

Turnover Management Method Based in Predictive Analysis

Valéria Moura
vrms@cesar.school
CESAR School
Recife, PE, Brazil

Antonio Willian Souza
aws@cesar.school
CESAR School
Recife, PE, Brazil

César França
franssa@ufrpe.br
Universidade Federal Rural de
Pernambuco
Recife, PE, Brazil

Abstract

Context: IT companies struggle with the high statistics of job rotation every year. The high employee turnover in Brazilian software development companies hampers the execution of long-term strategic planning, as well as hinders new business opportunities and software projects due to the loss of intellectual capital. **Problem:** This article aimed to investigate the turnover data of an IT company in Brazil, and develop a data-driven method to predict voluntary turnover, to opportunize the company to act to minimize it. **Solution:** We propose a turnover management process, supported by a predictive model using artificial intelligence (AI). **IS Theory:** Contingency Theory, TAM (Technology Acceptance Model), Knowledge Management, and Organizational Justice are some of the most relevant theories in the relationship between Information Systems (IS) and turnover. Proper management policies can improve communication, engagement, and organizational justice — factors that have a direct impact on employee retention, as showed in this study. **Method:** The research is prescriptive in nature, and its evaluation was conducted through a case study. The results were analyzed using a qualitative approach. **Summary of Results:** the feasibility of applying the proposed process has been confirmed according to each stage of the turnover management process and the results obtained. **Contributions and Impact on the IS Field:** The primary contribution of this work is to assist companies in developing strategies for effective turnover management to minimize it, using a process based on a predictive model supported by AI.

CCS Concepts

- **Computing methodologies** → *Model development and analysis*;
- **Social and professional topics** → **Socio-technical systems**.

Keywords

Turnover, Information Technology, Human Resources Management, Artificial Intelligence

1 Introdução

Com a emergência sanitária do covid-19, o setor de tecnologia da informação cresceu de forma expressiva em capital humano, [8] de acordo com um estudo realizado pelo Banco Interamericano de Desenvolvimento (BID) e pelo LinkedIn [18], este setor registrou uma taxa maior que 60% de contratação quando comparado com o período antes da pandemia. Um outro estudo realizado por [10] em Portugal, defendeu a ideia de que as organizações devem ter como objetivo controlar a rotatividade, o absenteísmo e a intenção de turnover. Estas estratégias construtivas devem existir de forma a reter o seu talento e a sua força de trabalho. A existência de equilíbrio entre os diferentes domínios da vida profissional, familiar

e de lazer, afeta o nível de satisfação e bem-estar dos indivíduos. Uma maior identificação organizacional é baseada na satisfação e menor intenção de turnover e têm um papel fundamental na compreensão da identidade dos indivíduos e na explicação das suas atitudes e comportamentos no local de trabalho.

Há um longo tempo as pesquisas na engenharia de software buscam mapear fatores que podem influenciar no turnover, com a finalidade de melhor informar gestores para que possam evitar ou mitigar os seus impactos [2, 14, 27, 29]. Porém, a popularização da inteligência artificial nos anos recentes possibilitou o surgimento de uma nova gama de estudos mais fortemente fundamentados em dados concretos da indústria [3]. Nesta pesquisa, exploramos a criação de um processo de gerenciamento de turnover, assistido por um modelo preditivo usando inteligência artificial (IA). Esta escolha de tema se justifica pela crescente importância da gestão de recursos humanos nas organizações modernas, bem como pela necessidade de abordagens inovadoras baseadas em dados para lidar com o desafio do turnover voluntário. Através de um processo de construção que seguiu o Design Science Research [4], esta pesquisa partiu de um caso real de uma empresa, passou pela construção de um modelo de predição de Turnover utilizando inteligência artificial, e pela proposição do processo de gerenciamento. Os resultados apontam para uma boa recepção do processo sugerido, por parte do capital humano da empresa estudada.

É importante destacar que, uma vez que o modelo preditivo desenvolvido neste estudo utilizou dados históricos de uma única organização, é provável que ele seja restrito exclusivamente ao comportamento dos dados desta única organização. Por isso, não defendemos neste artigo a generalidade deste modelo. Ao contrário, o principal ponto inovador deste trabalho consiste na proposta de uma organização qualquer possa construir os seus próprios modelos com base nos seus próprios dados como parte de um processo estruturado de gerenciamento do turnover, utilizando nosso trabalho como ponto de partida. O restante deste artigo está organizado da seguinte maneira: a Seção 2 aborda a fundamentação teórica, discutindo os diferentes tipos de turnover e alguns trabalhos relacionados; a Seção 3 detalha a metodologia de design do processo proposto; a Seção 4 apresenta os resultados; e a Seção 5 as conclusões da pesquisa, abordando limitações e trabalhos futuros. Os Apêndices A e B ao final do texto detalham todo o processo de criação e avaliação do modelo de IA utilizado como parte do processo.

2 Fundamentação Teórica

Pensando na retenção dos talentos, as empresas de TI precisam investir fortemente na melhoria do seu quadro de funcionários que atuam com gestão de pessoas. No relatório da PricewaterhouseCoopers [22] são exploradas as tendências e mudanças que moldarão

o futuro da força de trabalho até 2030, incluindo os desafios e oportunidades para os departamentos de RH e sobre gestão de pessoas.

2.1 Gestão de Pessoas

Uma empresa de TI deve ter implantada uma boa área de Gestão de pessoas. Esta área exerce um papel fundamental dentro das organizações, visto que tem como finalidade ser responsável por concatenar diversos processos organizacionais que atendem as demais áreas em uma empresa. Para se adaptar a uma mudança no modo de trabalho, como aconteceu com a pandemia, muitos foram os treinamentos fornecidos aos gestores para que todos fossem preparados para lidar com a liderança remota, gestão de talentos e recrutamento. Neste contexto, [7] indica que: a gestão de pessoas é uma área muito sensível à mentalidade que predomina nas organizações, sendo extremamente contingencial e situacional, uma vez que depende de alguns aspectos, como a cultura e a estrutura organizacional adotada, das características do contexto ambiental, do negócio da organização, da tecnologia adotada, dos processos adotados internamente, do estilo de gestão adotado e de outras variáveis importantes levadas em consideração.

Para [7], “Gestão de pessoas é a função gerencial que tem como objetivo a cooperação das pessoas atuantes nas organizações com o objetivo de alcançar metas organizacionais e individuais”. Segundo [1] apud Walker, 1980), a capacidade de uma empresa em atingir objetivos estratégicos é influenciada pelos recursos humanos em três itens fundamentais: os custos, a capacidade de operar efetivamente e a capacidade de empreendimento e inovação. O resultado da saída de alguns colaboradores e a entrada de outros para substituí-los no trabalho acaba provocando o efeito chamado de turnover. A análise conduzida por [12] traz que alguns dos principais prejuízos desta rotatividade de profissionais se dá pela queda de produtividade, falta de engajamento, comportamentos não produtivos e atitudes negativas com relação à empresa. De acordo com estudos baseados na literatura, estas perdas são medidas através do custo de 90% a 200% do salário anual das principais profissões, como é afirmado por [24] apud P. W. Hom, 2017), e até 213% do salário anual nos casos de cargos altamente especializados [24] apud D. G. Allen, 2010), preocupando gestores de empresas de todos os portes e setores.

2.2 O que é Turnover?

Segundo [28] Turnover é um termo em inglês que define o movimento de entrada e saída dos funcionários em uma organização, independente de ter sido motivado pelo empregador ou por decisão do empregado. Como sinônimo, na língua portuguesa, temos o termo rotatividade. Para fins desta pesquisa, o foco do estudo será o movimento de saída de funcionários.

Existem diferentes tipos de turnover de acordo com [13]. Dentre eles podem ser citados:

- Turnover Involuntário: Quando a organização acha por bem fazer o desligamento do funcionário.
- Turnover Voluntário: Quando por alguma razão, o funcionário pede desligamento da empresa. O Turnover Voluntário se divide em **Funcional**: quando o funcionário não se sente produtivo naquela empresa e pede desligamento em prol de alcançar desafios que os tornem mais produtivos; e **Disfuncional**: quando mesmo sendo um funcionário de alta performance,

o funcionário, não estando satisfeito com os desafios a que é submetido, pede desligamento da empresa.

Em seu estudo [16] afirma que dentro do turnover disfuncional pode-se encontrar o seguinte desdobramento: **Evitável**: trata de situações nas quais a empresa poderia ter atuado de forma efetiva para impedir o desligamento dos colaboradores. Por exemplo, melhoria no plano de carreira ou oferecendo benefícios, para que os colaboradores desejem continuar na empresa. **Inevitável**: são turnovers cujas causas a organização tem pouco ou nenhum controle. Por exemplo, se colaboradores desejam deixar a empresa devido a problemas de saúde ou por um desejo de voltar para os estudos. Neste caso talvez exista pouco o que possa ser feito pela empresa para manter o funcionário.

Portanto, os gestores devem reconhecer que os funcionários são os principais contribuintes para a realização eficiente do objetivo e sucesso da organização [20] apud Abbasi et al. (2000)). Os gerentes devem controlar a rotatividade de funcionários em benefício do sucesso da organização.

De acordo com o resultado da pesquisa de campo realizada por [11], é ressaltada a importância da gestão de turnover para a área de Recursos Humanos e reiterada a importância das organizações terem alguém que faça este controle de rotatividade, uma vez que o bom desempenho e o crescimento de uma empresa dependem dos recursos profissionais competentes que são a base do seu sucesso. A saída de funcionários competentes traz muitos prejuízos para a empresa, e a rotatividade de pessoal acarretará custos e desgastes, pois à medida que os funcionários qualificados saem, as empresas gastam tempo e dinheiro selecionando e desenvolvendo novos funcionários. A taxa de rotatividade é um dos indicadores mais importantes na área de recursos humanos. Atrasos e faltas, falta de comprometimento, falta de motivação e baixo desempenho são as principais razões pelas quais os funcionários são demitidos das organizações. No caso de pedido de demissão, os funcionários buscam obter uma posição melhor na concorrência e/ou fazer uma mudança no campo profissional. A conclusão é que o turnover é uma importante ferramenta de gestão. Para que uma organização não perca um bom funcionário, é importante determinar as circunstâncias em que ocorreu uma demissão. É necessário criar um ambiente de trabalho produtivo onde os funcionários sejam respeitados e se sintam seguros. Saber administrar conflitos, criar novas estruturas de planejamento de carreira e reforçar o feedback, são formas de reter os colaboradores, pois eles se sentem mais motivados a manter níveis de produção de alta qualidade.

Segundo [15], o cálculo do turnover é dado pela soma dos funcionários admitidos e desligados e o resultado dividido pela quantidade total de funcionários da empresa.

2.3 Trabalhos Relacionados

Esta sessão tem como objetivo demonstrar um resumo de alguns estudos de trabalhos relacionados com esta pesquisa e que puderam contribuir para o direcionamento dado a este trabalho. Alguns possuem um grau de similaridade maior com o tema da dissertação aqui descrita e outros nem tanto. No entanto, nenhum deles teve como objetivo a definição de um processo de gestão de turnover para facilitar a aplicação de um plano de ação para redução do

turnover voluntário em empresa de software, que é o objetivo deste trabalho.

[23] aborda a previsão da saída voluntária de funcionários de uma empresa. Propõem uma abordagem baseada em dados do RH para analisar a rotatividade, sem pesquisa adicional. [25], por sua vez, visa contribuir para a gestão do risco de rotatividade, propondo um modelo de previsão com base em características demográficas disponíveis nos sistemas de recursos humanos. Já em [30], os autores exploram métodos de aprendizado de máquina supervisionados para prever a rotatividade de funcionários em organizações. Em [21], é apresentado um modelo de regressão logística múltipla para a criação de um indicador que mede a intenção de turnover voluntário de cada funcionário de uma companhia. A maioria dos trabalhos relacionados tiveram influência relevante com este estudo no que tange a variáveis utilizadas e alguns métodos preditivos em comum. Ainda podem ser encontradas outras pesquisas similares em [6, 9].

3 Metodologia Científica

Dado que o Turnover é um problema enfrentado por muitas empresas, esta seção tem o objetivo de sistematizar a gestão do turnover através de um processo baseado em dados de uma empresa do ramo de tecnologia no nordeste do Brasil. Este trabalho se propõe a ajudar na identificação de pessoas que tenham maior propensão a realizar um pedido de desligamento voluntário, mas também em ajudar a identificar os principais fatores que guiaram o turnover voluntário na empresa estudada, através de um modelo preditivo. Como parte do processo é importante identificar a propensão e os fatores motivacionais. Neste trabalho faremos isso por meio do modelo preditivo e da análise de dados.

3.1 Design Science Research

Com o objetivo de realizar uma estruturação robusta a ser seguida no desenvolvimento do processo proposto neste trabalho, buscamos adotar uma metodologia de Design Science Research. A metodologia de Design Science Research é uma abordagem metodológica que orienta como pode-se criar artefatos para resolver determinados problemas em contextos específicos através de um meio sistemático de conhecimento. Como é um método orientado a problema, se preocupa em entender como as coisas funcionam, transforma uma situação estabelecendo a operacionalização de artefatos ou recomendações para resolver os problemas, entendendo sua relevância e identificando as soluções propostas partindo de uma situação para chegar a uma solução através de experimentações por meio de ciclos [4]. Neste trabalho não foi usado nenhuma metodologia exploratória de dados como CRISPDM uma vez que esta pesquisa trata do uso do Design Science Research para a definição do processo de gerenciamento do turnover, objetivo deste trabalho, e não para apoiar no modelo preditivo.

3.1.1 Identificação do Problema (Problem Identification). Nesta fase foi feita a revisão da literatura e diante da questão da alta de turnover na empresa estudada, foi percebida a necessidade de entender o problema a fim de buscar alguma forma de controlá-lo. O objetivo foi averiguar se existia alguma forma de prever os próximos pedidos de desligamento e poder proporcionar algo efetivo para que os funcionários pudessem permanecer na empresa e não fazer um pedido de demissão. A empresa estudada enfrentou um turnover

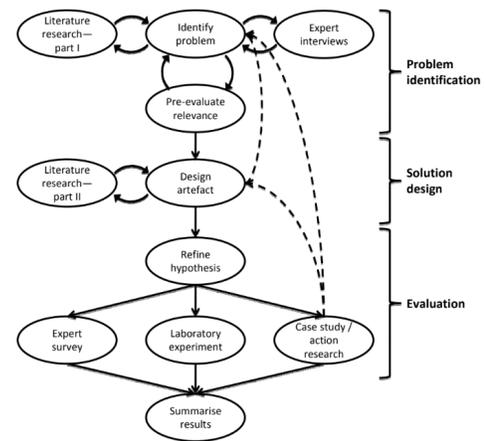


Figure 1: Etapas da Design Science Research [19]

alto no período pós-pandemia e para isso, será apresentado como resultado deste estudo um processo definido para ajudar o setor de capital humano desta empresa e de outras empresas que desejarem, a fazer uma melhor gestão do turnover dos seus colaboradores. No capítulo de Resultados será possível entender os resultados obtidos na construção do processo que foi definido para ajudar na gestão do turnover com suas fases e atividades a serem executadas, de forma mais detalhada, bem como entender todo o tratamento dado aos dados para o modelo preditivo. Para fazer uso deste processo, é importante salientar que é muito importante definir o melhor momento para sua execução. Minimamente foi sugerido um período de análise mensal onde o capital humano da empresa pudesse consultar estes dados de predição e definir o plano de ação desejado.

Detalhamento do Contexto. Para que os fatores do aumento do turnover pudessem ser avaliados através da execução do processo de gerenciamento de turnover e trouxesse os resultados desejados, inicialmente foram coletados os dados existentes de todos os funcionários da empresa entre os anos de 2002 a 2022 por cargo e posteriormente feita uma análise dos dados dos funcionários que pediram demissão entre 2019 e 2022 levando em consideração vários aspectos como quantidade de anos de empresa, modelo de trabalho, entre outras questões. Os resultados obtidos nesta análise podem ser vistos no apêndice B.

3.1.2 Design da Solução (Solution Design). Esta parte do Design Science Research aborda a fase da solução de design utilizada para a criação do processo de gestão do turnover e do modelo preditivo. Todo o design do processo de gerenciamento de turnover e do modelo preditivo que foi usado como base para a execução do processo, foi definido através de pesquisas na literatura. Inicialmente iremos explicar como foi feito todo o processo do modelo preditivo para depois explicar as fases e fluxos do processo de gestão do turnover. O objetivo geral do trabalho aqui apresentado foi o de sistematizar a gestão do turnover baseado nos fatores que motivaram os funcionários de uma empresa de tecnologia a deixarem os seus empregos. Para isso foi realizada uma análise preditiva através da

coleta de dados de todos os funcionários que ainda estão na empresa e dos que pediram desligamento entre os anos de 2019 e 2022, respeitando as regras de anonimidade requeridas pela LGPD ¹, onde o nome do colaborador foi trocado pela palavra Persona 1 e alguns outros dados que identificassem unicamente um colaborador, também foram convertidos para que não permitissem sua identificação. Através da ciência de dados, foram realizadas várias análises dos indicadores.

Construção do Modelo. Para a análise do turnover voluntário foi feita a aquisição dos dados de modo a obter informações dos colaboradores que pediram desligamento de forma voluntária considerando o período de quatro anos, sendo especificamente o período de 2019 até 2022. Também foram feitas entrevistas com especialistas do capital humano para embasar os dados considerados como entrada para o processo de gerenciamento de turnover. Foram avaliadas algumas variáveis categóricas e numéricas, de modo a propor um modelo preditivo que atuasse não apenas na identificação de pessoas que tinham maior propensão a realizar o pedido de desligamento, mas também nos principais fatores que guiaram o turnover voluntário na empresa estudada. Os dados de entrada de cada empresa servirão de direcionamento para o resultado que a empresa deseja obter com o uso no processo de gestão de turnover. Neste período foi realizada uma filtragem dos dados para considerar: i) funcionário em regime de trabalho CLT e ii) pedidos de desligamento por iniciativa do(a) funcionário(a). O pré-processamento foi realizado da seguinte forma: separação dos dados em variáveis categóricas, e booleanas. Os dados trouxeram informações como histórico salarial, dados de mobilidade, dados de carreira, dados de admissão e demais dados para uma avaliação posterior e que poderia trazer como resultado, a causa raiz dos índices de desligamento na empresa. O propósito deste estudo foi entender os elementos que motivaram os profissionais a continuarem em seus empregos ou a deixá-los e para isso foram realizadas as seguintes etapas: (i) *Criação de um dataset*: onde foi realizado o pré-processamento dos dados brutos; (ii) *Análise exploratória*: considerando o dataset gerado na etapa anterior foram identificados padrões relevantes que pudessem explicar o turnover; (iii) *Modelagem Preditiva*: neste momento foi criado um modelo preditivo a fim de classificar em Sim(1) ou Não(0) o pedido de demissão definindo assim uma probabilidade de certeza dessa classificação; e (iv) *API*: Esta etapa teve como objetivo a criação de uma API para identificar o turnover a partir de uma lista de funcionários.

Como resultado do trabalho de predição, foram sugeridas as variáveis mais relevantes para o pedido de demissão e uma proposta de processo que visa ajudar a identificar as possibilidades de turnover baseada na lista de funcionários ainda existentes na empresa. Este processo será melhor detalhado na seção de Resultados.

Design do Processo de Gestão do Turnover. O design do processo de gestão de turnover foi criado com fluxos definidos, com etapas de entrada e saída, a partir de estudos da literatura e tomando como base o resultado do modelo preditivo. Para esta fase, além do Design Science Research também foi analisado o método de Design Thinking para ajudar a identificar as fases e soluções a serem propostas.

Segundo [5], o design thinking é composto por 3 estágios: *Inspiração*, quando uma oportunidade é identificada; *Ideação*, quando se concebe algumas soluções gerais; e *Implementação*, onde a ideia é implementada.

É importante salientar que para a construção deste processo, foram analisadas as duas metodologias (Design Science Research e Design Thinking) mas a metodologia utilizada para a criação do processo seguiu o Design Science Research, uma vez que de acordo com [17], o Design Thinking tem como objetivo se concentrar mais na criação de uma experiência do usuário que inclui um artefato de TI, enquanto o Design Science Research se concentra na criação de um melhor artefato de TI e implica que a experiência do usuário decorre de um sistema melhor projetado.

3.1.3 Avaliação da Solução (Evaluation). Esta é a etapa do Design Science Research que aborda a avaliação do processo e do modelo preditivo. Esta fase de avaliação foi o momento em que um grupo focal composto por profissionais experientes do Capital Humano da empresa onde o estudo foi realizado, fez a avaliação do estudo a fim de verificar a coerência e a viabilidade do resultado. Para isso foram conduzidas entrevistas e criação de grupo focal para através dos resultados obtidos no caso de estudo, entender de forma resumida os feedbacks recebidos para ir retroalimentando o processo e consequentemente as ações a serem planejadas pela empresa a fim de reduzir o seu turnover.

3.2 Processo de Gerenciamento Do Turnover

Para a definição do processo de Gerenciamento de Turnover, onde toda a lógica de machine learning foi executada, foi feito o uso do ciclo PDCA ². O ciclo PDCA é uma ferramenta de gestão de processos que facilita a melhoria contínua das atividades empresariais por meio de quatro etapas claramente definidas: Planejamento, Execução, Controle e Ação. Para cada uma das três fases da metodologia de Design Science Research, foram definidas as seguintes etapas do processo de acordo com o ciclo PDCA e que devem ser seguidas pelas empresas a fim de nortear todo o fluxo de trabalho conforme pode ser visto na Figura 2.

Qualquer empresa pode usar o processo de gestão do turnover desenvolvido neste trabalho, e para isso é possível que seja executada a estrutura do processo desde o fluxo do planejamento, passando pela fase de execução até o controle. A fase de ação poderá ser definida e executada de acordo com a realidade de cada empresa.

1) *Planejamento: Definição da Periodicidade* - O processo apresentado pode ter início na periodicidade sugerida pelo capital humano da empresa. É importante estabelecer períodos cíclicos de execução que mais se adequem à realidade da empresa. *Aquisição dos Dados* - O processo inicia sua execução de fato a partir do acesso aos dados dos funcionários que pediram demissão de forma voluntária. Os dados trazem informações como histórico salarial, dados de mobilidade, dados de carreira, dados de admissão e demais dados que a empresa julgar necessário para uma avaliação posterior e que pode trazer como resultado, a causa raiz dos índices de desligamento na empresa. *Governança dos Dados de Colaboradores* - Os dados precisam respeitar a LGPD e com isso é feita uma Governança nos dados dos colaboradores desde o momento da coleta, garantindo

¹<https://www.gov.br/mds/pt-br/aceso-a-informacao/lgpd>

²<https://www.siteware.com.br/metodologias/ciclo-pdca/>

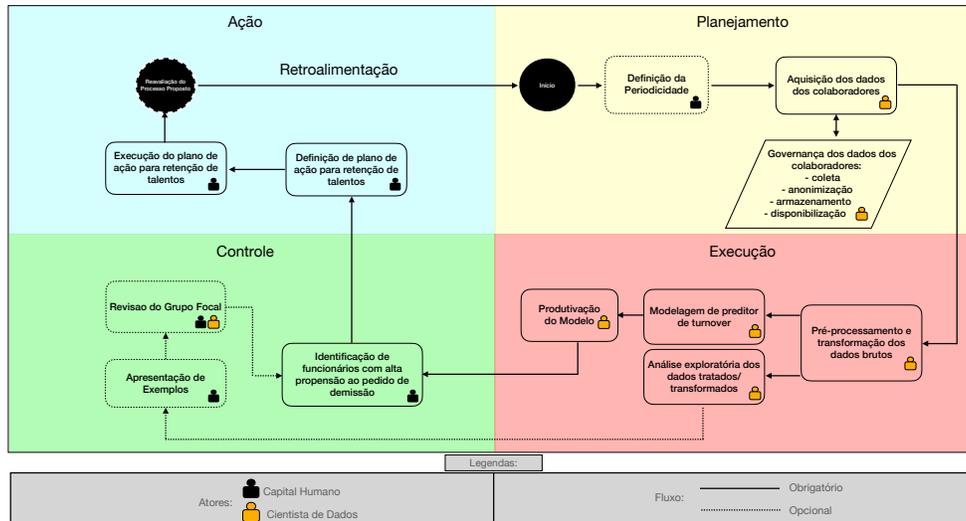


Figure 2: Processo de Gestão do Turnover.

sua anonimização. É necessário um cuidado com o armazenamento desta informação que deve seguir padrões rígidos de segurança bem como a manutenção de sua integridade. Em seguida é feita a disponibilização do dado para o pré-processamento.

2) Execução: Pré-processamento e Transformação dos Dados Brutos - Durante a fase do pré - processamento, é feito todo o trabalho de transformação dos dados brutos em dados mais refinados. É necessária muita cautela na hora de manusear estes dados a fim de não mudar sua integridade e consequentemente não ocasionar um desvio nos resultados. Modelagem de Preditor de Turnover - Logo após o pré-processamento, uma atividade importante a ser realizada é uma modelagem para entender quais variáveis podem ser significativas como preditor do turnover. Análise Exploratória nos Dados Tratados - Após o pré-processamento é feita uma análise exploratória em cima dos dados transformados para que sejam removidos aqueles que não trarão nenhum resultado válido ou que não contribuirão de forma significativa para o estudo. Nesta etapa também podem ser geradas amostras representativas dos perfis que servirão de exemplos na fase de Controle. Produtivação do Modelo - Em seguida é possível produtivizar o resultado do modelo que pode ser feito de várias formas. Um exemplo mais simples é através da criação de uma API onde os dados obtidos no modelo preditivo servem de entrada e ao fazer seu deploy, os dados podem ser consultados a qualquer momento pelos profissionais do capital humano da empresa.

3) Controle: Apresentação de Exemplos - A partir deste momento poderão ser apresentados alguns exemplos de amostras representativas dos perfis (personas) propensos a pedir demissão para servir para uma posterior análise e definição de um plano de ação para

retenção de talentos. Identificação de funcionários com alta propensão ao pedido de demissão - Logo após a apresentação de Exemplos, o setor que é responsável pela gestão do turnover na empresa, com base no resultado da pesquisa, fará a identificação dos funcionários com alta propensão ao pedido de demissão para um posterior trabalho de prevenção da saída do colaborador. Realização de Grupo Focal - Pode ser realizado um grupo focal com o capital humano da empresa com a intenção de validar os atributos encontrados pela execução do modelo preditivo.

4) Ação: Definição do Plano de Ação para Retenção de Talentos - Com base na Definição / Execução / Redefinição é possível executar um plano de ação para retenção de talentos. Nesta etapa os representantes do capital humano podem escolher fazer uso ou não de amostras representativas dos perfis (personas) propensos a pedir demissão ou fazer uso somente do resultado do modelo preditivo obtido com a produtivação do modelo para servir como base na definição de um plano de ação para retenção de talentos. Como parte do plano de ação, a empresa pode definir uma OKR 3 e fazer medições regulares a fim de avaliar os resultados obtidos na fase anterior. Os objetivos são descritos de forma a mensurar qualitativamente o que se deseja alcançar e precisam motivar e desafiar a equipe. Já as Key Results ³ são um conjunto de métricas que existem com a finalidade de medir o progresso da OKR em direção ao objetivo de forma qualitativa. É importante que para cada Objetivo exista um conjunto de 2 a 5 resultados principais para que sejam fáceis de serem lembrados. Execução do Plano de Ação para Retenção de Talentos - Uma vez definido o plano de ação para retenção de talentos, este é o momento de colocá-lo em prática

³<https://www.feedz.com.br/blog/key-results/>

seguinto todo o planejamento definido. É importante observar que todo processo é cíclico e sua retroalimentação pode ser realizada sempre que for necessário fazer algum ajuste nas variáveis a serem consideradas, dando início a um novo ciclo.

3.3 Avaliação

Esta fase de avaliação tem como objetivo avaliar diferentes aspectos: o modelo do ponto de vista técnico, de conhecimento de domínio, e o processo. O método de design da solução, objetivo deste trabalho, foi executado seguindo o processo. Fizemos uma avaliação das fases do Planejamento e Execução do PDCA com um cientista de dados e avaliamos o resultado obtido, que informava que a variável alvo mais representativa na decisão de pedir demissão, era o seguro de vida. Na fase de Controle foi feita entrevista com uma especialista do Capital Humano da empresa sobre este resultado. Ao questionarmos esse resultado, ela informou que este campo era zerado quando a pessoa pedia demissão. Sendo assim, mascarava a informação de que quem pedia demissão não possuía seguro de vida. Uma outra informação que se mostrou diferente dos dados extraídos, foi que quanto mais tempo de casa a pessoa tinha, mais propensa ela era de pedir demissão. Isso não foi confirmado com entrevistas com especialistas do capital humano. Pelo contrário, pudemos avaliar que a faixa onde o pedido de demissão foi mais acentuado foi com até 4 anos de empresa. Desta forma, decidimos retroalimentar o processo e iniciar um novo ciclo, onde pudemos ajustar alguns dados que ainda precisaram de tratamentos, como campos de datas com formatos variados. A seção de resultados trará os resultados do segundo ciclo de execução do processo após a validação das informações através do grupo focal com o capital humano da empresa. Vale salientar que o modelo é algo instrumental dentro do processo de gerenciamento, e ele não precisa ter a capacidade de ser transferido pra outras organizações! Cada uma precisa criar o seu próprio modelo, com base nos seus dados!

4 Resultados

Percebemos pelos resultados do modelo na segunda execução do processo e posteriormente pela avaliação do Capital Humano da empresa estudada, através de um grupo focal, que dentre as pessoas que pediram demissão e que o modelo previu que pediriam demissão, o valor salarial, mudança e progressão salarial e o nível do cargo parece ter tido maior influência no pedido de demissão do que o tempo de casa. Uma sugestão para o Capital Humano da empresa seria pegar o resultado do modelo e comparar as variáveis com os dados das pessoas que pediram demissão. Isso alertaria para a definição de ações que precisariam ser feitas com as pessoas da empresa, candidatas a pedir demissão. Após estes resultados foi feita uma validação através do grupo focal com o Capital Humano da empresa sobre as seguintes hipóteses: Possibilidade de colaboradores que não têm quinquênio⁴, pedirem mais demissão. Constatando que pessoas com menor tempo de casa são as que mais pedem demissão. O capital humano concordou que seria uma hipótese válida. Importância do banco de horas negativo como motivador a um pedido de demissão. Não foi constatada uma coerência lógica que confirmasse a hipótese como principal fator de pedido de demissão, mas teria influência somado com outras variáveis. Foi

levantado que até teria uma influência maior se o banco de horas fosse muito positivo, constatando que existiu uma sobrecarga de trabalho e o colaborador pode desejar mudar de carreira ou de empresa que forneça melhores condições de trabalho. Foi trazido através do grupo focal alguns fatores motivantes levantados nas entrevistas de desligamento (levantamento qualitativo) que mais pesaram na decisão de pedir demissão. São eles: o pessoal mais jovem, os milênios, salários mais atrativos ou propostas de salário mais arrojadas, dificuldade de dar fit com a cultura da empresa por não ter tido tempo de vivência devido ao home office, pessoas com mais de 5 anos de casa, cuja motivação do pedido de demissão foi viver outras experiências em empresas de outros segmentos, ou até com empresas de mesmo segmento mas mais conhecidas mundialmente. De acordo com cada uma das etapas do processo de gestão do turnover com apoio do PDCA, os dados obtidos foram coletados, os datasets foram criados, modelos de classificação foram apresentados para o grupo focal e dado os resultados obtidos pelo modelo, em comparação com os feedbacks do grupo focal, pode-se constatar a viabilidade da aplicação do processo proposto.

5 Conclusão

Este trabalho teve como objetivo geral prover um processo de gerenciamento de turnover a partir da análise das suas principais causas em uma empresa de software. Para isso teve como objetivo específico, explicado pela figura 3 a criação de um processo sistemático capaz de identificar o perfil de pessoas com alta propensão ao turnover voluntário, com a ajuda de um modelo preditivo de inteligência artificial. A criação deste processo de gerenciamento de turnover foi baseado numa metodologia científica de design chamada DSR (Design Science Research) fazendo uso do modelo de gestão PDCA. Foi feito uso do PDCA por se tratar de um método de melhoria contínua que é um fator primordial para ir refinando o processo. Baseado em dados históricos da empresa, através da execução do processo sugerido neste estudo e através de predições realizadas por um modelo preditivo, podem ser feitas análises dos indicadores por parte do setor de capital humano da empresa com o objetivo de entender o que pode nortear o pedido de demissão de seus colaboradores. Como indicado anteriormente, o modelo preditivo desenvolvido nesta pesquisa não tem o intuito de ser generalista, ou aplicável diretamente à outras organizações. A proposta deste estudo foi tão somente evidenciar que é possível integrar a prática de análise de dados e construção de modelos preditivos a um processo bem estruturado de gestão de turnover organizacional. No entanto, pesquisas futuras podem se beneficiar da integração de modelos gerados em diferentes organizações em busca de algo que tenha efetividade genérica pelo menos dentro deste grupo de empresas de tecnologia da informação similares entre si.

5.0.1 Limitações. O estudo, baseado em análise preditiva de dados internos de turnover em uma empresa de tecnologia, apresenta limitações que afetam a generalização dos resultados. Destacam-se o viés de seleção e a representatividade restrita da amostra, além de problemas metodológicos e de qualidade dos dados. Fatores importantes, como satisfação no trabalho e cultura organizacional, podem ter sido subestimados, e inconsistências nos registros comprometem a precisão. É importante reforçar também que este trabalhos

⁴<https://www.jusbrasil.com.br/artigos/o-que-e-quinquenio/805141122>

lida com a tentativa de usar tecnologias para predição de comportamento de pessoas, e assim como toda tecnologia e conhecimento desta natureza, deve ser utilizada de maneira estreitamente ética pelos gestores, para que não causem quaisquer tipos de prejuízos aos seus colaboradores. A ausência de informações subjetivas complementares limita a compreensão mais profunda das causas da rotatividade, e poderiam ser exploradas em pesquisas futuras. No entanto, ciente de que dados de natureza subjetiva também podem sofrer limitações, como da falta de honestidade de colaboradores na hora da entrevista de desligamento, por exemplo, o que pode afetar diretamente a qualidade dos dados que dizem respeito ao real motivo do pedido de demissão.

5.0.2 Trabalhos Futuros. Uma necessidade apresentada pelo Capital Humano através do grupo focal foi que este trabalho pode ser usado para fazer uma avaliação mais estratificada por perfil (desenvolvedor, engenheiro de testes, etc), por carreira (Júnior, Pleno, Sênior) e por gênero já que na área de TI, mulheres podem ser mais afetadas que homens. Se o processo for executado em locais diferentes é possível gerar um guia de boas práticas de retenção de talentos com base na evolução da análise dos dados reais dos achados em cada empresa. Uma sugestão para trabalhos futuros seria pesquisar como melhorar os sistemas de RH a capturar as informações relevantes que de fato podem impactar a decisão de demissão voluntária. Também pode ser feito uso de ferramentas como fonte de pesquisa como o Glassdor, que tenta cobrir o gap com relação aos motivos que levaram o funcionário a pedir demissão, já que este pode querer manter um bom relacionamento com a empresa, evitando dar opiniões mais sinceras quando no Glassdor suas opiniões podem ser compartilhadas anonimamente. Esta pesquisa também pode ser expandida para incluir a área de gestão de pessoas e como esta área trata o problema do turnover. Os tópicos de interesse do SBSI 2025 são os relacionados a: Gestão de processos e sistemas de informação organizacionais, Sistemas de informação para gestão de dados, informação e conhecimento e Sistemas de informação e Inteligência Artificial (Generativa, LLM, PLN, entre outros). A pesquisa relatada possui a temáticas e/ou contribuição mais relacionadas a Visão Sociotécnica de Sistemas de Informação.

References

- [1] Martinho Isnard Ribeiro de Almeida, Maria Luiza Mendes Teixeira, and Dante Pinheiro Martinelli. 1993. Por que administrar estrategicamente recursos humanos? *Revista de Administração de Empresas* 33 (1993), 12–24.
- [2] Narallyne Araújo, Tiago Massoni, Camila Sarmento, Francielle Santos, and Ruan Oliveira. 2022. Investigating the relationship between software team leadership styles and turnover intention. In *Proceedings of the XXXVI Brazilian Symposium on Software Engineering*. 106–111.
- [3] Dan Avrahami, Dana Pessach, Gonen Singer, and Hila Chalutz Ben-Gal. 2022. A human resources analytics and machine-learning examination of turnover: implications for theory and practice. *International Journal of Manpower* 43, 6 (2022), 1405–1424.
- [4] Monalissa Barcellos, Gleison Santos, Tayana Conte, Bianca Trinkenreich, and Patricia Matsubara. 2022. Organizing Empirical Studies as Learning Iterations in Design Science Research Projects. In *Proceedings of the XXI Brazilian Symposium on Software Quality*. 1–10.
- [5] Tim Brown and Barry Katz. 2011. Change by design. *Journal of product innovation management* 28, 3 (2011), 381–383.
- [6] Michelle Larissa Luciano Carvalho, Paulo da Silva Cruz, Eduardo Santana de Almeida, Paulo Anselmo da Mota Silveira Neto, and Rafael Prikladnicki. 2024. Please do not go: understanding turnover of software engineers from different perspectives. arXiv:2407.00273 [cs.SE] <https://arxiv.org/abs/2407.00273>
- [7] I CHIAVENATO. 2014. *Gestão de pessoas: o novo papel dos recursos humanos nas organizações*. Manole. (book).
- [8] CNN. 2021. *Procura por profissionais de tecnologia cresce 671% durante a pandemia*. Retrieved July 20, 2023 from <https://www.cnnbrasil.com.br/tecnologia/procura-por-profissionais-de-tecnologia-cresce-671-durante-a-pandemia/>
- [9] Luiz Alexandre Costa, Edson Dias, Danilo Monteiro Ribeiro, Awdren Fontão, Gustavo Pinto, Rodrigo Pereira dos Santos, and Alexander Serebrenik. 2024. An Actionable Framework for Understanding and Improving Talent Retention as a Competitive Advantage in IT Organizations. arXiv:2402.01573 [cs.SE] <https://arxiv.org/abs/2402.01573>
- [10] A. B. P. F. da. CRUZ. 2019. *Identificação Organizacional, Intenção de Turnover e Satisfação com os Papéis de Vida: Estudo com uma Amostra da Geração dos Millennials no Setor Tecnológico*. Ph. D. Dissertation. Universidade de Lisboa, Lisboa, PT.
- [11] Aline Gonçalves De Miranda, Adriana Chaves Andrade, Edinéia Dos Santos, Reginaldo Moreno, and Vanessa Gonçalves Luchetta. 2017. A IMPORTÂNCIA DA GESTÃO DO TURNOVER. *Revista Maiêutica, Indaial*, 5, 1 (2017), 105–116. issn=2525-8346.
- [12] Amanda Enander and José Cardoso. 2020. How is employee turnover related to employee retention? A systematic review on two sets of meta-analyses. (2020).
- [13] Alessandra Flávia da Silva Ferreira. 2022. Turnover e o custo da rotatividade: um estudo de caso em uma startup de tecnologia no Estado do Rio Grande do Norte.
- [14] Guilherme Luiz Frufrek and Luciano Tadeu Esteves Pansanato. 2015. Employee turnover: An analysis of Brazilian software development professionals. In *8th IADIS International Conference on Information Systems, IS 2015*. 145–152.
- [15] Bruna Guimarães Gupy. 2022. Saiba como calcular o custo do turnover de forma simples. Disponível em: <https://www.gupy.io/blog/como-calcular-turnover>. Acesso em: 20 Julho 2023.
- [16] Mariana Dias Gupy. 2023. Turnover: o que é, como calcular e qual o impacto da rotatividade na empresa. Disponível em: <https://www.gupy.io/blog/turnover>. Acesso em 10 Junho 2023.
- [17] Sebastian Kaiser. 2019. The difference between Design Thinking & Design Science. Disponível em: https://medium.com/@spacefactor/sebastian_kaiser/the-difference-between-design-thinking-design-science-482244cfb327. Acesso em 12 Agosto 2023.
- [18] B. MACEDO. 2022. *Setor de tecnologia cresce mais de 60% durante a pandemia, aponta estudo*. Retrieved July 20, 2023 from <https://www.cnnbrasil.com.br/economia/setor-de-tecnologia-cresce-mais-de-60-durante-a-pandemia-aponta-estudo/>
- [19] Philipp Offermann, Olga Levina, Marten Schönherr, and Udo Bub. 2009. Outline of a design science research process. In *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*. 1–11.
- [20] Henry Ongori. 2007. A review of the literature on employee turnover. *Academic Journals*, <http://www.academicjournals.org/ajbm> (2007).
- [21] Glaucia Jardim Pissinelli, Leonardo Tomazelli Duarte, and Cristiano Torezzan. [n. d.]. UM MODELO DE REGRESSÃO LOGÍSTICA MULTIPLA PARA PREDICÃO DE TURNOVER—UMA APLICACÃO EM PEOPLE ANALYTICS.
- [22] PRICEWATERHOUSECOOPERS. 2017. *Workforce of the future. The competing forces shaping 2030*. Retrieved Mar 2, 2024 from <https://www.pwc.com/gx/en/services/workforce/publications/workforce-of-the-future.html>
- [23] Evy Rombaut and Marie-Anne Guerry. 2018. Predicting voluntary turnover through human resources database analysis. *Management Research Review* 41, 1 (jan 2018), 96–112. DOI 10.1108/MRR-04-2017-0098.
- [24] Aline Cavalcante Santana, Cleyton Mário de Oliveira Rodrigues, Ivaldir Honório de Farias Junior, and Wyllyams Barbosa Santos. 2023. OntoTurnover:: A Lightweight Domain Ontology for Modeling Employee Turnover. In *2023 18th Iberian Conference on Information Systems and Technologies (CISTI)*. IEEE, 1–4.
- [25] Anton F. Schlechter, Chantal Syce, and Mark Bussin. 2016. Predicting voluntary turnover in employees using demographic characteristics: A South African case study. *Acta Commercii - Independent Research Journal in the Management Sciences* (jan 2016), 1–10. ISSN: (Online) 1684-1999, (Print) 2413-1903.
- [26] Sharecare. 2020. Modelagem preditiva: o que é, para que serve e como implementar? Disponível em: <https://sharecare.com.br/blog/o-que-e-modelagem-preditiva/>. Acesso em 6 Julho 2023.
- [27] Lynn McFarlane Shore and Harry J Martin. 1989. Job satisfaction and organizational commitment in relation to work performance and turnover intentions. *Human relations* 42, 7 (1989), 625–638.
- [28] Cibele Cardoso da Silveira. 2011. Análise de turnover na química Brasil LTDA. <https://lume.ufrgs.br/handle/10183/40447>
- [29] Godwin J Udo, Tor Guimaraes, and Magid Igbaria. 1997. An investigation of the antecedents of turnover intention for manufacturing plant managers. *International Journal of Operations & Production Management* 17, 9 (1997), 912–930.
- [30] Yue Zhao, Maciej Hryniewicki, Francesca Cheng, Boyang Fu, and Xiaoyu Zhu. 2018. Employee Turnover Prediction with Machine Learning: A Reliable Approach. In *Conference: 2018 Intelligent System Conference (Intellisys)*, IEEEAT: London, UK. 737–758. <https://doi.org/10.1007/978-3-030-01057-7>

A Apêndice: Etapas da Construção do Modelo

A.1 Criação de um Dataset

Foi utilizado como instrumento para coleta de dados, a criação de um dataset (pré-processamento) com o histórico de movimentações de funcionários desde o ano de 2019 até o ano de 2022. Estes dados foram agrupados num arquivo csv e trabalhado via ferramenta Colab⁵ do google. A partir do histórico salarial foram geradas as Variáveis do Histórico Salarial, de Valor (em Reais) de Crescimento Salarial, de Valor em Percentual de Crescimento Salarial e da Diferença (em Anos) de Movimentações Salariais. Em seguida, foi feita uma preparação dos dados através de um dicionário de dados. Na sequência foi feita a conversão de dados referentes a Grau de Instrução. Para cada campo encontrado nestes termos, foram atribuídos valores de 1 a 10. Depois da preparação e da conversão do dicionário de dados, foi feita a carga de dados, criando a regra de que todos os valores numéricos teriam os valores decimais separados por vírgula (,) e valores de milhares separados por ponto (.). Na sequência foi feito um cálculo do histórico salarial de cada funcionário, baseando-se nas informações de Mérito, Promoção ou Adequação, obtidos ao longo da jornada do colaborador na empresa. A média de aumento, o percentual de aumento e o tempo que o funcionário levou para receber o reconhecimento foram usados como base e em seguida, cada persona teve seus resultados computados. Após a extração destes dados, foram selecionados os dados das personas que pediram desligamento da empresa. Para cada persona foi verificada a senioridade. Após a extração de todas as informações de cada persona, foi feita uma remoção de dados que não serviriam para análise a fim de diminuir a quantidade de dados efetivamente válidos. Também houve a preocupação com a governança dos dados dos colaboradores no momento da coleta dos dados, respeitando a LGPD e garantindo assim a sua anonimização. Foi necessário um cuidado com o armazenamento desta informação que seguiu padrões rígidos de segurança bem como a manutenção de sua integridade. Em seguida foi feita a disponibilização do dado para o pré-processamento. Durante a fase do pré-processamento, foi feito todo o trabalho de transformação dos dados brutos em dados mais refinados. Foi necessária muita cautela na hora de manusear estes dados a fim de não mudar sua integridade e consequentemente não ocasionar um desvio nos resultados. Finalmente, o preenchimento de dados faltantes se deu pelas categorias predominantes.

A.2 Análise Exploratória

Após construir o dataset com as etapas de pré-processamento anteriormente mencionadas, foi realizada uma análise exploratória em cima dos dados transformados para que fossem removidos aqueles que não trariam nenhum resultado válido ou que não contribuiriam de forma significativa para o estudo. A partir deste momento puderam ser apresentados alguns resultados obtidos com a análise dos dados significativos para a definição de um plano de ação para retenção de talentos.

A.3 Modelagem Preditiva

De acordo com [26], a modelagem preditiva é um método criado a partir de modelos matemáticos e aplicado por meio do aprendizado de máquina. Seu funcionamento geral tem um objetivo muito simples: calcular a probabilidade de um evento ocorrer no futuro e, em alguns casos, o custo associado a ele. Logo após o pré-processamento, uma atividade importante realizada foi uma modelagem para entender quais variáveis poderiam ser significativas no resultado do turnover.

Seleção de atributos Para definir quais atributos utilizar, foi aplicado um classificador do tipo Floresta Randômica para valorar a importância dos atributos, considerando apenas os 20 primeiros retornados pelo classificador.

Separação dos conjuntos de treino e testes A base de dados foi separada em treino e testes na proporção de 70% para treino e 30% para testes.

Definição dos Modelos Como diferentes modelos podem trazer diferentes resultados de classificação, foram escolhidos os seguintes classificadores para avaliação: Regressão Logística, Suport Vector Machine, Árvore de Decisão, Multilayer Perceptron, Floresta Randômica e Gaussian Naive Bayes

Execução dos Modelos Os diferentes modelos foram treinados considerando um processo de validação cruzada com 5 folds e foram avaliados de acordo com as seguintes métricas: Acurácia, Precisão, Recall e F1-Score.

Não foi feito nenhum refinamento de hiperparâmetro.

API Em seguida foi possível criar uma API e fazer seu deploy e com base nesta implantação, identificar os colaboradores com alta propensão ao turnover. Com base nestas informações é possível executar um plano de ação para retenção de talentos.

B Apêndice: Resultado Da Fase de Design Da Solução

Preparação E Exploração Dos Dados Os dados iniciais foram compostos de um montante de 1172 funcionários que trabalhavam na empresa entre 2002 e 2022. Lembrando que em cima destes dados foram feitos todos os tratamentos necessários para uma melhor acurácia do estudo preditivo do turnover, que serão explicados posteriormente nesta seção. Para o período entre 2019 e 2022, 297 profissionais pediram demissão, e a distribuição foi feita pelos cargos que ocupavam. Observou-se que o cargo que proporcionalmente teve um maior pedido de demissão foi o de Engenheiro de Software com 34% e na sequência, o de Engenheiro de Testes com 24%. O cargo que pediu mais demissão proporcionalmente ao tempo de empresa (até 2 anos) foi o de Engenheiro de Software com uma média de 19,2%. O cargo onde as pessoas ficaram menos tempo e pediram demissão também foi o de Engenheiro de Software (menos de 1 ano de empresa) com uma percentual de 5,1%. Uma outra análise realizada nos dados de quem pediu demissão foi a quantidade de anos que esta pessoa trabalhava na empresa. Deste montante, 48,10% pediram demissão com até 2 anos de empresa, 60,6% pediram demissão com até 3 anos de empresa e 73,10% pediram demissão com até 4 anos de empresa. Acima de 5 anos de empresa, caiu drasticamente o pedido de demissão voluntária. É importante constatar que 85% das pessoas que pediram demissão, participavam do modelo de trabalho em regime presencial e 15% trabalhavam no modelo Home Office.

⁵<https://www.alura.com.br/artigos/google-colab-o-que-e-e-como-usar>

Com este cenário percebemos que com os dados existentes e usando um processo de gestão de turnover, poderíamos tentar minimizar o turnover voluntário da empresa usando uma predição sobre qual(is) variável(is) poderia(m) ser trabalhada(s) a fim de tentar fazer com o que o colaborador pudesse permanecer na empresa ao invés de pedir desligamento. A seguir explicaremos como foi feita a fase de execução do modelo preditivo aplicado à fase de design da solução, que auxiliou a entender os benefícios da execução do processo de gestão do turnover (objetivo deste estudo).

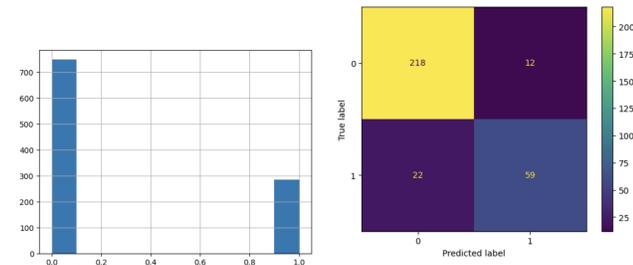
Execução do Modelo Quando observamos como estava distribuída a variável de pedido de demissão (variável alvo), observou-se o seguinte resultado: no treino do modelo de machine learning tivemos mais dados de uma variável false do que true. Sendo assim, o algoritmo iria aprender muito mais informações sobre a variável false e ficaria tendencioso e criaria um viés gerando um overfitting⁶ no modelo. Por isso precisou ser feito um balanceamento na nossa variável alvo, para que o algoritmo pudesse aprender tanto com a variável false quanto com a variável true. Depois do balanceamento, quando analisamos as variáveis mais correlacionadas com o alvo (pedido de demissão), foi identificada que aquela que tinha mais importância era o SEGURODEVIDA. Quando realizamos uma análise não supervisionada, (sem utilizar a variável alvo), a fim de segmentar os dados, não houve uma separação clara entre os funcionários que pediram demissão dos funcionários ativos. Quando realizamos uma análise supervisionada (utilizando a variável alvo) foi possível identificar as principais características que ajudaram a entender o turnover, de acordo com o grau de importância. Estas variáveis ajudaram a definir as fronteiras de separação dos grupos por parte do modelo. No treinamento do modelo de classificação foram selecionadas as variáveis mais importantes para construção de um modelo preditivo. Em seguida utilizamos uma abordagem baseada em árvore de decisão, criamos algumas personas para testar o modelo. No final, a ideia seria passar uma lista de dados de funcionários e uma API mostraria o resultado desta predição. No entanto, os resultados obtidos na primeira pesquisa com IA⁷ não trouxeram informações muito coerentes e foi necessária uma nova execução do ciclo do processo, uma vez que a variável cujo resultado foi o mais significativo "SEGURODEVIDA" não pode ser considerada válida, uma vez que este dado é zerado pelo departamento pessoal, para todo colaborador que pede demissão da empresa. Após a segunda execução do processo, foram feitos ajustes nos dicionários de dados para padronizar os campos de data de admissão, data de nascimento, data da última movimentação salarial e data de demissão que estavam contidas na base de dados com formatos diferentes. Após a correção dos campos de datas, foi feita a criação de atributos de histórico salarial a fim de serem usados como dados comparativos ao avaliar para cada colaborador, o menor e o maior valor de crescimento salarial, a taxa de aumento salarial, o mínimo e o máximo valor de salário, a média de crescimento, entre outros e estes atributos foram incluídos no processo de classificação para separar quem teve ou não promoção, para saber se houve aumento salarial ou não. Na sequência foi feita a geração de um dataframe cujo resultado foi um arquivo csv de histórico salarial que serviu de entrada para uma análise de estatística básica (Exploratory Data

Analysis - EDA) onde foi feita uma análise dos atributos mais correlacionados (em absoluto) com o alvo, indicando quem poderia ser utilizado como bom candidato para fazer a separação dos atributos mais relevantes. Foi possível fazer uma avaliação das correlações de atributos criados a partir do histórico salarial. Fazendo a análise de distribuição dos dados, foi perceptível um certo desbalanceamento entre os colaboradores ainda ativos na empresa daqueles que pediram demissão (dados alvos do estudo). Após a validação da importância dos atributos, foi feito um comparativo entre as variáveis e percebeu-se que existe uma variabilidade pequena entre elas. Algumas outras variáveis possuem variabilidade entre si, mas está restrito a um pequeno grupo de pessoas que possuem alguns benefícios a mais a depender da senioridade como por exemplo: Previdência Privada, Ifood Benefícios, Alelo Mobilidade e Desconto de Previdência Privada e por isso não foram levadas em consideração. Existe uma variabilidade maior quando analisamos a média do crescimento de valor salarial do que quando avaliamos pelos máximos e mínimos valores de salário. Existe uma variabilidade maior no tempo da última movimentação financeira, se compararmos os dados das distribuições dos ativos e dos inativos. É bom destacar que quando olhamos quem é líder e quem não é, existe uma diferença considerável entre os ativos e os inativos gerando uma hipótese de que a maioria de quem pediu demissão não era líder enquanto que é visualizado que existe uma boa quantidade de ativos que exercem papel de liderança. A variável Quinquênio é maior na distribuição dos ativos do que dos inativos corroborando com a hipótese de quem mais pede demissão tem menos tempo de casa uma vez que quem tem quinquênio tem minimamente mais de 5 anos de empresa. Após a distribuição, foi aplicado o método K-S Test (Kolmogorov Smirnov Test) para identificar a comparação entre as distribuições (ativas e inativas) e baseado nas amostras das duas distribuições foi verificado se teve alguma diferença significativa entre elas e o percentual de confiança da divergência entre elas. Considerando que existe uma possibilidade das distribuições serem diferentes mas com pelo menos 95% de confiança podemos dizer que elas não divergem tanto uma da outra. Foi feita uma avaliação da distribuição dos dados ativos e inativos para que não existisse uma variação tão significativa na distribuição destes atributos. Quando olhamos para estes atributos não existe uma diferenciação tão grande entre quem pediu e quem não pediu demissão. Alguns atributos, foram descartados do processo de aprendizagem. Na sequência foi reproduzida uma análise de clustering pela variável alvo na projeção em duas dimensões. Não ficou clara uma separação de grupos. Depois selecionamos os dois componentes que geraram a melhor formação de clusters mas que não geraram nenhuma informação relevante baseada na análise de duas dimensões. Após a análise dos atributos realizada pela Floresta Randômica, foram verificadas as 20 variáveis apresentadas por grau de importância. O grau de importância citado se refere a importância que cada uma destas variáveis apresenta na definição da fronteira de separação que o modelo aprende. Não necessariamente isso irá implicar que estes atributos em si, sejam a principal causa da pessoa ser demitida mas geram uma dica de onde pode partir a análise. Estes atributos foram levados para validação com o grupo focal de colaboradores do capital humano da empresa. Fizemos uma avaliação de importância dos atributos para estabelecer a fronteira de separação entre eles através da floresta randômica.

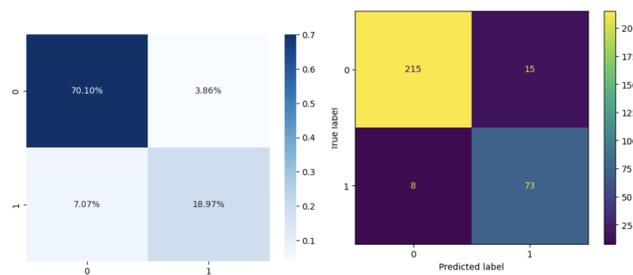
⁶<https://codecrush.com.br/glossario/overfitting>

⁷<https://www.alura.com.br/artigos/inteligencia-artificial-ia>

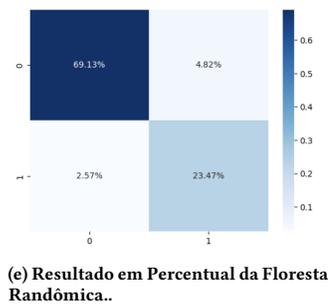
Figure 3: Resultados em Gráficos



(a) Desbalanceamento entre Ativos e Inativos..



(c) Resultado em Percentual do SVM. (d) Resultado em Números da Floresta Randômica.



Conforme pode ser visto na figura 4 (a), existe um desbalanceamento entre o conjunto de ativos e dos inativos. O conjunto de inativos tem menos da metade dos ativos. O primeiro treinamento no mesmo conjunto de dados e no mesmo conjunto de testes foi baseado em cinco métodos diferentes, conforme pode ser visto na Figura 4: Logistic regression, Support Vector Classifier, Multilayer Perceptron, Random Forest e Gaussian Naive Bayers. Analisamos a acurácia, a precisão, o recall e o F1 Score uma vez que estávamos interessados em prever quem pediria demissão e precisaríamos errar menos isso do que quem pediria demissão e de fato não pediu. De acordo com os resultados: O Random Forest teve uma precisão alta e um recall baixo. O perceptron teve o F1 Score baixo apesar de uma precisão razoável. Neste caso em cada uma das métricas, o SVM pareceu melhor em acurácia e precisão. Naive base teve o melhor recall. E o Decision Tree o melhor F1 Score.

Conforme pode ser visto na Figura 4 (b) obtivemos o seguinte resultado do método SVM em números. No treinamento de modelo

	Logistic Regression	Support Vector Classifier	Decision Tree	Multilayer Perceptron	Random Forest	Gaussian Naive Bayes	Best Score
Accuracy	0.865489	0.874194	0.864584	0.706993	0.860663	0.569556	Support Vector Classifier
Precision	0.840875	0.871785	0.767124	0.700808	0.866904	0.428244	Support Vector Classifier
Recall	0.686207	0.696673	0.787114	0.505021	0.661464	0.912523	Gaussian Naive Bayes
F1 Score	0.705698	0.726900	0.754435	0.406568	0.691635	0.560255	Decision Tree

Figure 4: Método de Avaliação com Base Desbalanceada.

x causalidade pudemos observar que o modelo preditivo trouxe como resultado 22 falsos negativos (número de colaboradores que não seriam candidatos a pedir demissão mas que de fato pediram) e 12 falsos positivos (significa que o modelo previu que 12 pessoas pediriam demissão mas que de fato não pediram). Esta análise foi feita com a base completa, com desbalanceamento. Faremos nova análise com base balanceada posteriormente. O gráfico da Figura 4 (c) traz a informação em termos percentuais, dos falsos negativos (7,07%) e dos falsos positivos (3,86%) encontrados pelo modelo preditivo. Este resultado foi aceitável mas ainda existia um número alto de falsos negativos. Então como tentativa de melhorar o modelo tomamos como estratégia trabalhar com base balanceada. Para balancear a base foi utilizado o processo de balanceamento SMOTE (Synthetic Minority Oversampling Technique). Em seguida foi executada uma nova rodada de treinamento do modelo avaliando novamente os mesmos seis métodos avaliados anteriormente e obtivemos os seguintes resultados que podem ser vistos na figura 9.

	Logistic Regression	Support Vector Classifier	Decision Tree	Multilayer Perceptron	Random Forest	Gaussian Naive Bayes	Best Score
Accuracy	0.852744	0.877011	0.846793	0.683327	0.893185	0.760071	Random Forest
Precision	0.868941	0.889731	0.823289	0.681140	0.912909	0.690915	Random Forest
Recall	0.834362	0.862670	0.862946	0.830787	0.870900	0.943383	Gaussian Naive Bayes
F1 Score	0.849382	0.874694	0.852056	0.724381	0.890316	0.797168	Random Forest

Figure 5: Métodos de Avaliação com Base Balanceada.

Após o balanceamento da base de dados o modelo que apresentou o melhor resultado foi o Random Forest. A acurácia e a precisão ficaram próximos. Conforme pode ser visto na Figura 4 (d) obtivemos o seguinte resultado do método Random Forest em números. No treinamento de modelo x causalidade pudemos observar que o modelo preditivo trouxe como resultado 8 falsos negativos (número de colaboradores que não seriam candidatos a pedir demissão mas que de fato pediram) e 15 falsos positivos (significa que o modelo previu que 15 pessoas pediriam demissão mas que de fato não pediram). Esta análise foi feita com a base balanceada e houve uma melhora no modelo uma vez que o nosso interesse era que ele errasse menos os falsos negativos e foi constatado que com a base balanceada ele passou a errar menos. Os falsos negativos passaram de 22 (base não balanceada) para 8 (base balanceada). O gráfico da Figura 4 (e) traz a informação em termos percentuais, dos falsos negativos (2,57%) e dos falsos positivos (4,82%) encontrados pelo modelo preditivo. Ao final, tivemos uma melhoria considerável no número de falsos negativos que são mais importantes de 22 para 8, ainda que tivemos um aumento dos falsos positivos de 12 para 15. O percentual de acertos saiu de 89,07% para 92,6%. Estes resultados foram validados através de um grupo focal com profissionais do capital humano da empresa, para entender o que parecia não estar coerente com a predição dos dados tanto os falsos negativos quanto os falsos positivos.