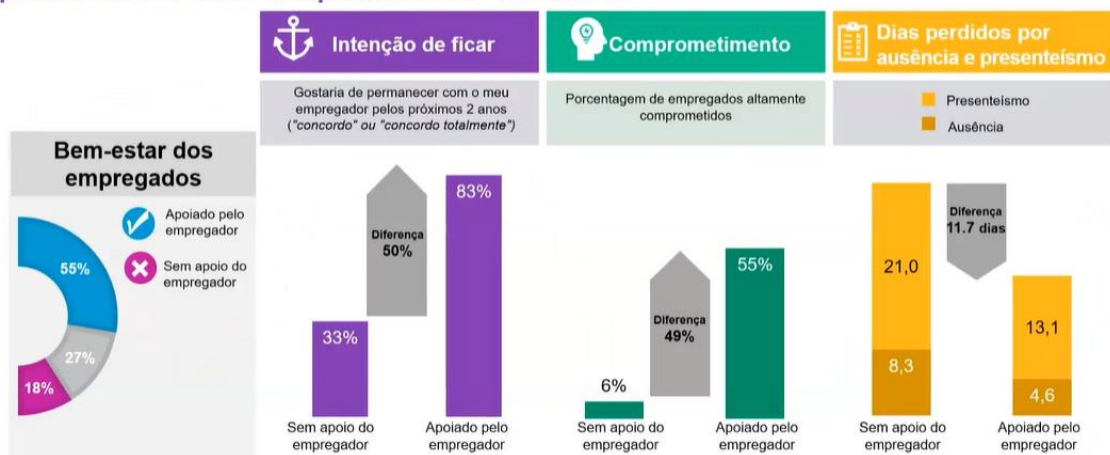


Figura 1: Comprometimento e bem-estar de funcionários.
Fonte: WTW (2022).

2.3. REPENSANDO RETENÇÃO

De acordo com FINNEGAN (2018), alguns fatores são de conhecimento comum, como exemplo, disponibilidade de profissionais qualificados em relação a demanda de mercado e, é natural que o não equilíbrio pode acarretar em mais profissionais a procura de emprego do que ofertas de oportunidades ou quando a oferta de oportunidades é alta pode haver maior demanda para profissionais e, talvez a disponibilidade de profissionais qualificados na mesma área de formação desejada ou com os conhecimentos necessários não sejam suficientes, o que torna mais desafiador a manutenção dos funcionários na empresa.

É de conhecimento comum que podem existir diversas causas que levam os funcionários a decidirem solicitar o desligamento, Klein e Mascarenhas (2016), colaboram com a informação que a remuneração é um dos fatores, porém vai muito além disso.

ALLEN et al., (2010), relacionam também outras razões para que corroboram com a decisão de deixar o emprego, tais como o compromisso com a empresa, oportunidades de crescimento e desenvolvimento individual.

2.4. CUSTO DA ROTATIVIDADE

Profissionais de recursos humanos apresentam alto percentual de *turnover* nas empresas, todavia, é pouco comum correlacionar o custo financeiro causado em relação ao *turnover* propriamente dito (FINNEGAN, 2018).

De fato, qualquer tipo de rotatividade impacta em altos custos para empresas. Reposição de profissionais com baixa performance, é indesejado pois será impactado em encargos sociais como a multa do fundo de garantia por tempo de serviço em torno de 40% de acordo com leis trabalhistas em vigor no momento da conclusão deste projeto. Reposição de profissionais de alta performance devido a pedido de desligamento não possui custos elevados com encargos sociais, porém um dos maiores custos quase que intangíveis é em relação ao desenvolvimento profissional na qualificação do funcionário de acordo com os padrões da empresa. De qualquer forma, a contratação de funcionários devido ao *turnover* poderia ser reduzido se as empresas apresentassem maior dedicação para identificar formas de reter os profissionais qualificados.

Finnegan (2018), desenvolveu uma metodologia própria para representar o custo com o *turnover* e como é possível agir de forma preditiva para ter uma melhor identificação do período de permanência dos funcionários na empresa. O primeiro foco é o custo, pois é a palavra que mais chama atenção dos executivos e do departamento financeiro, pois entende que o custo da rotatividade ou falta de engajamento depende de mudanças que precisam ser iniciadas pelos executivos da organização.

A metodologia, conforme ilustra a Figura 2, sugere que sejam estabelecidos os custos de rotatividade, que passam por uma série de etapas, recrutamento, divulgação da vaga, cobertura das atividades realizadas por outros profissionais, tempo para realização de entrevistas, valor envolvido nas horas dos profissionais envolvidos no recrutamento ou reposição de atividades. Além disso quando um novo profissional é admitido na empresa, ele demanda uma atenção dedicada para entender as formas de funcionamento, cultura e muitas vezes treinamentos específicos fazem parte dessa ambientação inicial até que o profissional comece a dar resultados por si só.



Figura 2: Metodologia Finnegan.
Fonte: FINNEGAN (2018).

Para Finnegan (2018), a proximidade da liderança com sua equipe permite identificar os objetivos dos funcionários, os próximos passos e objetivos de cada um de forma preditiva e transparente, identificar o tempo de permanência na função ou na empresa, sendo essa uma forma prática de planejar o plano de sucessão e definição de mão de obra necessária para entrega dos objetivos da empresa.

2.5. DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS

Goldschmit e Bezerra (2015), definem *Data Mining*, ou mineração e dados, como uma das etapas de descoberta do conhecimento em base de dados, também conhecida como *Knowledge Discovery in Data Base*, que faz parte de um contexto maior que está vinculado a

KDD, contemplando as etapas principais de dados, informações e conhecimento. Para os autores, o valor dos dados armazenados está tipicamente ligado à capacidade de se extrair conhecimento de mais alto nível a partir deles, ou seja, informação útil que sirva para apoio à tomada de decisão, e/ou para exploração e melhor entendimento do fenômeno gerador dos dados. Os dados podem apresentar padrões ou tendências úteis e interessantes que, se descobertos, possuem potencial, por exemplo, para otimizar procedimentos em uma empresa, ajudar na compreensão dos resultados de um experimento científico, auxiliar médicos a interpretar os efeitos de um tratamento, para citar alguns casos.

O processo de *KDD* ressalta a importância de ter uma clara definição do problema através definição de:

1. Conjunto de dados: os dados, em geral o processo *KDD* pressupõe que os dados estejam organizados em uma estrutura tabular dimensional, contendo casos e características do problema a ser analisado.
2. Especialista do domínio: normalmente representa a pessoa ou grupo de pessoas que conhece o assunto e o ambiente que deverá ser realizada a aplicação do *KDD*.
3. Objetivos da Aplicação: compreendem basicamente as tarefas a serem realizadas e características esperadas da aplicação.

2.5.1. ETAPAS OPERACIONAIS DO PROCESSO DE KDD

De acordo com Goldschmit e Bezerra (2015), o processo de *Data Mining* envolve etapas operacionais essenciais, incluindo o pré-processamento, mineração e pós-processamento dos dados, conforme ilustra a Figura 3.

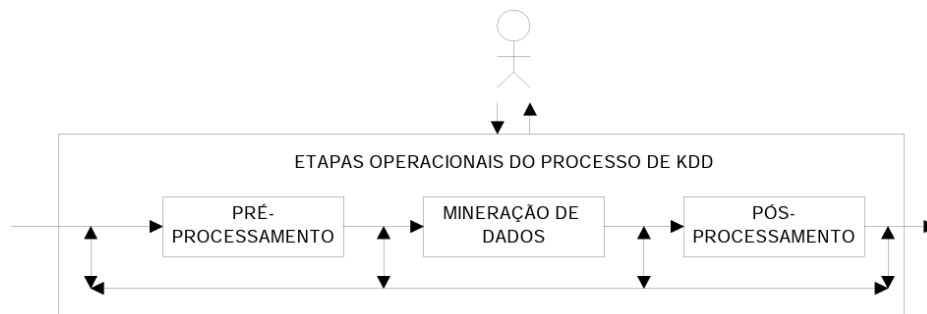


Figura 3: Etapas operacionais do processo de KDD.

Fonte: Boente, Goldschmit e Estrela (2008).

O Pré-processamento é a fase inicial do processo e abrange atividades de captação, organização e tratamento dos dados, com o objetivo de prepará-los para os algoritmos que serão aplicados na etapa de mineração. Segundo Boente, Goldschmit e Estrela (2008), essa etapa é composta por diversas sub etapas, conforme se segue:

Seleção dos dados: Consiste em identificar e selecionar um subconjunto relevante das bases de dados disponíveis para o projeto. Isso permite focar nos dados mais pertinentes e evitar o processamento desnecessário de informações irrelevantes.

Limpeza dos dados: Nessa etapa, o objetivo é garantir a qualidade dos dados representados. Isso envolve a verificação e correção de possíveis erros, como dados ausentes, inconsistentes ou duplicados. A qualidade dos dados é fundamental para obter resultados precisos e confiáveis na etapa de mineração.

Codificação dos dados: Os dados precisam ser codificados de forma apropriada para que possam ser utilizados como entrada pelos algoritmos de mineração de dados. Isso pode

envolver a transformação de variáveis categóricas em numéricas, a normalização dos dados ou outras técnicas de codificação necessárias para a aplicação dos algoritmos.

Enriquecimento dos dados: Essa etapa tem como objetivo agregar mais informações aos registros existentes, tornando-os mais ricos e completos para o processo de descoberta do conhecimento. Isso pode incluir a incorporação de dados externos, a geração de novas variáveis derivadas dos dados existentes ou a obtenção de informações adicionais por meio de fontes complementares.

O pré-processamento dos dados é uma etapa crucial do processo de *Data Mining*, pois influencia diretamente a qualidade e a eficácia dos resultados obtidos na etapa de mineração. Ao realizar essas atividades de forma adequada, é possível garantir dados de alta qualidade e prepará-los para a extração de conhecimento valioso.

2.5.2. MINERAÇÃO DE DADOS

De acordo com Goldschmit e Bezerra (2015), a etapa de mineração de dados é uma parte fundamental do processo de *KDD*. Essa etapa envolve a busca ativa por conhecimentos úteis dentro do contexto da aplicação de *KDD*. Alguns autores associam a mineração de dados à descoberta de conhecimento em bases de dados, considerando-os sinônimos e detalhando-os em diversas tarefas. As principais tarefas da mineração de dados incluem:

Descoberta de associações: O objetivo dessa tarefa é identificar associações entre as informações que são similares nos dados. Isso envolve a identificação de padrões de ocorrência de itens ou eventos, permitindo descobrir relações de dependência entre eles.

Classificação: Consiste em descobrir uma função que mapeie um conjunto de registros em um conjunto de classes. Uma vez descoberta, essa função, chamada de atributo predictor, pode ser aplicada a novos registros para prever a classe a qual esses registros pertencem. Nessa tarefa, o atributo-alvo é uma classe discreta.

Regressão: Essa tarefa busca encontrar uma função que mapeie os registros de um banco de dados em intervalos reais. A diferença em relação à classificação é que o atributo-alvo assume valores numéricos. A regressão é utilizada para fazer previsões de valores contínuos com base em dados históricos.

Agrupamento: O agrupamento é usado para segmentar os registros de uma base de dados em subconjuntos ou clusters, nos quais os elementos compartilham características comuns que os distinguem dos elementos de outros clusters. Essa tarefa é útil para identificar grupos naturais ou padrões intrínsecos nos dados, sem a necessidade de ter classes pré-definidas.

Sumarização: Essa tarefa consiste em identificar e indicar características comuns entre os registros do conjunto de dados. A sumarização ajuda a extrair informações-chave, resumindo os dados de forma a facilitar a compreensão de padrões e tendências.

Detecção de desvios: Essa tarefa tem como objetivo identificar registros do conjunto de dados cujas características não atendem aos padrões considerados normais no contexto analisado. Isso permite a identificação de anomalias, outliers ou comportamentos atípicos que possam ser de interesse para investigação.

Descoberta de sequências: Essa tarefa é uma extensão da descoberta de associações, onde são buscados itens frequentes levando em consideração várias transações ocorridas ao longo de um período. O objetivo é identificar padrões sequenciais ou de sequência temporal nos dados.

Essas tarefas de mineração de dados são fundamentais para explorar e extrair conhecimento valioso a partir dos dados disponíveis, permitindo tomar decisões informadas e obter *insights* significativos.

3. ESTUDO DE CASO NUMA EMPRESA MULTINACIONAL DE ÓLEO E GÁS

Neste trabalho utilizou-se o método *CRISP-DM*, um dos principais modelos utilizados na metodologia *KDD*, através da aplicação de cada fase proposta associando aos objetivos específicos do projeto. Foi utilizada a base de conhecimento adquirida através de atividades realizadas numa empresa multinacional do ramo de Óleo e Gás, onde um dos principais desafios enfrentados é a redução da rotatividade de funcionários qualificados para execução de atividades e, esse setor da indústria apresenta grande dificuldade de formação de profissionais, visto a área de atuação ser muito específica.

Endente-se que essa é uma das etapas mais trabalhosas e que demanda um tempo relevante do projeto, pois é a fase que requer um estudo minucioso das informações disponíveis, incluindo sua estrutura, qualidade e possíveis relações entre os dados. Então, para alcançar o objetivo específico do projeto, foram utilizadas bases de dados de diversas fontes:

- a) Base dados extraídas dos sistemas de informação da empresa;
- b) API para identificação de inflação de determinado período;
- c) O período de interesse do projeto registrado: janeiro/2020 até maio/2023.

As análises preditivas foram realizadas com base em dados extraídos por meio de combinação com acesso a APIs e bibliotecas de *machine learning* do *python*.

Uma vez que as variáveis de interesse foram identificadas, um dos desafios relevantes é preparar os dados, que consiste desde a obtenção dos dados, por meio de coleta eletrônica e APIS, até a investigação de dados internos da própria empresa. Utilizou-se ainda como fonte de dados a API do governo disponível no site do IBGE, conforme ilustra a Figura 4:

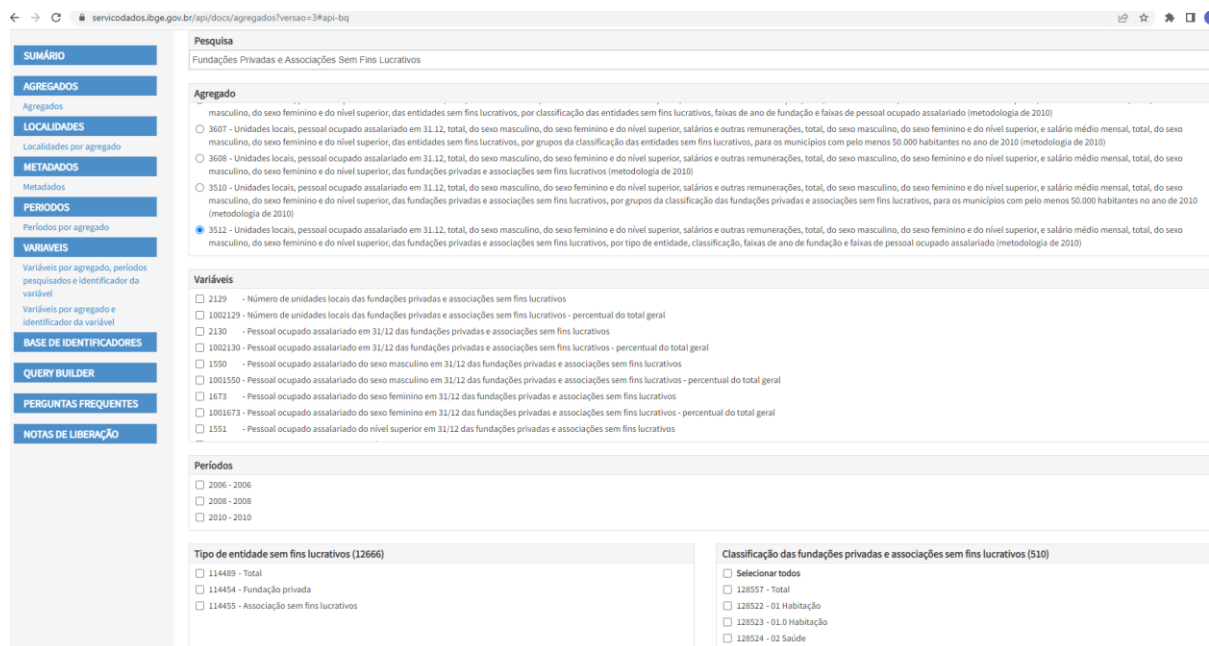


Figura 4: Query Builder.
Fonte: Site Query Builder (2023).

A seguir, foi utilizado o filtro de interesse para obter referência a respeito dos salários e outras remunerações de profissionais de nível superior de fundações privadas e, além disso

foi possível realizar o teste através da utilização do aplicativo *Postman* para checar os resultados obtidos na URL, conforme ilustra a Figura 5:

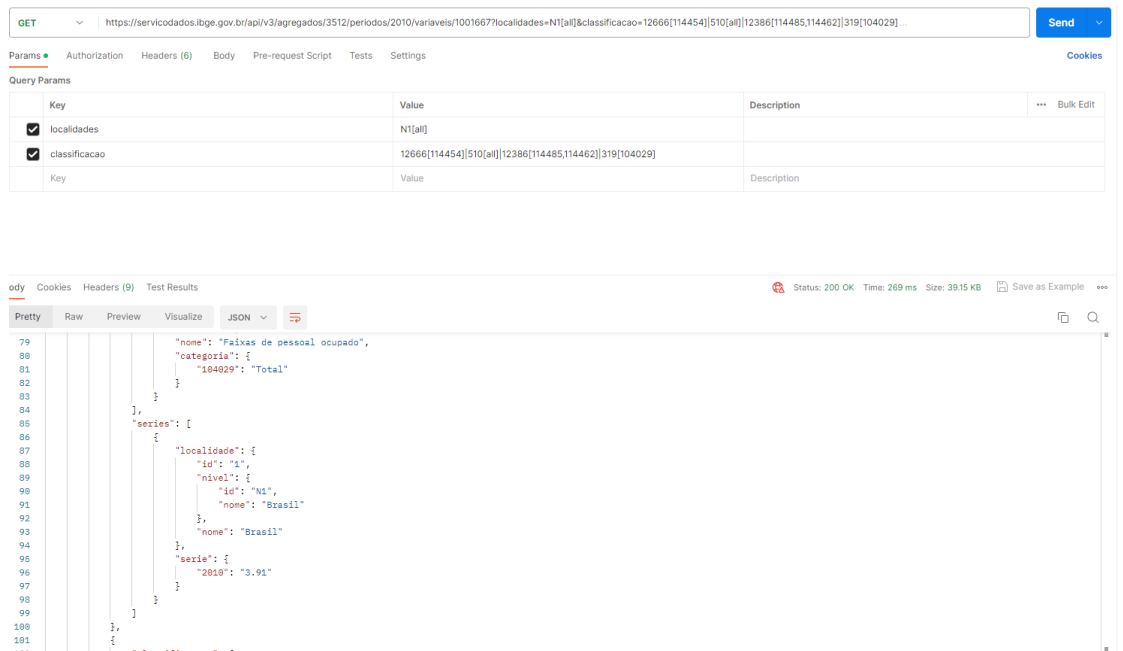


Figura 5: Simulação API do *Postman*.

Fonte: Elaboração própria.

Outra fonte de dados utilizado no projeto foi o *Novo CAGED* através de uma interface desenvolvida em *Power BI*, onde apresentou dados atualizados com 100% de adesão ao interesse do projeto, conforme ilustra a Figura 6:

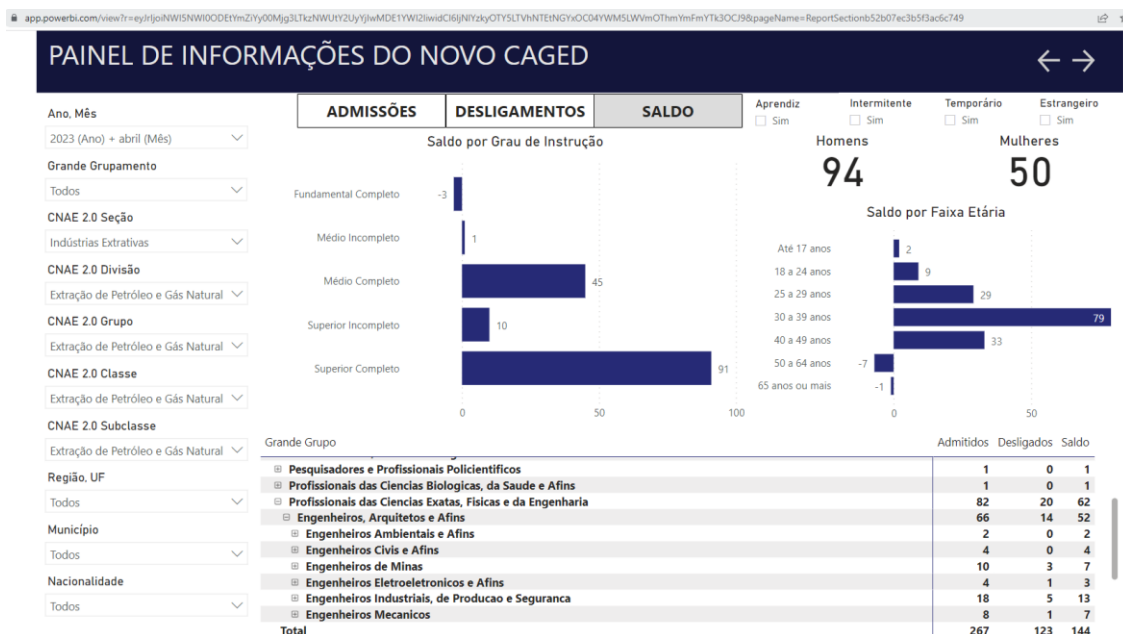


Figura 6: Novo CAGED.

Fonte: Site Novo CAGED (2023).

O desafio, no entanto, foi a extração e obtenção de dados que ainda não estavam disponíveis ao público de forma automatizada, ou seja, não foi possível ser feito até a data

desenvolvimento do projeto em junho de 2023. Embora até o momento da entrega do projeto ainda não tivesse disponibilizada forma *online* de acesso aos dados de maneira que flexibilizasse a utilização da técnica de *webscrapping*, que é a técnica de coleta de dados na internet de forma automática e customizada para armazenar esses dados em arquivos, banco de dados, nuvem ou em qualquer dispositivo válido, a prioridade para tomada de decisão foi a precisão dos dados com base num formato claro e atual, o que permitiu uma melhor análise dos dados disponíveis do governo em relação aos dados da empresa.

A consolidação dos dados históricos obtidos através do site do *Novo CAGED* e da empresa onde foram observadas as variáveis referentes ao percentual de *turnover*, assim como a realização de projeções de percentual de *turnover* para determinado período no futuro, seguiu as seguintes diretrizes: (a) Brasil: Percentual de *turnover* observado no período considerando todo os universos de trabalhadores do Brasil; (b) Setor Óleo e Gás: Percentual de *turnover* de todas as empresas do mesmo ramo de atuação da empresa através da classificação obtida por meio do *CNAE*; (c) Empresa de Óleo e Gás: Percentual de *turnover* registrado da empresa e (d) Função: Percentual de *turnover* de uma função escolhida aleatoriamente.

O primeiro passo, foi projetar um gráfico através do recurso *Matplotlib*, permitindo visualizar e comparar o *turnover* em 4 diferentes perspectivas (Brasil, Óleo e Gás, Empresa e Função) de forma prática e intuitiva, apresentando no eixo x, o período, e, no eixo y, o percentual de *turnover* no período, conforme ilustra o Gráfico 1:

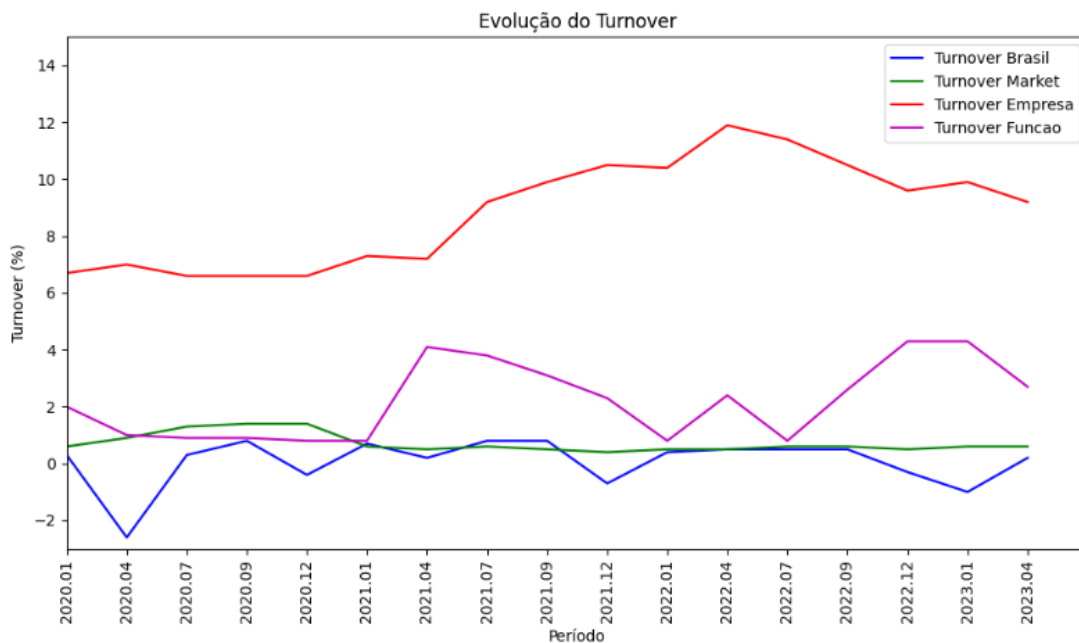


Gráfico 1: Evolução do *Turnover*.

Fonte: Elaboração própria.

Nota-se que o percentual de *turnover* da empresa está significativamente maior que o Setor de Óleo e Gás, e Brasil também. Quando observado o percentual de *turnover* da empresa em relação a uma função específica, nota-se que a função possui *turnover* elevado em relação ao Setor e Brasil, porém não é uma diferença tão significativa.

Através da utilização do método *fit* da biblioteca *SARIMAX* utilizando dados de histórico como base de treinamento aplicando métodos estatísticos para encontrar os valores ideais dos parâmetros do modelo, de modo a minimizar a discrepância entre as previsões do modelo e os dados observados, essa função é importante pois permite que o modelo capture

os padrões e a estrutura dos dados históricos para realizar previsões futuras mais precisas. As projeções de *turnover* são realizadas através do método *predict_mean* da mesma biblioteca, conforme ilustra a Figura 7:

```

Turnover médio período:
Turnover médio Brasil: 0.06
Turnover médio Óleo e Gás: 0.71
Turnover médio Empresa: 8.85
Turnover médio Função: 2.21
-----
Forecast Turnover considerando ano de 2024.12:
Forecast Turnover Brasil: -0.02
Forecast Turnover Óleo e Gás: 0.56
Forecast Turnover Empresa: 9.16
Forecast Turnover Função: 2.31
    
```

Figura 7: Projeções de *turnover*.
Fonte: Elaboração própria.

Os dados dos funcionários que solicitaram desligamento no período do estudo do projeto foram organizados através da utilização dos campos relativos a período, função, sexo, grau de escolaridade, tempo de permanência na empresa e idade do funcionário. Então, a partir desses dados, foram extraídas algumas informações importantes, onde o objetivo foi identificar, classificar e projetar o número de pedido de desligamento dentro do período, conforme ilustra o Gráfico 2:

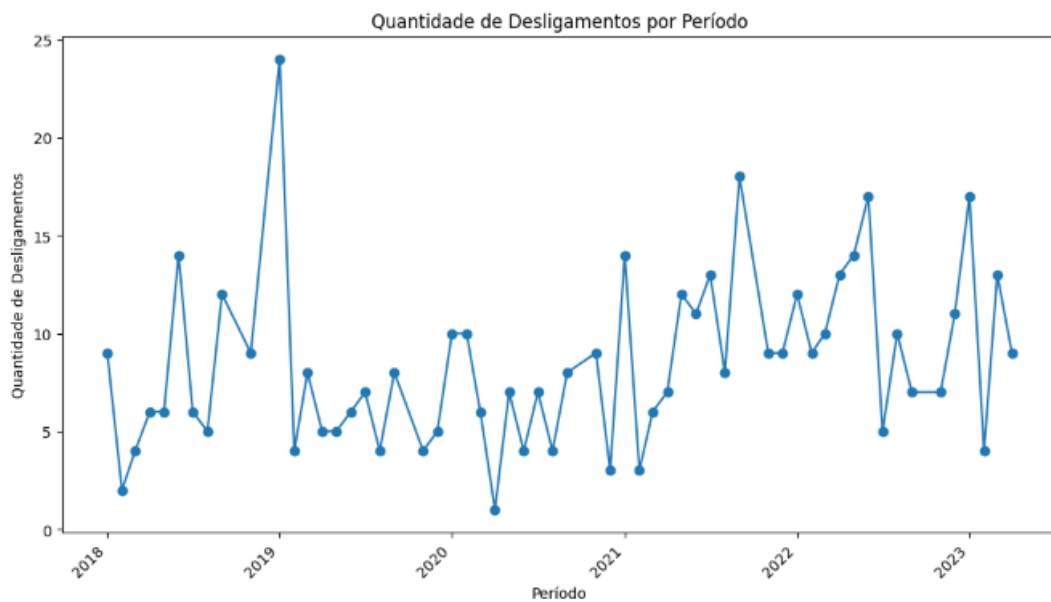


Gráfico 2: Desligamento por período.
Fonte: Elaboração própria.

Em seguida, projetou-se um gráfico com o tempo médio de permanência do funcionário na empresa com escalabilidade baseada na idade do funcionário, conforme ilustra o Gráfico 3:

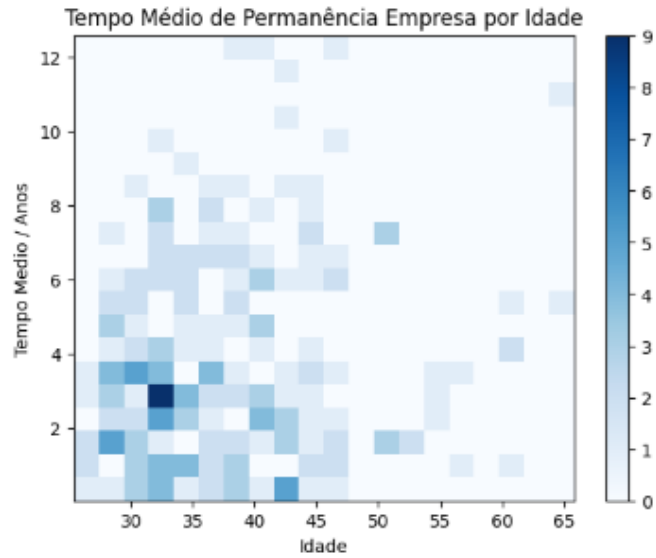


Gráfico 3: Tempo de permanência na empresa por idade.
Fonte: Elaboração própria.

A ilustração do Gráfico 4, mostra um gráfico por regressão linear baseado na idade do funcionário e no tempo de serviço, em anos, desse funcionário na empresa.

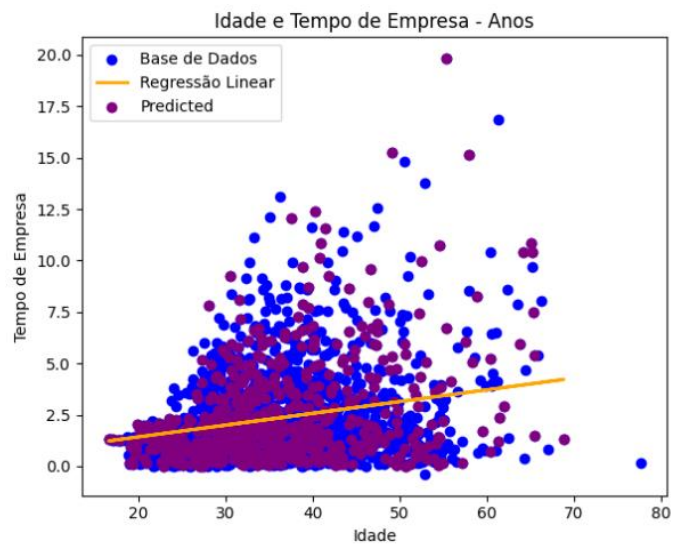


Gráfico 4: Idade e tempo da empresa em anos.
Fonte: Elaboração própria.

Outra extração de dados muito interessante de ser considerada, conforme ilustra o Gráfico 5, mostra uma projeção do tempo de empresa do funcionário e sua respectiva frequência.

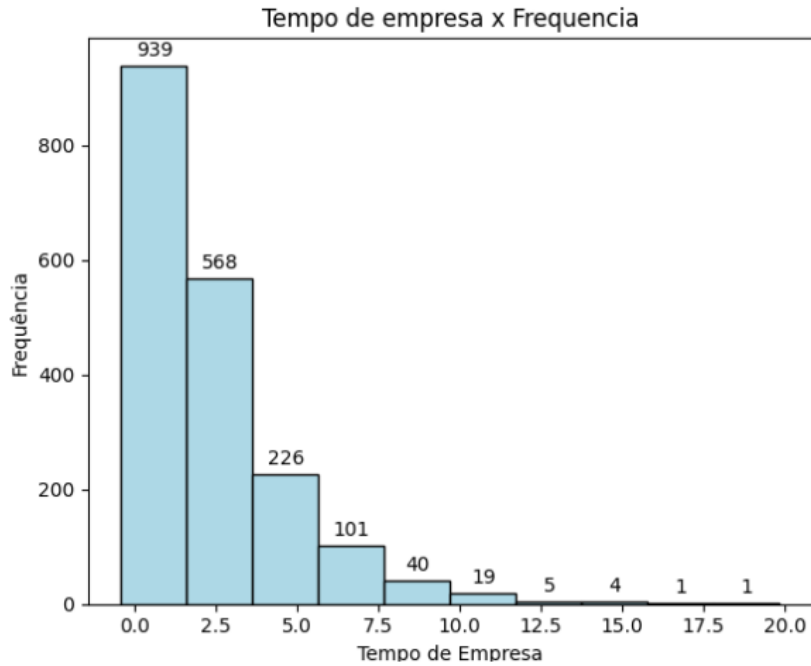


Gráfico 5: Tempo de empresa x frequência.
Fonte: Elaboração própria.

Para a projeção salarial e cálculos do impacto financeiro, foi gerado um gráfico onde o percentual de inflação está no eixo y, e o período, está representado no eixo x, conforme ilustra o Gráfico 6, onde foi registrado o período inicial do projeto, janeiro/2020, e o período final, registrado em maio/2023.

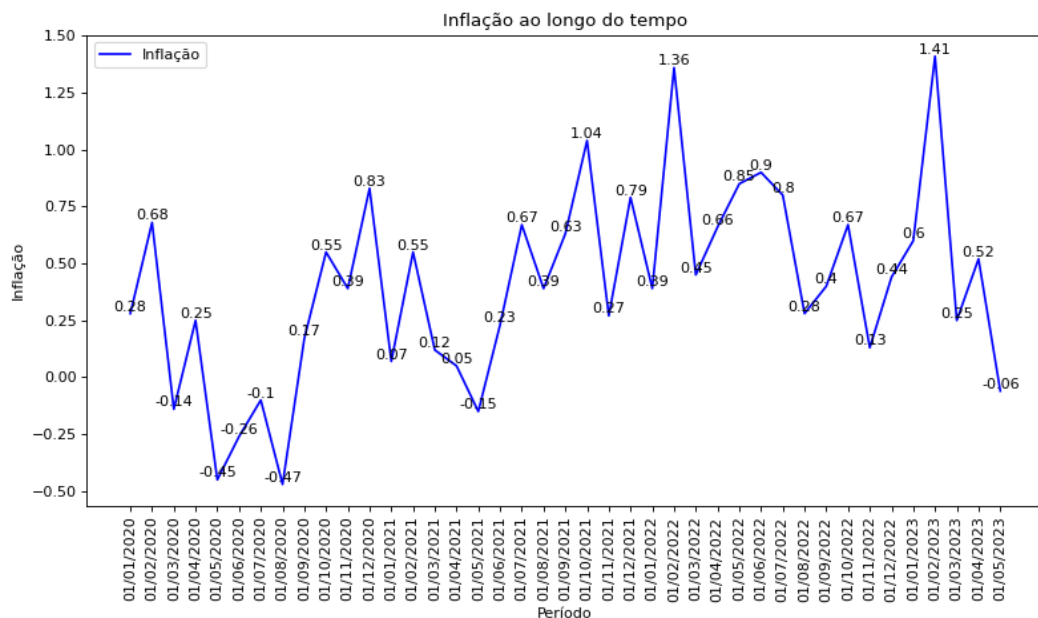


Gráfico 6: Inflação ao longo do tempo.
Fonte: Elaboração própria.

Acessando o arquivo *excel* que contém a base de dados de salários de funcionários durante o período do projeto, foi possível identificar e gerar informações estatísticas, como salário médio, e realizar as projeções de interesse do projeto.

Inflação acumulada a partir de Janeiro / 2023: 2.72 %

Projeções salario em 2024:

Ano	Salario Estimado
0 2024.01	4840.63
1 2024.04	4853.14
2 2024.07	4865.66
3 2024.09	4874.00
4 2024.12	4886.51

Salário médio: 3902.4

Realizando projeções:

Considerando o periodo: 2025.12

Qual salário projeto no período: 5303.66

Projeção do crescimento percentual Dez/2025 em relação ao salario médio atual: 1.36

Projeção da inflação acumulada: 16.73 %

Nova estimativa usando salario e inflação projetada: 6190.87

Projeção do crescimento percentual Dez/2025 em relação ao salario estimado: 1.17

Possibilidade de crescimento salarial considerando Dezembro/2025 em relação ao salário médio atual: 1.59

Figura 8: Projeção salarial com base na inflação prevista.

Fonte: Elaboração própria.

Portanto, utilizando recursos de estatísticas e *machine learning* foi possível obter as seguintes informações, conforme ilustra a Figura 8.

4. ANÁLISE DE RESULTADOS

A partir das análises preditivas realizadas por meio do método *KDD*, técnicas de estatísticas, *machine learn* e projeções de dados, em relação as projeções de percentual de *turnover*, foi possível projetar gráfico de projeção de *turnover* usando o histórico da inflação do período, bem como comparar o percentual do *turnover* considerando dados do Brasil, Setor de Óleo e Gás, empresa e função do funcionário, e projetar os percentuais estimados para essas mesmas variáveis considerando o período sugerido para fins de avaliação da metodologia. Uma vez que analisou-se os percentuais de *turnover* em relação ao Brasil e Setor de Óleo e Gás, observou-se que a empresa possui um percentual de *turnover* mais elevado. Também realizou-se projeções para avaliar as diversas funções da empresa e identificar as funções que apresentaram maior percentual de *turnover*, assim como identificar a média de idade e tempo médio que o funcionário permanece na empresa até solicitar desligamento.

No gráfico de barras teve o objetivo de apresentar a quantidade de funcionários que deixaram a empresa em relação ao tempo de empresa. Também pode-se gerar projeção salarial e impactos financeiros, utilizando como base o conhecimento analisado anteriormente, através da utilização as técnicas de estatística e funções, *fit*, *predict* e *forecast*, em programas *python*, onde identificou-se que a função operador apresentou maior incidência de *turnover* na empresa e, selecionando essa função, foi possível realizar a média salarial da função no período, descrevendo projeções de crescimentos salarial considerando período hipotéticos, além de projetar o salário, num período específico, assim como, a projeção do crescimento salarial na ordem de 36% se comparado com o salário médio atual.

Para uma análise mais robusta, decidiu-se projetar o valor do salário para o mesmo período caso seja considerado o impacto da inflação acumulada desde janeiro/2020 até Dezembro/2025 observando-se que a projeção nesse caso sugere um aumento de 59% para essa função.

5. CONSIDERAÇÕES FINAIS

Pode-se concluir que as projeções realizadas possuem um alto potencial de relevância para serem aplicadas de forma estratégica na empresa. Os dados revelaram um percentual de *turnover* mais elevado quando comparado com o Brasil e o Setor de Óleo e Gás, em que a empresa está inserida, sugerindo a necessidade de verificar se as estratégias de recursos humanos estão alinhadas com os objetivos estratégicos da empresa.

As análises preditivas apresentadas fornecem *insights* importantes para a empresa focar suas ações no aumento da habilidade de retenção dos funcionários, levando em consideração diferentes variáveis, como faixa etária, tempo de empresa e função. Com base nesses resultados, a empresa pode desenvolver programas de desenvolvimento, remuneração e benefícios que sejam atrativos aos funcionários desde o momento da admissão, aumentando assim a probabilidade de mantê-los por mais tempo na empresa.

Ainda, as análises realizadas permitem estimar custos e tempo médio de permanência de funcionários na empresa de acordo com a faixa etária dos profissionais. Isso possibilita um planejamento mais eficiente e direcionado, direcionando recursos de forma específica e evitando investimentos em áreas que não demandam atenção.

Vale ressaltar que os resultados e projeções obtidos neste projeto devem ser avaliados por profissionais especialistas nos diferentes ramos de atuação da empresa, especialmente na gestão de pessoas e recursos humanos. É altamente recomendável que essas análises sejam conciliadas com o planejamento estratégico da organização, garantindo assim o alinhamento das ações de atração, engajamento e retenção com os objetivos e metas estabelecidos.

Embora não seja possível chegar a uma conclusão específica com base apenas nos dados analisados neste projeto, os resultados e projeções sugerem que a empresa deve prestar atenção especial na estratégia de contratação em relação à idade e experiência prévia dos candidatos. É importante compreender que a retenção dos funcionários não é influenciada por um único fator, mas sim pela combinação e correlação de diversos elementos. Portanto, a empresa deve desenvolver estratégias de desenvolvimento, atração e remuneração que sejam adequadas ao perfil de seus funcionários.

Dessa forma, as informações obtidas oferecem suporte significativo para a tomada de decisão da empresa no que diz respeito ao aumento do foco em ferramentas de atração, engajamento e retenção de talentos. Com dados específicos e análises aprofundadas, a organização será capaz de alocar recursos de forma mais precisa e evitar investimentos desnecessários em áreas que não requerem atenção. No entanto, é fundamental que essas conclusões sejam avaliadas e validadas por profissionais especializados, levando em consideração as particularidades da empresa e seu planejamento estratégico.

6. REFERÊNCIAS

ALLEN, D. G.; BRYANT, P. C.; VARDAMAN, J. M. Retaining talent: replacing misconceptions with evidence-based strategies. *The Academy of Management Perspectives*, v. 24, n. 2, p. 48-64, (2010).

BOENTE, A. N. P., GOLDSCHMIT, R. R. e ESTRELA, V.V. Uma metodologia de suporte ao processo de descoberta de conhecimento em bases de dados. In: V SEGeT - Simpósio de Excelência em Gestão e Tecnologia, Resende, 2008.

CHIAVENATO, I. *Gestão de Pessoas*. 3 ed. Rio de Janeiro: Elsevier, (2010).

FINNEGAN, P. R. *Second Edition, Power of Stay Interviews for Engagement and Retention*, USA, SHRM, (2018).

KLEIN, F. A.; MASCARENHAS, A. O. Motivação, satisfação profissional e evasão no 97 serviço público: o caso da carreira de especialistas em Políticas Públicas e Gestão Governamental. *Revista de Administração Pública*, v. 50, n. 1, p. 17-39, (2016)



GOLDSCHMIDT, R. R.; BEZERRA, E. Inicial do nome do meio.” Data mining: conceitos, técnicas, algoritmos, orientações e aplicações. 2ª ed. - Rio de Janeiro: Elsevier, (2015).

TOWERS WATSON. Estudo sobre intenção de desligamento e dificuldades de atração, Brasil, (2022).