



Inteli

NOME DO PROJETO

Everymind



Controle do Documento

Histórico de revisões

Data	Autor	Versão	Resumo da atividade
09/08/2022	Kil Mateus	1.1	atualização da seção 4 (4.1)
12/08/2022	Kil Mateus	1.2	atualização das seções 1, 2, 3 e 4 (4.2)
22/08/2022	Gabriel Nascimento	2.1	atualização da seção 4 (4.1.6)
26/08/2022	Gabriel Nascimento	2.2	atualização da seção 4 (4.1.6, 4.1.7, 4.2,4.3)
08/09/2022	Ana Clara Zaidan	3.1	atualização da seção 4 (4.4)
09/09/2022	Ana Clara Zaidan	3.2	atualização da seção 4 (4.5)

Sumário

1. Introdução	5
2. Objetivos e Justificativa	6
2.1. Objetivos	6
2.2. Justificativa	6
3. Metodologia	7
3.1. CRISP-DM	7
3.2. Ferramentas	7
3.3. Principais técnicas empregadas	7
4. Desenvolvimento e Resultados	8
4.1. Compreensão do Problema	8
4.1.1. Contexto da indústria	8
4.1.2. Análise SWOT	8
4.1.3. Planejamento Geral da Solução	8
4.1.4. Value Proposition Canvas	8
4.1.5. Matriz de Riscos	8
4.1.6. Personas	9
4.1.7. Jornadas do Usuário	9
4.2. Compreensão dos Dados	10
4.3. Preparação dos Dados	11
4.4. Modelagem	12
4.5. Avaliação	13
4.6. Comparação de Modelos	14
5. Conclusões e Recomendações	14
6. Referências	15
Anexos	16

1. Introdução

Apresente de forma sucinta o parceiro de negócio, seu porte, local, área de atuação e posicionamento no mercado. Maiores detalhes deverão ser descritos na seção 4

O parceiro de negócio é a Everymind, uma consultoria de Salesforce. É uma empresa muito bem reconhecida pela própria Salesforce, e uma das maiores nesse mercado.

A Everymind foca em comercializar soluções utilizantes de tecnologias da análise de dados - tecnologias cuja origem é a Salesforce. Esse processo, como dita o conceito de boutique que muito inspira a Everymind, é alfaiatado conforme a demanda de cada cliente, que são, em maioria, grandes empresas. Por consequência de um modelo de negócio tão polido, os produtos oferecidos pela Everymind apresentam alto grau de eficiência.

A Everymind não possui práticas de verificação de turnover que apresentam índices de eficácia suficientemente elevados. Em tal contexto, a Everymind sofre substancial perda de lucratividade, posto que gastos com contratação de novos colaboradores é igualmente substancial, visto que envolve investimento em fatores desde treinamento até tempo de adaptação.

O problema proposto envolve a taxa de rotatividade de funcionários. Encontrar as possíveis causas que englobam esse índice e as relações entre cada variável baseada no nosso banco de dados faz parte do processo.

O alto índice de rotatividade de funcionários é o problema cuja resolução nos foi alocada. Encontrar-la-emos via construção de algoritmo de machine learning (ML) que, após identificar padrões nos dados relacionados ao contexto da saída de funcionários da empresa, possibilitará ação imediata sobre eles: "Propor um modelo preditivo que possibilite ter a visibilidade de tendência de risco de saída dos colaboradores e desta forma contribua para ações de retenção e redução de taxa de turnover, [tanto como] revisitar os demais processos de carreira e [de] desenvolvimento" (descrição oficial da demanda).

2. Objetivos e Justificativa

2.1. Objetivos

Classificaria-se o projeto como um sucesso no contexto ideal em que as altas taxas de saída de funcionários diminuiria de forma marcante, de maneira que, simultaneamente, os índices de satisfação desses mesmos funcionários cresceriam de forma excepcional.

2.2. Justificativa

A proposta de solução é a construção de um modelo preditivo (algoritmo de machine learning) que, após identificar padrões nos dados relacionados ao contexto da saída de funcionários da empresa, possibilitará ação imediata sobre eles.

O sucesso do modelo que propomos engendra, dentre outros benefícios, a redução da rotatividade (turnover) de funcionários, maior alinhamento dos funcionários à cultura da empresa, e maior orientação ao possível impacto de mudanças na governança corporativa.

O método usado em nossa solução possui destaque sobre o dos competidores por sua natureza inherentemente ágil, flexível, e rizomática. Com a capacidade de realizar análises sem intervenção humana; com potencial de processamento de grande quantidade de dados; e eficaz em identificar padrões em um período de tempo extremamente curto e com grande precisão, indubitavelmente, o algoritmo que desenvolvemos não pode ser subestimado.

3. Metodologia

Descreva as etapas metodológicas que foram utilizadas para o desenvolvimento, citando o referencial teórico. Você deve apenas enunciar os métodos, sem dizer ainda como ele foi aplicado e quais resultados obtidos.

3.1. CRISP-DM

O "Cross Industry Standard [for] Data Mining", ou CRISP-DM, é, em síntese, a norma universal para realização de mineração de dados. Possui como protocolo um processo que segue uma hierarquia de crescentes níveis de abstração. O primeiro nível dessa hierarquia,

denominado "phase", consiste na ordenação do processo de mineração em determinado número de fases, sendo cada uma dessas fases um conjunto de determinada quantidade de tarefas. O segundo nível é o conjunto de tarefas em questão, e é denominado "generic", posto que tais tarefas são classificadas como genéricas. Já o terceiro nível, "specialized task", é onde determina-se como serão realizadas as tarefas em cada contexto específico. O nome do quarto nível aptamente resume a sua função: "process instance" - o local em que são gravadas as decisões, ações, e resultados de cada engajamento que ocorre em cada um dos demais níveis durante o processo de mineração. Não obstante, em todos os níveis deve haver tanto completude quanto estabilidade. Isto é, respectivamente, tanto cobertura para com todo o processo de mineração, quanto conservação da validade do modelo em contexto de desenvolvimentos futuros que alicercem possíveis mudanças no protocolo. Além disso, é importante mencionar que, segundo especialistas, existem dois pré-requisitos para aplicar mineração de dados em um empreendimento: o primeiro é o entendimento do negócio; o segundo, o entendimento dos dados sobre os quais ocorrerá a mineração.

3.2. Ferramentas

Descreva brevemente as ferramentas utilizadas e seus respectivos papéis

- Adalove - direcionamento do desenvolvimento do projeto e criação do backlog
- Google Colaboratory - desenvolvimento do código
- GitHub - setor de entregas do código e da documentação
- Jira/Notion - organização do backlog de cada sprint, incluindo user stories, estimativa de dificuldade e responsáveis
- Google Docs - documentação

3.3. Principais técnicas empregadas

Descreva brevemente as principais técnicas empregadas, algoritmos e seus benefícios

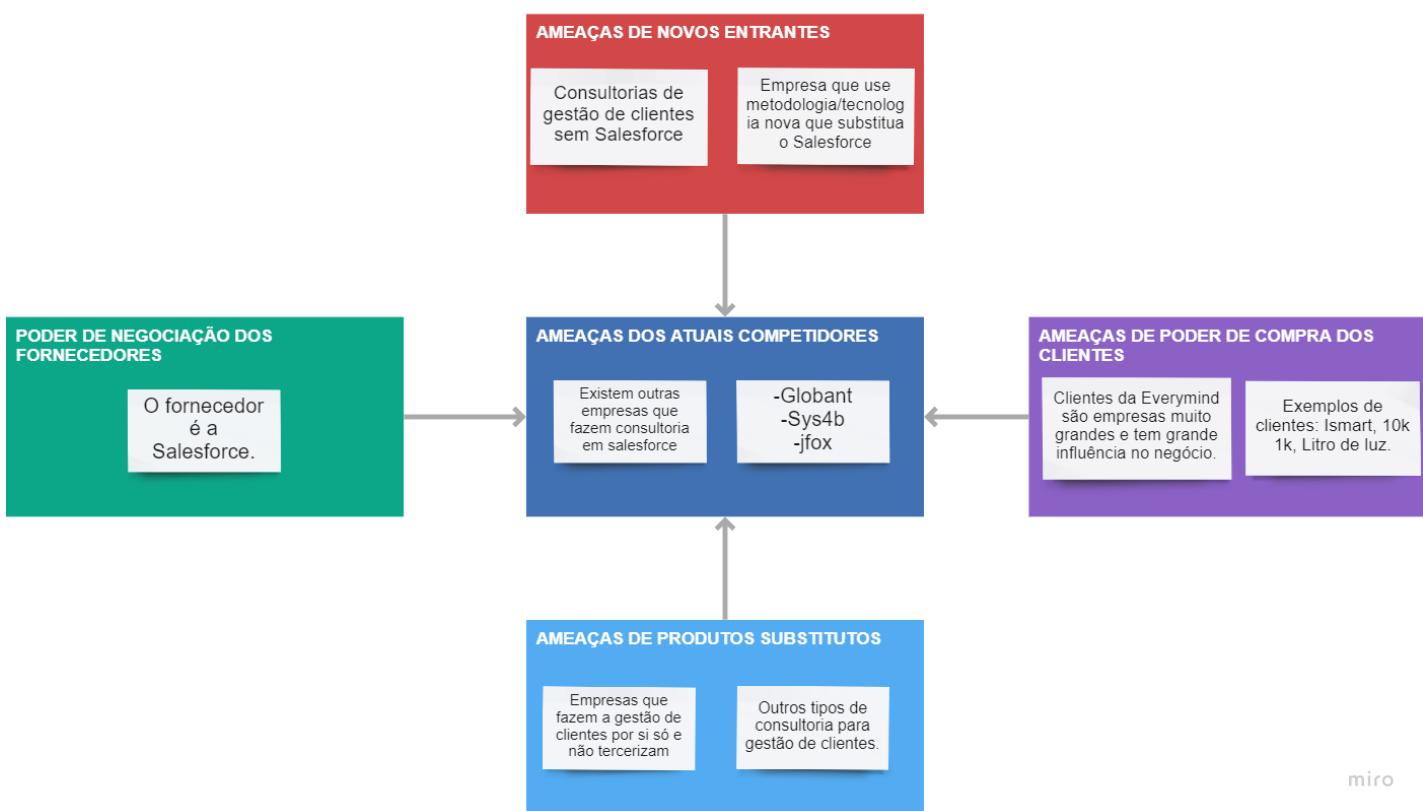
4. Desenvolvimento e Resultados

De maneira geral, você deve descrever nesta seção a aplicação dos métodos aprendidos e os resultados obtidos por seu grupo em seu projeto

4.1. Compreensão do Problema

4.1.1. Contexto da indústria

5 Forças de Porter



https://miro.com/app/board/uXjVOqCyebl/?share_link_id=925777377534

Para compreender o contexto de indústria no qual se encontra a Everymind, deve-se de início entender o modelo de negócio da Everymind, que é, essencialmente, o comércio de

soluções utilizantes de tecnologias da análise de dados - tecnologias cuja origem é a Salesforce. Esse processo, como dita o conceito de boutique que muito inspira a Everymind, é alfaiatado conforme a demanda de cada cliente, que são, em maioria, grandes empresas. Por consequência de um modelo de negócio tão polido, os produtos oferecidos pela Everymind apresentam alto grau de eficiência, especialmente se comparados com aqueles de seus principais competidores: Sys4b, Globant, e JFOX, que são outras consultorias de Salesforce, que prestam um serviço semelhante ao da Everymind.

Já para compreender o contexto de indústria de inteligências artificiais (no eixo do mercado relevante para este documento) é, em síntese, necessário entender onde são majoritariamente utilizadas. Mais comumente, o mercado de consultorias as utiliza para melhorar a experiência de usuários; para conseguir identificar potenciais compradores; para analisar o comportamento de clientes; para monitorar o marketplace; e como alavanca para o início do uso de Salesforce. Além disso, os ativos de TI recorrem à inteligência artificial para anteciparem problemas de desempenho, automatizando as devidas correções antes que tais problemas sejam detrimenosos à performance.

4.1.2. Análise SWOT



https://miro.com/app/board/uXjVOgCyebl/?share_link_id=925777377534

4.1.3. Planejamento Geral da Solução

a) Problema a ser resolvido

O problema a ser resolvido é o alto índice de rotatividade de funcionários. Tal problema engendra outros problemas, dentre eles: desconhecimento do motivo de saída de cada funcionário; gastos com contratação de novos colaboradores (que envolve investimento em fatores desde treinamento até tempo de adaptação); e dificuldade dos líderes de projeto em identificar quais funcionários têm mais chance de sair.

b) Dados disponíveis

2 Spreadsheets com dados básicos sobre funcionários que saíram e que foram demitidos; 1 Spreadsheet com dados de pesquisa de satisfação no ambiente de trabalho por setor da empresa

→ Dados básicos dos funcionários são: nome completo, data de admissão, data de saída (se estiver desligado), tipo de saída (dispensa, demissão, etc.), cargo, salário mensal, data de nascimento (i.e. idade), gênero, etnia, estado civil, grau de escolaridade, área (e.g. vendas), Estado (e.g. SP), cidade, situação (ativo, desligado, afastado), e se recebeu alguma promoção ou troca de cargo.

→ Dados da pesquisa de satisfação incluem perguntas que abordam fatores como: colaboração, compensação, comunicação, confiança, Diversidade e Responsabilidade Social, qualidade e frequência do reconhecimento, saúde pessoal, propósito e direcionamento, estresse, frequência, saúde mental, valores, desenvolvimento profissional, confiança, comunicação e colaboração com o gestor, autonomia, qualidade, promotor, equilíbrio entre vida profissional e pessoal, ambiente de trabalho, felicidade no trabalho, função dentro da empresa, orgulho e sugestões.

c) Solução proposta

Construção de algoritmo de machine learning que, após identificar padrões nos dados relacionados ao contexto da saída de funcionários da empresa, possibilitará ação imediata sobre eles: "Propor um modelo preditivo que possibilite ter a visibilidade de tendência de risco de saída dos colaboradores e desta forma contribua para ações de retenção e redução de taxa de turnover, [tanto como] revisitar os demais processos de carreira e [de] desenvolvimento" (descrição oficial da demanda).

d) Tipo de tarefa (regressão ou classificação)

O método de classificação mostra-se como o mais adequado para o desenvolvimento da AI requisitada, visto que ele os rótulos (y) pertencerão a um conjunto discreto e finito de categorias - "tem tendência de sair" ou "não tem tendência de sair".

e) Como a solução proposta deverá ser utilizada

Em um contexto ideal, a solução deverá ser utilizada da seguinte maneira:

1- AI é construída;

- 1- Preparação dos dados (organização, análise, tratamento);
- 3- Dados são minerados (inserção dos dados no modelo);
- 4- Conclusões são alcançadas - resposta sobre risco de o colaborador sair da empresa;
- 4- Com base nas conclusões, a Everymind engendra ações que visem solucionar a problemática do alto índice de saída de funcionários - como ações de reconhecimento sobre os colaboradores com maior chance de sair;
- 5- Ações mostram-se efetivas, problema é resolvido.

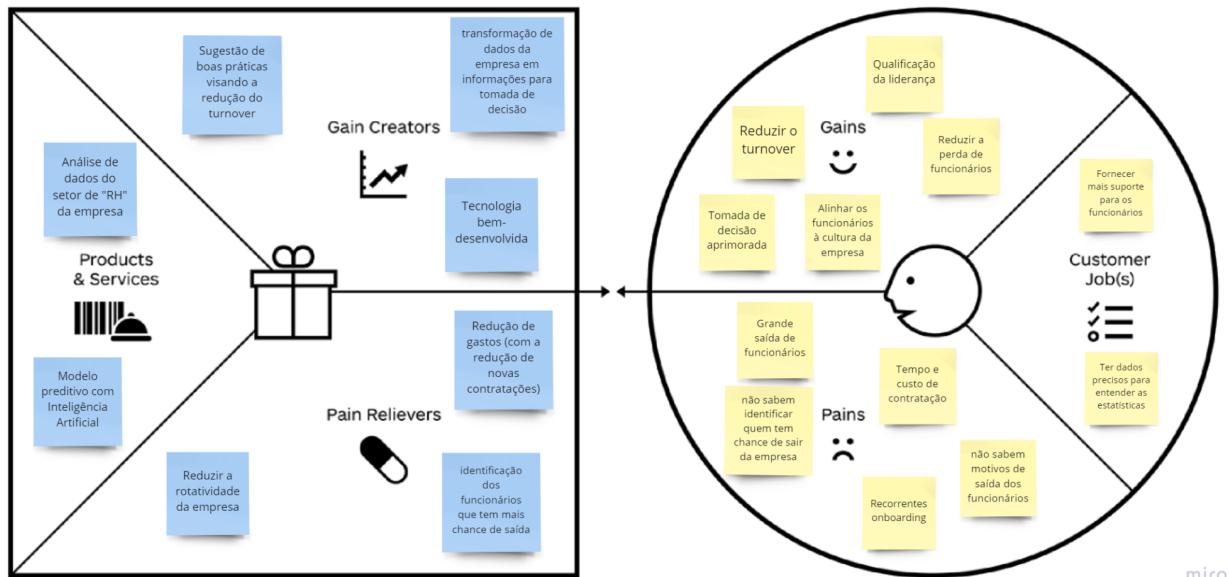
f) Benefícios trazidos pela solução proposta

O objetivo da solução proposta é resolver o problema do cliente (o alto índice de saída de funcionários). Portanto, os benefícios esperados incluem redução do turnover (rotatividade) de funcionários, contribuição para um aprimoramento da tomada de decisão pelas lideranças da empresa (ou seja, aumento da qualidade dessas lideranças), e funcionários mais bem alinhados à cultura da empresa.

g) Critério de sucesso + medida que será utilizada para o avaliar

No contexto em que o índice de saída de funcionários da Everymind, atualmente alto, torna-se baixo após a aplicação da solução que propomos, consideraremos que sucesso foi obtido.

4.1.4. Value Proposition Canvas



https://miro.com/app/board/uXjVOgCyebl/?share_link_id=925777377534

4.1.5. Matriz de Riscos

		Matriz de Risco									
Probabilidade		Ameaças					Oportunidade				
Muito	Alta	5	7	3	11, 12	1	15	13	4	1	
Alta		4	2	5	9	6/10	14				
Médio		3	7								
Baixa		2									4
Muito	Baixa	1			8						
		1	2	3	4	5	5	4	3	2	1
Muito Baixo		Baixo	Médio	Alta	Muito Alta	Muito Alta	Alta	Médio	Baixo	Muito Baixo	
Impacto											

https://miro.com/app/board/uXjVOgCyebl/?share_link_id=925777377534

Lista de riscos:

- 1- Variáveis pouco claras
- 2- Falta de dados necessários
- 3- Resposta pouco específica/subjetiva
- 4- Falta de experiência do time ao utilizar as ferramentas novas
- 5- Não alcançar expectativas do cliente
- 6- Falta de organização e gestão de tempo
- 7- Mau entendimento sobre o contexto da indústria de Sales force
- 8- Tecnologias pouco eficientes
- 9- Problemas com o Github ser open source
- 10- Falta de comunicação entre o grupo
- 11- Falta de proatividade dos integrantes
- 12- Má divisão de tarefas, sobrecarregando poucos
- 13- Complexidade alta demais do projeto
- 14- Perda/roubo do código e/ou banco de dados
- 15- Mudança de escopos constantes

A matriz de riscos visa prever e analisar os riscos que podem afetar um negócio/projeto, e ver o quanto cada um deles o afeta, e também a probabilidade de cada um acontecer. Os números na matriz representam cada item da lista.

4.1.6. Personas

Posicione aqui suas Personas (as que utilizam o modelo e as que são afetadas pelo modelo)

1^a persona

(funcionário do RH - utiliza o modelo)



+ :::

Nome: Luisa

- **Idade:** 27 anos
- **Ocupação:** Funcionária do setor de Pessoas e Cultura do EveryMind
- **Biografia:** gosta muito do seu trabalho, e acredita nos valores da empresa; trabalha com a parte de recrutamento, inclusão...
- **Características (personalidade, conhecimentos, interesses, habilidades):** Racional, com habilidades de resolução de problemas e link entre parte lógica e humana de processos
- **Motivações com modelos preditivos:** resposta de dados bem visual (com diferentes gráficos), e, mesmo não sendo profissional em tech, tem facilidade em entender a lógica
- **Dores com modelos preditivos:** às vezes as respostas são subjetivas e pouco claras
- **Motivações/necessidades com o problema:** Deseja analisar as respostas do modelo preditivo, podendo, assim, desenvolver um plano de ação juntamente aos líderes para diminuir o turnover
- **Dores com o problema:** como trabalha com recrutamento, ela tem que organizar eventos de onboarding constantemente, tendo um cenário de funcionários pouco alinhados aos valores da empresa; quer melhorar a imagem da empresa

2^a persona

(squad líder de projeto da empresa - utiliza o modelo)



Nome: Janice

- **Idade:** 30 anos
- **Ocupação:** squad líder de projeto do setor de desenvolvimento do Everymind
- **Biografia:** estuda constantemente sobre novas metodologias e busca sempre se aprimorar pessoal e profissionalmente
- **Características (personalidade, conhecimentos, interesses, habilidades):** mente inovadora e sempre aberta a novas soluções.
- **Motivações com modelos preditivos:** Deseja analisar as respostas do modelo preditivo, podendo, assim, desenvolver um plano de ação para evitar o turnover
- **Dores com modelos preditivos:** é uma tecnologia nova, e tem medo de não receber instruções suficientes para como lidar com as respostas do modelo
- **Motivações/necessidades com o problema:** quer melhorar a gestão de sua equipe, através da utilização das respostas do modelo, para, assim, melhorar o rendimento do time e aumentar a qualidade da experiência dos funcionários na empresa
- **Dores com o problema:** não consegue identificar as necessidades específicas de cada funcionário da equipe, para ter um tratamento e reconhecimento personalizado com cada um; rendimento da equipe está baixo

3^a persona

(funcionário da empresa - afetado pelo modelo)



Nome: Jonas

- **Idade:** 23 anos
- **Ocupação:** funcionário do setor de desenvolvimento do Everymind
- **Biografia:** foi contratado a menos de 6 meses e ainda está se adaptando à empresa
- **Características (personalidade, conhecimentos, interesses, habilidades):** grande habilidade em lógica e programação, se enxergando em um mercado de trabalho muito movimentado e com muitas oportunidades
- **Motivações com modelos preditivos:** gosta de usar a alexa em casa para despertadores e lembretes, mas não aproveita de features mais complexas
- **Dores com modelos preditivos:** não sabe se pode confiar nos resultados desses modelos; se são assertivos o suficiente
- **Motivações/necessidades com o problema:** reconhecerem sua insatisfação e a alta chance de se demitir, fará com que a empresa enxergue suas dores e invista em sua trajetória na empresa, mostrando o porque de ela estar ali (valores da empresa)
- **Dores com o problema:** ainda se sente meio perdido na empresa e não está muito satisfeito, pois não se sente visto pelos superiores e acha que não tem o reconhecimento que merece

4.1.7. Jornadas do Usuário



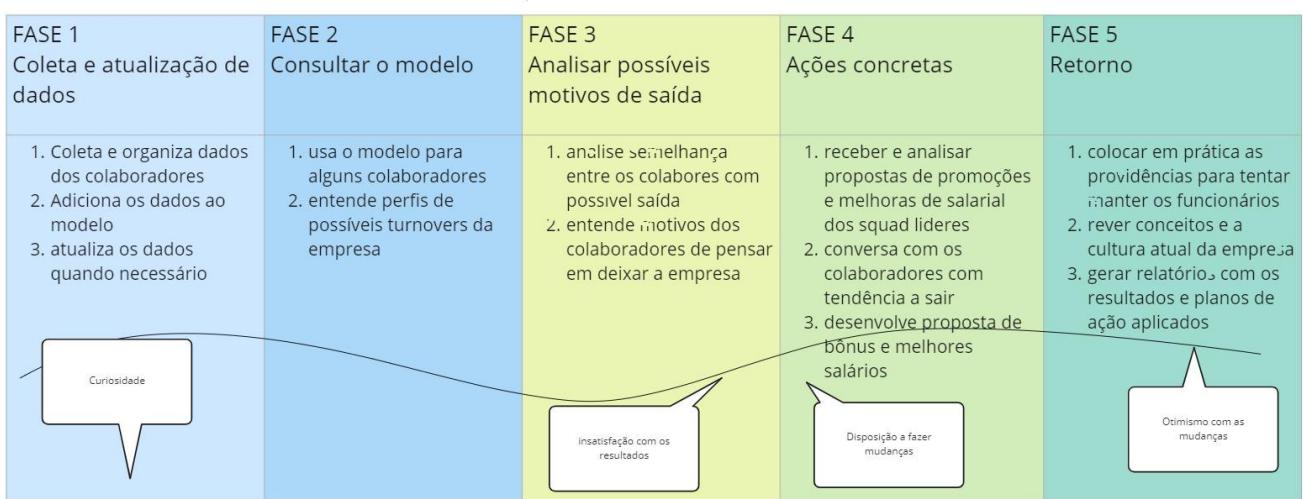
Luisa

Cenário: Funcionária do setor de Pessoas e Cultura do EveryMind, responsável pelo recrutamento de novos funcionários. Identifica um cenário de colaboradores com valores pouco alinhados

Expectativas

Desenvolver um plano de ação para reduzir o turnover e conseguir melhorar a imagem da empresa

FASE 1 Coleta e atualização de dados	FASE 2 Consultar o modelo	FASE 3 Analizar possíveis motivos de saída	FASE 4 Ações concretas	FASE 5 Retorno
1. Coleta e organiza dados dos colaboradores 2. Adiciona os dados ao modelo 3. atualiza os dados quando necessário	1. usa o modelo para alguns colaboradores 2. entende perfis de possíveis turnovers da empresa	1. analise semelhança entre os colabores com possível saída 2. entende motivos dos colaboradores de pensar em deixar a empresa	1. receber e analisar propostas de promoções e melhorias de salarial dos squad líderes 2. conversa com os colaboradores com tendência a sair 3. desenvolve proposta de bônus e melhores salários	1. colocar em prática as providências para tentar manter os funcionários 2. rever conceitos e a cultura atual da empresa 3. gerar relatórios com os resultados e planos de ação aplicados



Oportunidades

Se utilizar do modelo para entender a alta do turnover de funcionários, ao trabalhar contribuindo com o squad líder, existe as oportunidades de promover planos de ação concretos e personalizados, diminuindo as demissões.

Responsabilidades

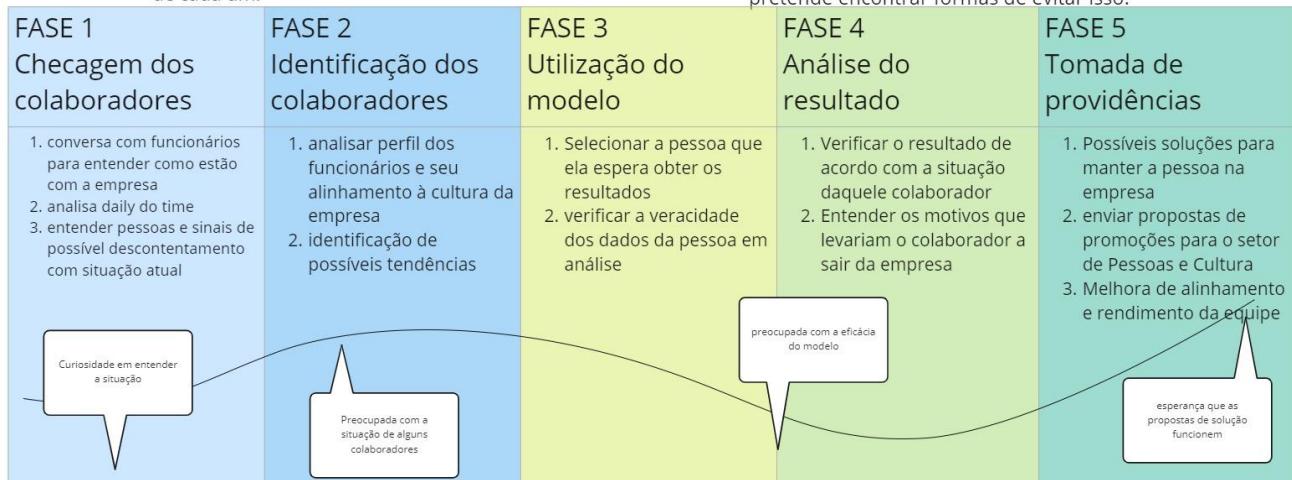
Como funcionário do setor de Pessoas e Cultura, tem a responsabilidade de coletar e alimentar os dados e garantir que sejam atualizados para que o modelo seja o mais preciso o possível.

miro



Janice

Cenário: Squad líder do setor de desenvolvimento, tem dificuldade em identificar quais funcionários têm maior potencial de deixar a empresa e em entender os motivos de cada um.



Oportunidades

Melhorar a gestão da sua equipe, melhorar o rendimento do time, e aumentar a qualidade da experiência dos funcionários da empresa, tornando-a mais personalizada.

Expectativas

Através do uso do modelo, pretende conseguir entender o que leva os funcionários a saírem, saber quais são os colaboradores com potencial de deixar a empresa, e pretende encontrar formas de evitar isso.

Responsabilidades

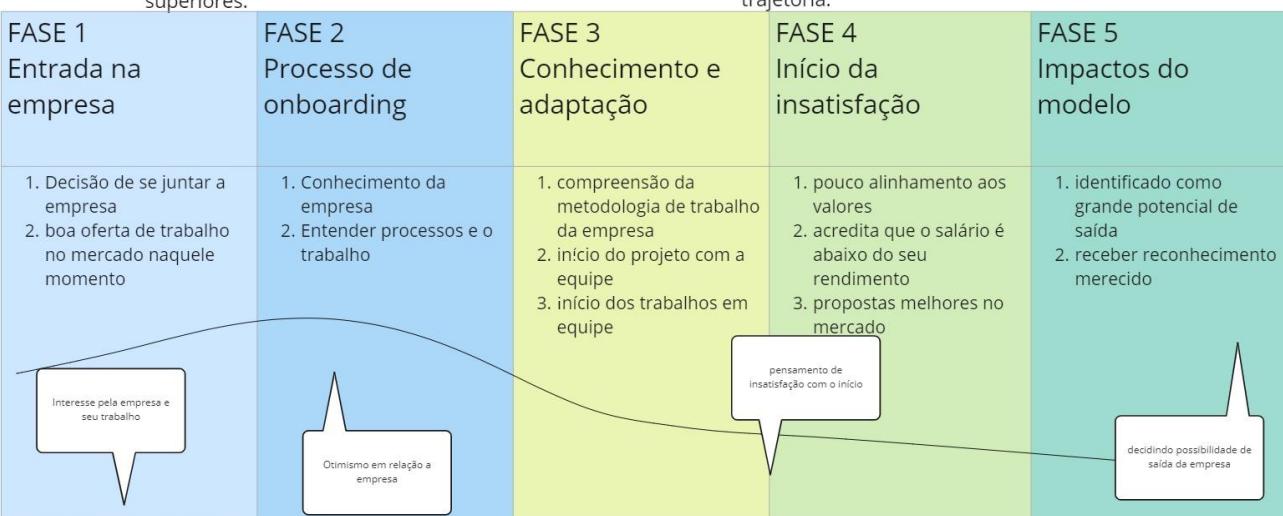
Analizar a resposta do modelo em relação a cada colaborador do seu squad, e partir para a ação, identificando as necessidades específicas de cada funcionário

miro



Jonas

Cenário: Funcionário do setor de desenvolvimento há 6 meses, com grande habilidade técnica e em um mercado de trabalho muito quente. não se sente visto pelos superiores.



Oportunidades

A partir da identificação de sua insatisfação e um plano de ação, as oportunidades são inúmeras, como: alinhamento de seus valores com a empresa, fazendo com que se sinta como parte significativa, condições de trabalho melhoradas e mais reconhecimento.

Expectativas

Espera que, com o resultado do modelo, receba o reconhecimento que merece e que suas dores sejam enxergadas pelo squad líder, e que invistam em sua trajetória.

Responsabilidades

Disponibilização dos dados necessários para o desenvolvimento do modelo preditivo e estar aberto à aplicação desse modelo e os possíveis impactos.

miro

4.2. Compreensão dos Dados

4.2.1 Descrição dos dados a serem utilizados

Planilha XLSX com as informações dos colaboradores que saíram e que foram contratados. A planilha tem 475 colaboradores com informações como: data de admissão, data e tipo de saída, cargo e área de atuação.

4.2.2 Dados disponíveis:

➡ 3 Spreadsheets

 ➡ Spreadsheet "Everymind"

 ➡ Dados são: data de admissão, data de saída, tipo de saída (dispensa, demissão, etc.), cargo, salário mensal, data de nascimento (i.e. idade), gênero, etnia, estado civil, grau de escolaridade, área (e.g. vendas), Estado (e.g. SP), cidade.

 ➡ Spreadsheet "Reconhecimento"

 ➡ Dados são: situação (ativo, afastado, ou desligado), data de admissão, data de vigência, novo cargo, novo salário, motivo ("promoção" ou "mérito"), "alterou função" ("sim" ou "não").

 ➡ Spreadsheet "Ambiente de Trabalho 27.07"

 ➡ Dados incluem perguntas que abordam fatores como: colaboração, compensação, comunicação, confiança, Diversidade e Responsabilidade Social, qualidade e frequência do reconhecimento, saúde pessoal, propósito e direcionamento, estresse, frequência, saúde mental, valores, desenvolvimento profissional, confiança, comunicação e colaboração com o gestor, autonomia, qualidade, promotor, equilíbrio entre vida profissional e pessoal, ambiente de trabalho, felicidade no trabalho, função dentro da empresa, orgulho e sugestões.

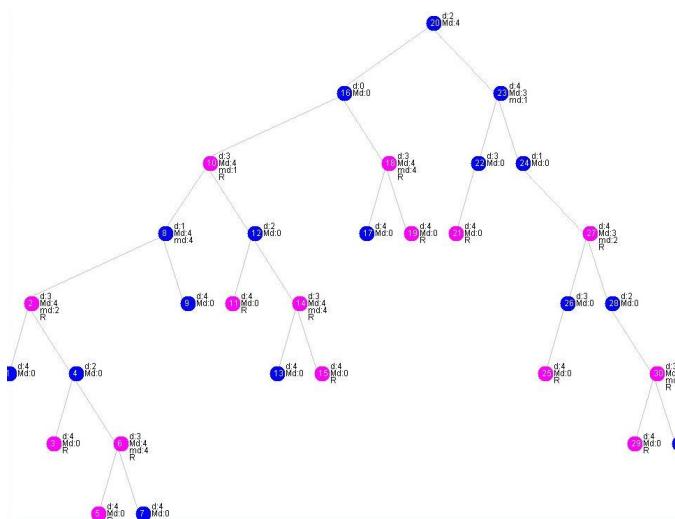
→ Dados referentes às perguntas: divisão, pilar, pontuação, fator, pontuação, pergunta, pulou, muito insatisfeito, insatisfeito, neutro, satisfeito, muito satisfeito, taxa de confiabilidade.

→ Spreadsheet "Chart1"

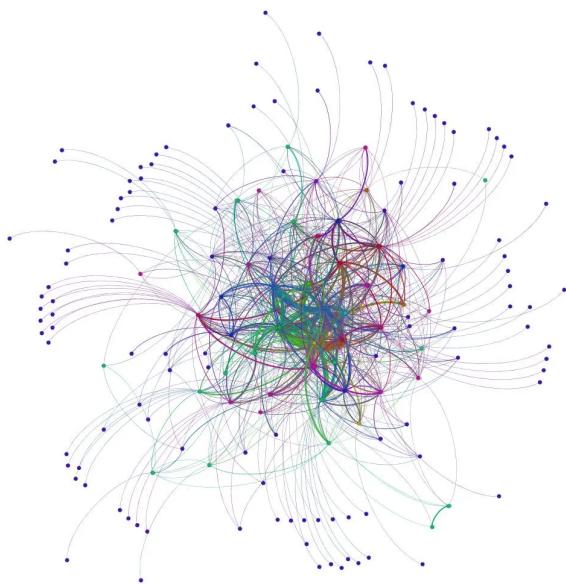
→ Dado é: gráfico demonstrando trendline para respectivas quantidades de "muito satisfeito" (vide seção Dados do Spreadsheet "Ambiente de Trabalho 27.07").

a) Se houver mais de um conjunto de dados, descrição de como serão agregados/mesclados.

A interpretação dos spreadsheets será feita a partir das instruções dadas pelo Cliente, e, no momento de escrita (12 de agosto, 2022), mostra-se ideal que sua agregação seja feita de maneira rizomática (em oposição à maneira arborescente).



A imagem acima retrata uma estrutura arborescente: caracterizada por sua orientação por princípios totalizantes, binarismo, e dualismo. Progresso unidirecional, sem a possibilidade de retroatividade e de cortes binários contínuos.



Rizomas, representados pela imagem acima, ao contrário de árvores, são pontos de entrada e saída não-hierárquicos na representação e na interpretação de dados. Isto é, uma concepção horizontal e não-hierárquica em que qualquer coisa pode estar ligada a qualquer outra, sem priorização por espécies. De acordo com Deleuze & Guattari, os princípios de um rizoma são:

- 1 e 2. Princípios de conexão e heterogeneidade: qualquer ponto de um rizoma pode ser conectado a qualquer outro, e assim deve ser.
- 3. Princípio da multiplicidade: só quando o múltiplo é efetivamente tratado como substantivo, "multiplicidade", deixa de ter qualquer relação com o Um;
- 4. Princípio da ruptura significante: um rizoma pode ser rompido, mas recomeçará em uma de suas velhas linhas, ou em novas linhas;
- 5 e 6. Princípios de cartografia e decalcomania: um rizoma não é passível de nenhum modelo estrutural ou generativo; é um mapa, e não um traçado. O que distingue o mapa do traçado é que ele é inteiramente orientado para uma experimentação em contato com o real.

Parafraseando Nick Land, "Schizoanalysis works differently. It avoids Ideas, and sticks to diagrams: networking software for accessing bodies without organs. BWOs, machinic singularities, or tractor fields emerge through the combination of parts with (rather than into) their whole; arranging composite individuations in a virtual/ actual circuit. They are additive rather than substitutive, and immanent rather than transcendent: executed by functional complexes of currents, switches, and loops, caught in scaling reverberations, and fleeing

through intercommunications, from the level of the integrated planetary system to that of atomic assemblages. Multiplicities captured by singularities interconnect as desiring-machines; dissipating entropy by dissociating flows, and recycling their machinism as self-assembling chronogenic circuitry.”.

Tendo em mente os fatos mencionados, pode-se concluir que interpretar os dados de acordo a partir das anotações realizadas pelo Cliente na lousa do Inteli, e que agregá-los de acordo com uma lógica rizomática, no momento, mostra-se como a decisão mais adequada.

b) Descrição dos riscos e contingências relacionados a esses dados (qualidade, cobertura/diversidade e acesso).

Os dados não são de qualidade exímia posto que para construir uma AI que providencie resultados muito exatos, muito mais dados são pré-requisitados. Isto é, fatores que são indubitavelmente de extrema relevância não nos foram fornecidos, a exemplo daquele que engloba o número de filhos de cada ex-funcionário, ou o que engloba preferência partidária, e o que mostra quais políticas foram adotadas ou abandonadas pela empresa e em qual data. Analogamente, pode-se afirmar que a diversidade dos dados também não é ótima. Quanto ao acesso, pode ser classificado como de boa qualidade, pois o spreadsheet e as imagens são, pela própria natureza de seus respectivos formatos, facilmente acessíveis.

c) Se aplicável, descrição de como será selecionado o subconjunto para análises iniciais (quando o tamanho do conjunto de dados impossibilita a utilização do conjunto completo em todas as etapas da definição do modelo a ser usado).

[no momento de escrita (dia 10 de agosto, 2022), não é aplicável]

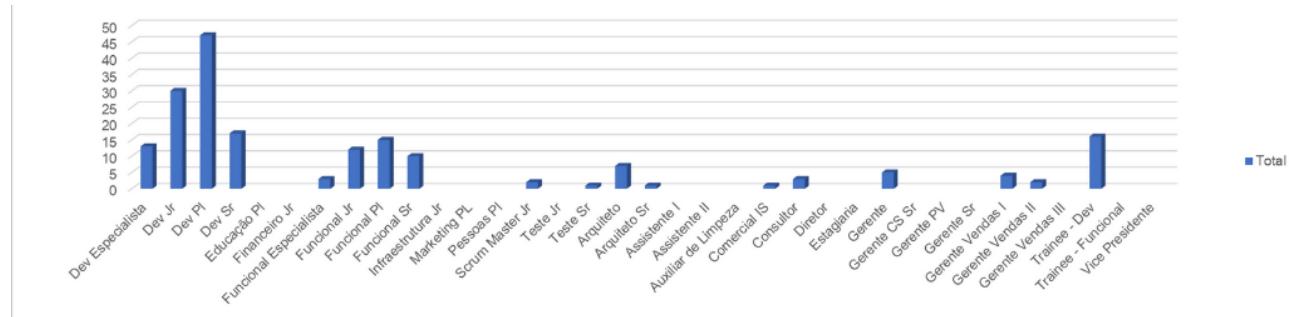
d) Se houver, descrição das restrições de segurança.

No momento de escrita (dia 10 de agosto, 2022), a segurança restringe-se àquela regularmente aplicada aos documentos confidenciais da Everymind e do Inteli. Tal segurança engloba protocolos como a proibição de publicá-los no GitHub.

4.2.3 Descrição estatística básica dos dados, principalmente dos atributos de interesse, com inclusão de visualizações gráficas e como essas análises embasam suas hipóteses.

1. Total de saídas por cargo

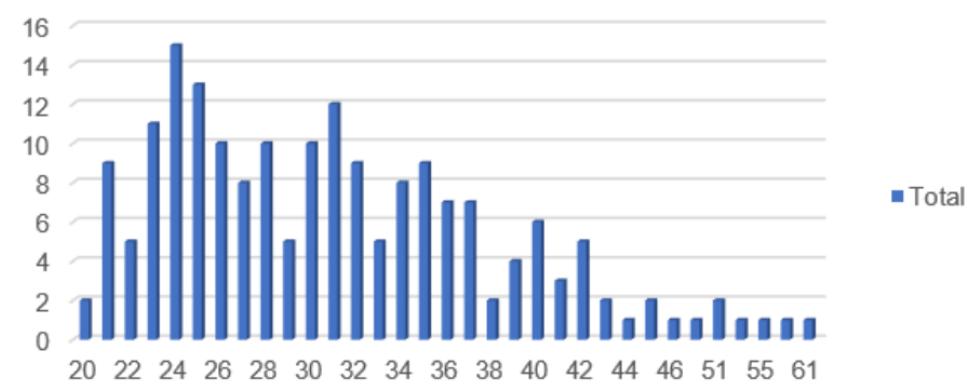
Este gráfico exibe a quantidade de pessoas que deixaram a empresa por cargo.



2. Saídas por idade

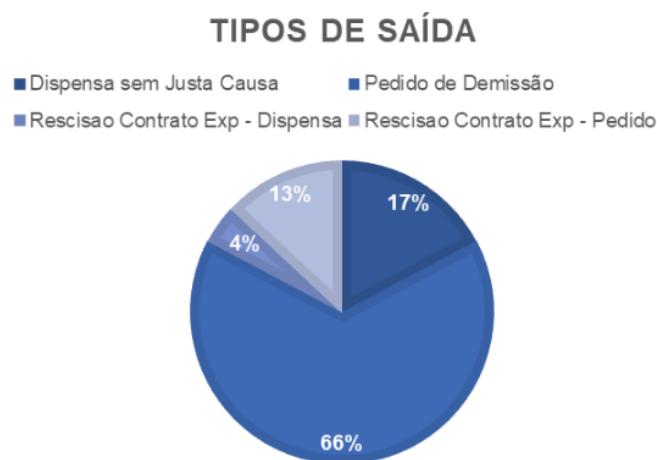
Este gráfico exibe a quantidade de pessoas que saíram por idade.

Saída por idade



3. Tipos de saída

Este gráfico exibe a porcentagem de saídas em cada um dos tipos de saída da empresa.



4.2.4 Descrição da predição desejada (“target”), identificando sua natureza (binária, contínua, etc.)

Para essa etapa do projeto será necessário entender principalmente a parte de saídas da tabela, se o funcionário saiu ou não da empresa, e qual foi a forma de saída

parte da tabela que o modelo vai “responder” (no caso, se saiu ou não saiu)

por enquanto o modelo preditivo é binário (de classificação)

4.3. Preparação dos Dados

Descreva as etapas realizadas para definir os dados e os atributos descritivos dos dados (“features”) a serem utilizados. Essa descrição deve ser feita de modo a garantir uma futura reprodução do processo por outras pessoas, e deve conter:

a) Descrição de quaisquer manipulações necessárias nos registros e suas respectivas features.

As manipulações de dados que fizemos retornam dados filtrados a fim de gerenciar melhor o modelo.

Salário Ativos:

Pegamos os dados da coluna 'Tipo Saída', e filtramos usando o comando '.isnull()' que retorna todos os colaboradores que possuem dados nulos, ou seja, quando pegamos também o salário, ele volta com os salários de todos os funcionários ativos.

```
[6] #Função retorna apenas os colaboradores que possuem o valor null na coluna de 'Tipo Saída'.
#.isnull() Ela pega todos os valores Null
Salario_ativos = planilha[planilha['Tipo Saída'].isnull()]

[7] Salario_ativos['Salario Mês']

 191    3304.80
 192    2000.00
 193    3000.00
 194    1750.00
 195    1017.00
 ...
 470    1716.48
 471    1716.48
 472    1716.48
 473    8250.00
 474    3491.66
Name: Salario Mês, Length: 284, dtype: float64
```

Salario_desativos:

Pegamos os dados da coluna 'Tipo Saída', e filtramos usando o comando 'dropna()' que retorna apenas os valores que não possuem campos vazios, ou seja, quando pegamos também o salário, ele volta com os salários dos antigos funcionários.

```
[ ] #Função que filtra todos que não tem valor NaN, ou seja, ele retornar apenas os funcionários desligados.
Salario_desativos = planilha.dropna()

▶ Salario_desativos['Salario Mês']

  0    4406.965
  1    4660.530
  2    7000.000
  3    7772.155
  4    2750.000
 ...
 186   1750.000
 187   7250.000
 188   8500.000
 189   6000.000
 190   1500.000
Name: Salario Mês, Length: 191, dtype: float64
```

Além disso, fizemos outras manipulações de dados, facilitando a sua visualização, como a separação dos estados e regiões, e transformando em valores numéricos. Calculamos também a idade dos colaboradores subtraindo a sua data de nascimento da data atual, e transformamos a coluna de gênero em 0 e 1, e proporção de gênero por saída de forma relativa.

Esse estrutura código pode ser utilizado em outras buscas simples.

b) Se aplicável, como deve ser feita a agregação de registros e/ou derivação de novos atributos.

As manipulações de dados que precisamos fazer para validar o nosso modelo preditivo, foram as seguintes.

Jornada de Trabalho:

Pegamos os dados das colunas 'Dt Saída' e 'Dt Admissão' onde geramos uma nova coluna. 'Tempo de Trabalho', onde retorna o tempo entre essas datas, em dias.

```
[5] #Função pega a data de admissão do colaborador e a data do seu desligamento, e encontra o período entre elas.
#dropna() Ela remove todos os dados que possuem valores NaN...
Jornada = (pd.to_datetime(planilha['Dt Saída']) - pd.to_datetime(planilha['Dt Admissão'])).dropna()

[5] Jornada
0    4594 days
1    2573 days
2    2250 days
3    2416 days
4    1635 days
.. .
186   60 days
187   35 days
188   30 days
189    7 days
190    7 days
Length: 191, dtype: timedelta64[ns]
```

Idade:

Pegamos os dados da coluna 'Dt Nascimento' de cada colaborador, e fizemos a diferença com a data atual para descobrimos a idade do funcionário.

```
[1] #Ele pega a data de hoje e subtrai da data de nascimento, retornando a idade, np.timedelta64 retorna a data em ano.
df1['Idade'] = ((pd.to_datetime('today')-pd.to_datetime(df1['Dt Nascimento']))/ np.timedelta64(1, 'Y')).astype(int)
```

c) Se aplicável, como devem ser removidos ou substituídos valores ausentes/em branco.

Em questão dos dados aplicados com valores ausentes e em branco, dependendo da coluna selecionada, conseguimos usá-los com o modelo de substituição.

Como exemplo temos os dados de tipo de saída: os valores em branco representavam que a pessoa estava ativa, sendo assim, criamos uma coluna numérica com os valores “0” para “ativos” e “1”, “2”, “3”, “4” para desativados.

Isso nos possibilitou criar uma relação entre data de admissão e saída e assim receber a quantidade de dias trabalhados descartando as pessoas que não foram demitidas.

Por fim, analisamos os dados sobre etnia e não utilizamos essa coluna pois não temos informações suficientes, os dados já vieram enviesados.

d) Identificação das features selecionadas, com descrição dos motivos de seleção.

As features selecionadas foram: Nome (ou Codinome, os nomes estão diferentes nas tabelas), Dt Saída (se houver), Situação (Ativo ou Desativado), Salário, Novo Salário

(relacionado ao reconhecimento), Cargo, Novo Cargo (relacionado ao reconhecimento), Dt Vigência (relacionado ao reconhecimento), Gênero, Motivo Saída.

As razões para cada uma foram respectivamente, identificar colaboradores, saber se saíram (Dt Saída e Situação), quanto o colaborador ganha por mês, ou o quanto ele passa a ganhar em um determinado período, identificar o cargo que atua, ou vai passar a atuar, data que recebeu o reconhecimento, identificar qual o gênero predominante na empresa, e o motivo da saída de um colaborador.

Não deixe de usar tabelas e gráficos de visualização de dados para melhor ilustrar suas descrições.

IMPORTANTE: Crie tópicos utilizando a formatação “Heading 3” (ou menor) para que o Google Docs identifique e atualize o Sumário (é necessário apertar o botão Refresh no Sumário para ele coletar as atualizações)

4.4. Modelagem

4.4.1 Modelos testados

Realizamos experimentos que englobam 6 tipos de modelos que consideramos válidos e com potencial para o nosso projeto de predição de Classificação: Regressão Logística, KNN, Árvore de decisão, SVM, Naive Bayes e Redes Neurais.

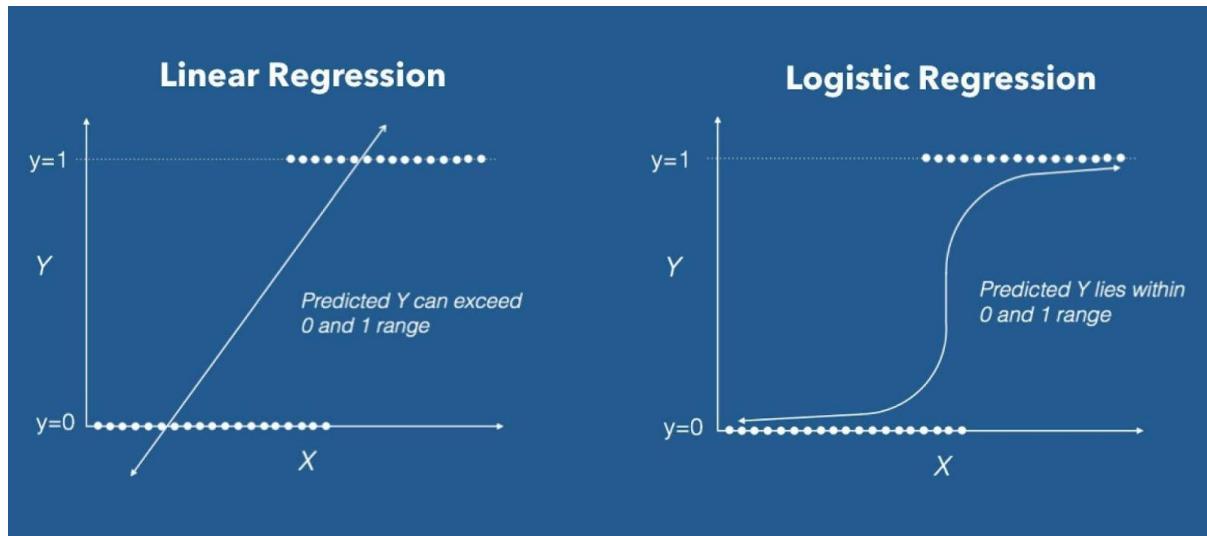
4.4.1.1 Modelo de Regressão Logística

O Modelo de Regressão Logística é o modelo estatístico mais utilizado para modelar variáveis categóricas. Ele é usado no aprendizado de máquina (ML) para ajudar a criar previsões precisas. O modelo tenta criar uma função matemática que visa predizer valores em relação às variáveis categóricas. É muito parecido com uma função de regressão linear, porém ela se diferencia pela utilização do valor Y, onde ao invés dele assumir um valor específico, ele retorna um valor binário (de 0 a 1).

Sendo assim, esse modelo utiliza operações estatísticas.

Em outras palavras, ele é indicado para situações em que a resposta é **binária**.

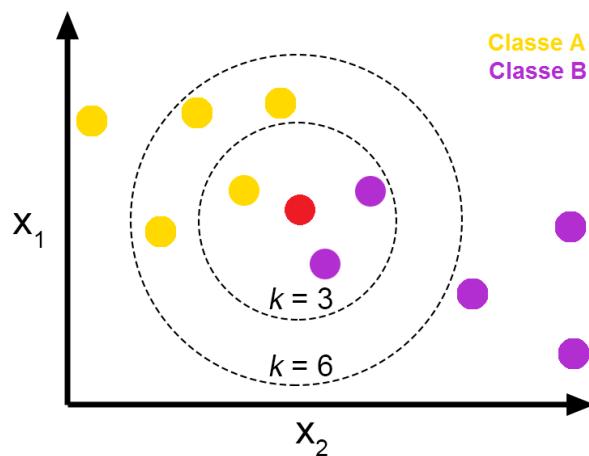
As vantagens da utilização desse modelo são constituídas por: requerer pequeno número de suposições, fornecer resultados em termos de probabilidade, classificação de indivíduos em categorias, facilidade com variáveis independentes categóricas, não precisa de escala de recursos de entrada e não necessita de grandes quantidades de recursos computacionais.



4.4.1.2 Modelo de KNN (K-Nearest Neighbors)

O modelo de KNN (K-Nearest Neighbors, em tradução literal “K-vizinhos mais próximos”) é um aprendizado baseado em instâncias, que compara um exemplar com classe não conhecida com os com classe já conhecida para fazer a classificação. O algoritmo consiste em estocar o conjunto de treinamento e realizar comparações entre cada exemplar de teste e os exemplares estocados, mas uma desvantagem é que, se comparado a outros modelos, esse processo pode levar muito tempo se o conjunto de treinamento for muito grande.

Essa comparação entre o novo exemplar com os de treinamento consiste em calcular a distância entre esse exemplar com os já conhecidos, utilizando métricas como distância de Manhattan e Euclidiana, encontrando os “k” exemplares mais próximos do novo. Logo, a classe da maioria dos “k” exemplares mais próximos ao exemplar de teste é aquela que deve ser atribuída a ele.



Na prática, o processo de experimentação de diferentes features foi dividido em 4 etapas: a divisão do dataset, o dimensionamento da feature, a definição do “k” e a avaliação do modelo.

1. A divisão do dataset para treino e teste foi feita através do “train_test_split”, importado do “sklearn.model_selection”, em que, para todos os testes, definimos como padrão 0.3 do dataset para teste.
2. Para otimizar a divisão dos dados e padronizar as features, utilizamos o “StandardScaler()”;
3. Para definir o parâmetro “k”, foi utilizado o padrão em que k é igual à raiz quadrada do tamanho do conjunto de testes. Nesse contexto, os outros parâmetros serão mais explorados na próxima sprint;
4. Para avaliar os resultados do modelo KNN, foi utilizado a taxa de acurácia e matriz de confusão, que serão melhor explicados na seção 4.5

4.4.1.3 Modelo de Árvore de Decisão

A árvore de decisão funciona iniciando com um único nó, e se divide em possíveis resultados. Cada um desses resultados leva a nós adicionais, que se ramificam em outras possibilidades. Uma árvore de decisão se ramifica em diversas escolhas que podem ser tomadas, que levam a outras escolhas, e assim por diante. A árvore continua fazendo isso até achar uma solução para o problema.

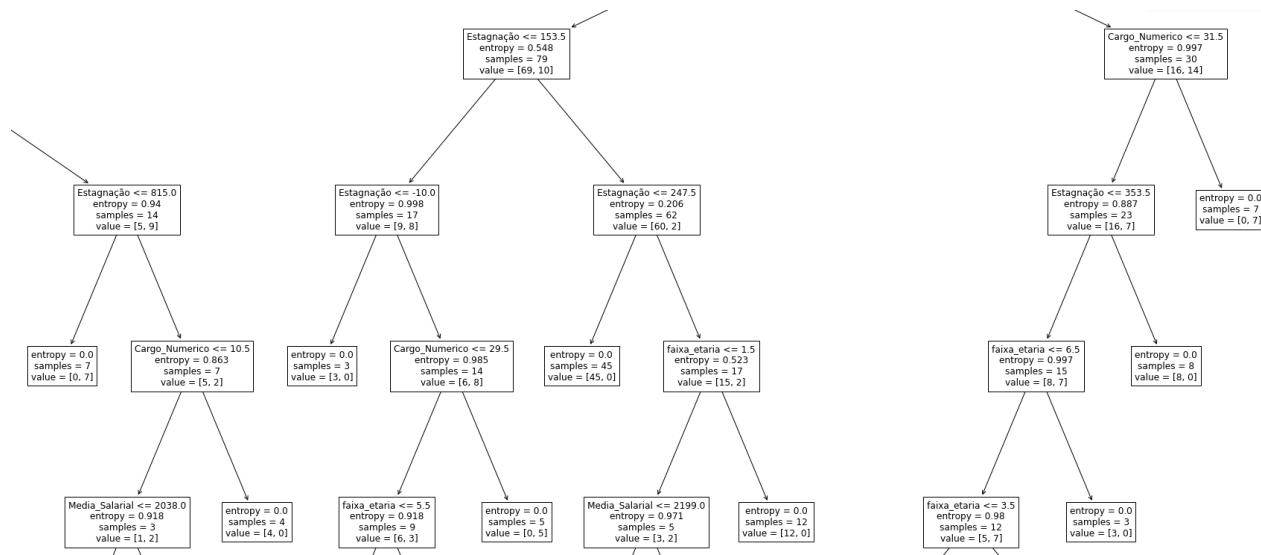


Imagen de parcela da árvore de decisão usada pelo grupo.

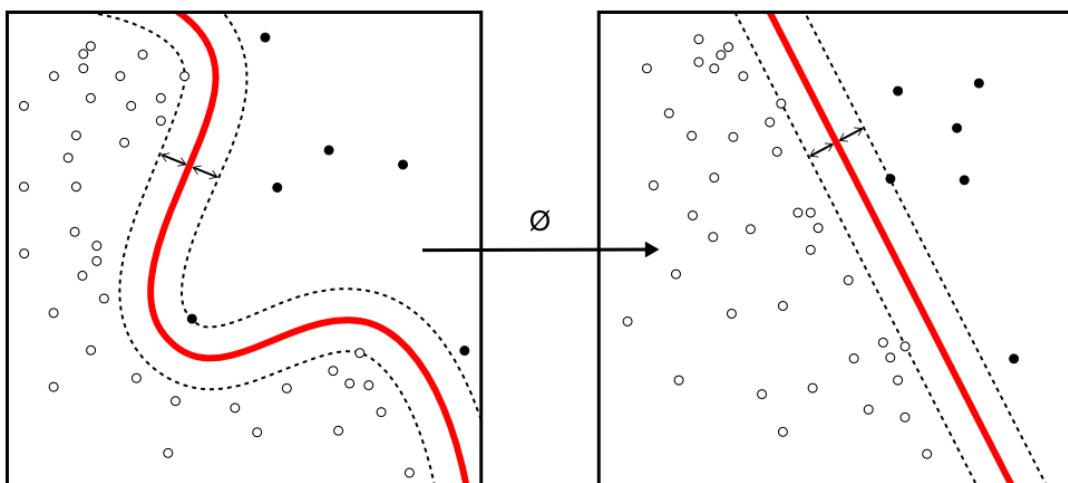
A aplicação desse modelo no projeto é verificar as possibilidades de características que um colaborador pode ter, e assim, verificar se com essas características, o colaborador ficaria na empresa, ou saíria.

A maior acurácia dentre os testes feitos desse modelo foi de 77%, utilizando as variáveis: 'faixa_etaria', 'Cargo_Numerico', 'Regiao_Numerico', 'Salario_Comparado' e 'Estagnação'. Além disso, foi observado que o modelo acerta mais os funcionários que sairiam do que os que ainda estão na empresa.

4.4.1.4 Modelo de SVM

O modelo SVM (Support Vector Machine) é um algoritmo de aprendizado supervisionado que contribui muito para tarefas de classificação e categorização. O algoritmo busca uma linha de separação entre duas classes distintas analisando um ponto de cada grupo que mais estão próximos um do outro. Ou seja, o SVM escolhe uma reta entre dois grupos que está equidistante de ambos. Um detalhe importante: também existe o SVM não linear, que separa os grupos sem necessariamente uma reta. Isso acontece por meio de uma transformação não-linear do espaço.

Para o nosso contexto, hipoteticamente, o modelo SVM faz sentido por classificarmos binariamente a tendência de turnover dos colaboradores, nesse sentido, o modelo foi testado para verificarmos essa hipótese.



Fonte: [Wikimedia Commons](#).

Contudo, após realizarmos os testes, verificamos que a acurácia média do modelo SVM em relação à nossa tabela de testes foi de 64% que não é uma boa acurácia em relação aos outros modelos testados. Aqui os resultados:

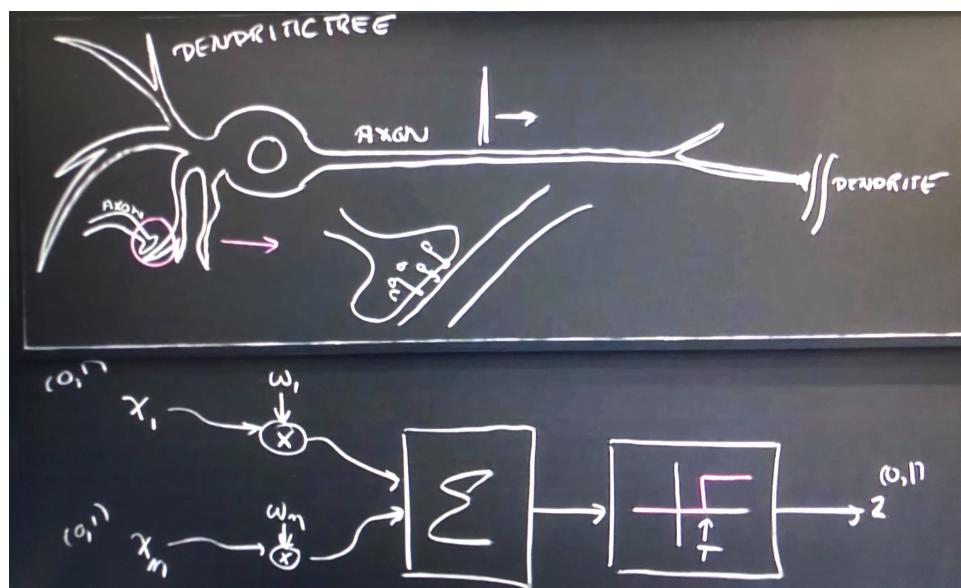
4.4.1.5 Modelo de Naive Bayes

Naive Bayes é um algoritmo de probabilidade a partir de uma técnica de classificação de dados, são essencialmente, previsores de atributos categóricos ou discretos. Usada em Machine Learning é uma aplicação do Teorema de Bayes, que é uma fórmula de probabilidade que calcula a possibilidade de um evento ocorrer.

A aplicação desse modelo é fazer o cálculo da probabilidade baseado nos dados que são disponibilizados ao algoritmo, é um modelo simples e rápido, que em geral possui um bom desempenho de classificação. No projeto é bom, pois só precisa de um pequeno número de dados para concluir classificações com uma boa precisão.

4.4.1.6 Modelo de Redes Neurais

O modelo de redes neurais é estruturado sobre o conceito de que a lógica de funcionamento subjacente ao processo cognitivo humano é praticamente superior às demais. Isto manifesta-se fisicamente, em síntese, no traduzir da biologia presente na arquitetura de um neurônio e de conjuntos de neurônios em fórmulas matemáticas que podem ser interpretadas por um computador de maneira que é originada AI.



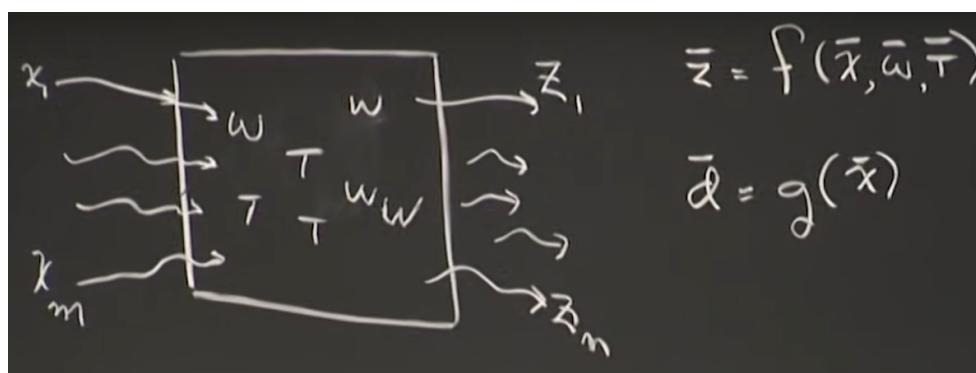
Screenshot de palestra do MIT disponibilizada publicamente

Na imagem:

- X_1 e X_n : input value - 1 ou 0.
- W_1 e W_n : weights ("pesos") - pode ser mais ou pode ser menos forte; se for mais, w_1 aumenta; se for menos W_n diminui. Isso reflete a influência da sinapse na decisão do axon de ser estimulado por inteiro.

- X: multiplicado por peso.
- E: executa inputs pelo somatório para agrupá-los e adquirir força coletiva.
- T: Para decidir se tal força coletiva de todos esses inputs é suficiente para fazer o neurônio disparar, execute o somatório via um threshold box que exibe a relação entre o input e o output. Nada acontece até que o input exceda um threshold 'T'. Se exceder →
- Output Z é um 1; se não, é um 0.
- Isto é, binário entra, binário sai. Nós modelamos os pesos sinápticos a partir desses multiplicadores; nós modelamos os efeitos cumulativos de todo esse input até o neurônio por um somatório, nós decidimos se será um tudo-ou-nada ao executá-lo por um somatório via o threshold box e verificar se a soma dos produtos são maiores do que o threshold. Se sim, recebemos um 1.

E o que faz um conjunto desses neurônios? Para mais fácil explicação, consideremos um crânio - uma grande caixa, repleta de neurônios, que, por sua vez, são repletos de weights e de thresholds.



Screenshot de palestra do MIT disponibilizada publicamente

Nessa caixa, entram uma variedade de inputs, de X_1 até X_m , que são capazes de orientar-se dentro da entropia. E do outro lado saem uma variedade de outputs, de c_1 até Z_n . Ou seja, por meio da influência dos weights e dos thresholds, esses inputs saem como outputs. Matematicamente, isso equivale a " $z = f(x, w, t)$ ", isto é, "z" é uma função do vetor input, do vetor weight, e do vetor threshold.

Isso é tudo o que é uma rede neural. E quando treinarmos uma rede neural, tudo que poderemos fazer consiste em ajustar tais weights e tais thresholds de maneira que o que sai é o que queremos. Em síntese, uma rede neural é um aproximador de funções.

$d = g(x)$: Por exemplo, talvez tenhamos sample data que nos cede um output vector que é desejado como outra função do input, esquecendo sobre o que são os weights e os thresholds. E é isso o que queremos que saia.

4.5. Avaliação

Nesta seção, descreva a solução de modelo preditivo, e justifique a escolha. Alinhe sua justificativa com a seção 4.1, resgatando o entendimento do negócio e explicando de que formas seu modelo atende os requisitos. Não deixe de usar equações, tabelas e gráficos de visualização de dados para melhor ilustrar seus argumentos.

Utilizamos os mesmos métodos de avaliação para todos os modelos, para tornar possível uma comparação mais clara entre eles. Logo, as ferramentas utilizadas foram a Matriz de confusão e a Taxa de acurácia.

4.5.1 Matriz de confusão

A Matriz de confusão consiste em uma matriz - que pode ser transformada em uma tabela - que exibe as frequências de classificação para cada classe de um modelo (mostra o número de previsões corretas e incorretas em cada classe), permitindo uma análise mais visual do desempenho de um algoritmo.

		Valor Preditivo	
		Sim	Não
Real	Sim	Verdadeiro Positivo (TP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (TN)

Fonte: <https://diegonogare.net/2020/04/performance-de-machine-learning-matriz-de-confusao/>

No nosso contexto, a tabela possui, em coluna, o que o modelo respondeu (rótulo), ou seja, se o colaborador possui tendência a sair ou não; e em linha, o que realmente aconteceu - se o colaborador saiu ou não, de fato.

Nesse sentido, nossas frequências seguem a seguinte regra:

- TP - true positive:
 - Quando o modelo diz que há tendência de sair e o colaborador realmente saiu.
- FP - false positive:
 - Quando o modelo diz que há tendência de sair e o colaborador não saiu.
- TN - true negative:

- Quando o modelo diz que não há tendência de sair e o colaborador não sai.
- FN - false negative:
 - Quando o modelo diz que não há tendência de sair e o colaborador sai.

4.5.2 Taxa de Acurácia

A partir dos conceitos apresentados na matriz de confusão, é possível calcular a taxa de acertos/acurácia dos modelos.

O cálculo da acurácia consiste na fórmula “**acurácia**= $\frac{VN+VP}{VP+FN+VN+FP}$ ”, em que: VP = Verdadeiros Positivos, VN = Verdadeiros Negativos, FP = Falso Positivos e FN = Falso Negativo. Para esse cálculo, utilizamos “dataset.score” no código dos modelos

4.5.3 Experimentações

4.5.4 Comparação de Modelos

A partir da avaliação dos modelos, tornou-se possível comparar os resultados das experimentações de cada modelo, através da criação de tabelas e ferramentas visuais. Dessa forma, conclusões foram alcançadas em relação tanto à escolha das feature engineerings como à escolha dos modelos mais precisos para o objetivo do projeto - classificar os funcionários para saber se eles têm ou não chance de saírem da empresa.

Uma etapa importante do processo de comparação dos modelos foi a comparação das **taxas de acurácia** de teste que cada um apresentou para cada combinação de possíveis variáveis que se mostraram mais relevantes. Dessa forma, foi gerada uma tabela (tabela 1), em que as linhas representam cada modelo testado, e as colunas representam cada combinação de variáveis. Além disso, a coluna “Média de acurácia” apresenta a acurácia média de cada modelo diante das experimentações feitas.

tabela 1:

	Média de acurácia	tamanho do teste	Idade, Cargo, Regiao, SP, Salario Comparado	faixa etaria, Cargo, Regiao, EstadoSP, Salario Mês, Salario Comparado	faixa etaria, Salario Comparado, SP	faixa etaria, Salario Comparado, SP, Estagnação	faixa etaria, Salario Comparado, SP, Estagnação, Cargo	Idade, Cargo, Região, Salario Comparado, Estagnação	Idade, Salario Comparado, SP, Estagnação	faixa etaria, Cargo, Regiao, Salario Comparado, Estagnação
KNN	67%	0.3	0.65 - 65%	0.63 - 63%	0.69 - 69%	0.69 - 69%	0.67 - 67%	0.69 - 69%	0.69 - 69%	0.67 - 67%
Árvore de decisão	72%	0.3	0.6783 67%	0.6783 67%	0.7343 73%	0.7482 74%	0.7622 76%	0.7203 72%	0.7063 70%	0.7762 77%
SVM	64%	0.3	0.678321 - 68%	0.608391 - 60.8%	0.699300 - 70%	0.62937 - 63%	0.62937 - 63%	0.62937 - 63%	0.62937 - 63%	0.62937 - 63%
Naive Bayers	67%	0.3	0.650349 - 65%	0.601398 - 60%	0.678321 - 67%	0.67832 - 67%	0.67132 - 67%	0.69930 - aprox 70%	0.67132 - 67%	0.72027 - 72%
Regressão logística	71%	0.3	0.699300 - aprox 70%	0.748251 - 75%	0.727272 - 72%	0.748251 - 74%	0.720279 - 72%	0.706293 - 70%	0.741258 - 74%	0.692307 - 69%

Nesse contexto, é possível analisar que os modelos de Árvore de decisão e de Regressão Logística foram os que apresentaram os melhores desempenhos em relação aos outros, com média de acurácia de 72% e 71%, respectivamente, e atingiram as maiores taxas individuais da tabela (a Árvore de decisão com 77%). A partir disso, vamos para a análise dos conjuntos de variáveis, especialmente nesses dois modelos, e percebe-se que o uso das variáveis “Salário Mês” e “idade” diminuem a acurácia, levando à não consideração delas nos próximos experimentos. Além delas, anterior a essas testagens, foi percebida uma não eficácia da variável “Gênero”, que foi descartada antes mesmo da rodada oficial de testes.

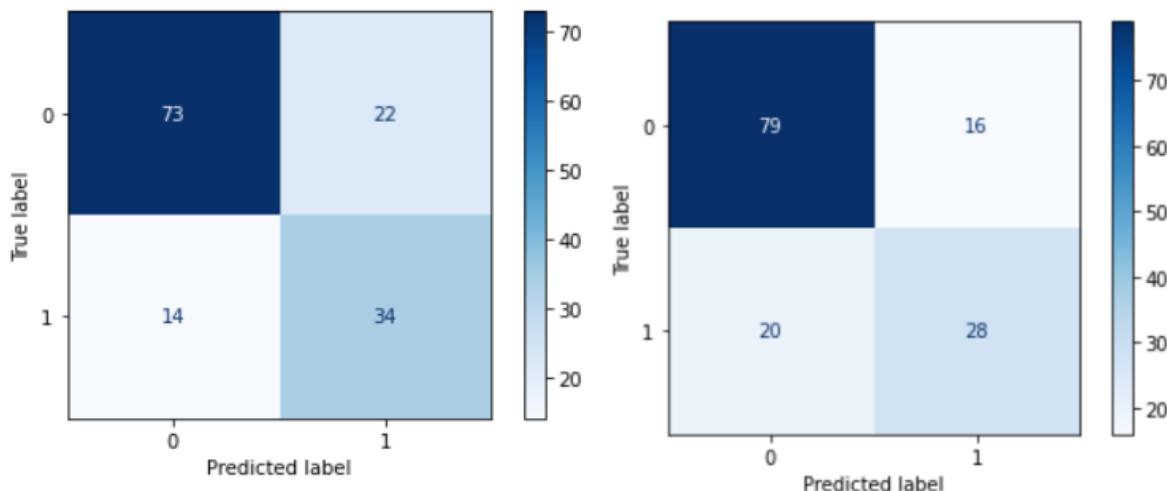
Dessa forma, as 3 experimentações de melhor desempenho estão em destaque em azul na tabela 1, e foram separadas para melhor visualização, gerando uma nova tabela (tabela 2).

tabela 2:

	<u>faixa etaria, Salario Comparado, SP, Estagnação</u>	<u>faixa etaria, Salario Comparado, SP, Estagnação, Cargo</u>	<u>faixa etaria, Salario Comparado, Região, Estagnação, Cargo</u>
Árvore de decisão	0.7482 74%	0.7622 76%	0.7762 77%
Regressão logística	0.7482 - 74%	0.7202 - 72%	0.6923 - 69%

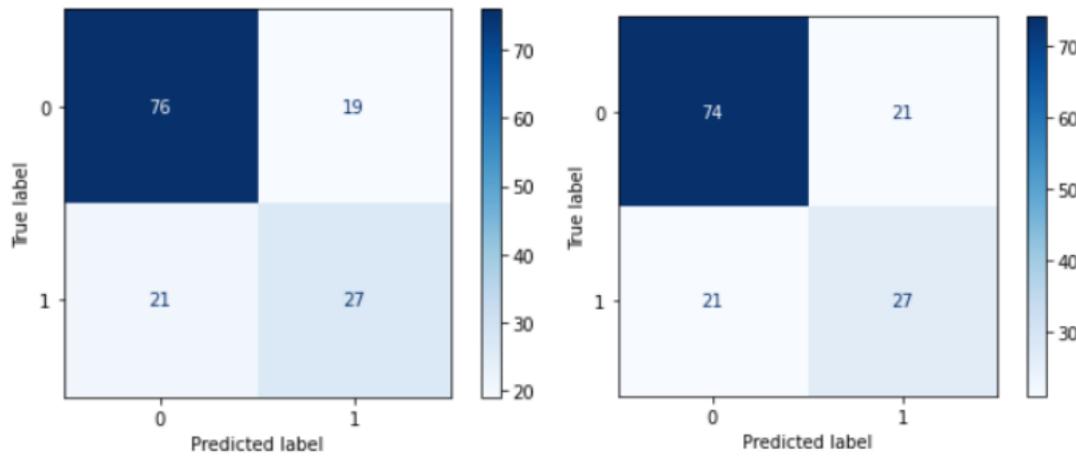
Pode-se perceber que o conjunto de variáveis referentes à “faixa etária, salário comparado, SP e estagnação” gerou uma acurácia de aproximadamente 74% para os dois modelos (1º experimento), e, com a adição da variável “cargo”, a acurácia da Árvore de decisão aumentou em aproximadamente 2%, mas a acurácia da regressão logística diminuiu em 2%, ilustrando como as variáveis podem diferir um pouco de comportamento entre modelos diferentes quando falamos de acurácia, e é por isso que utilizamos outras formas de avaliação para complementar a anterior, como a matriz de confusão. Segundo na análise, o mesmo acontece com a substituição da variável “SP”, pela variável “Região”, que aumenta a acurácia da árvore de decisão, mas a da regressão logística diminui.

Logo, diante das análises feitas através da acurácia, seguimos para as análises pela **matriz de confusão**, para decisões mais profundas sobre as ocorrências dos erros e acertos do modelo.

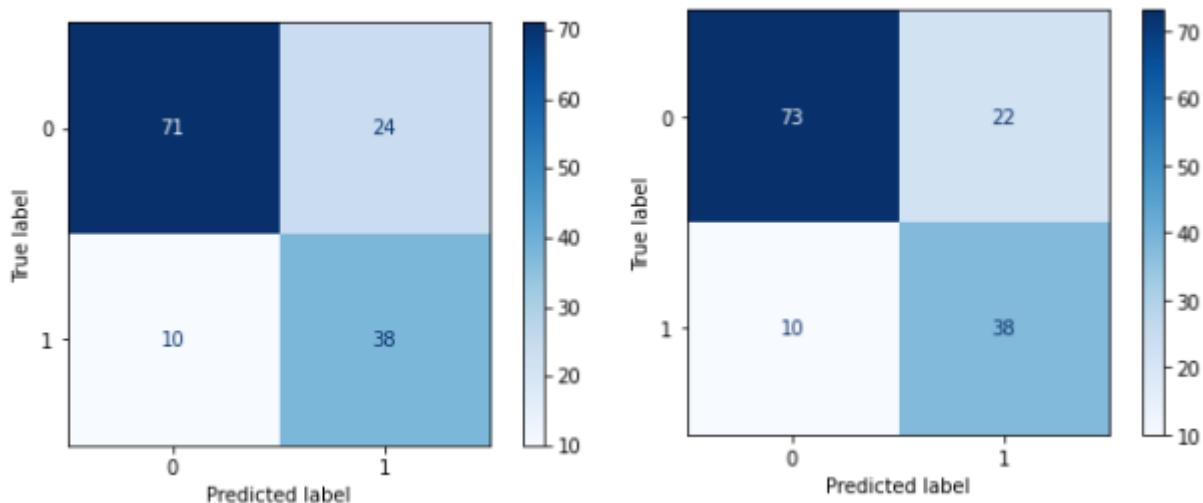


A matriz de confusão da esquerda tem acurácia de 74,8% é de árvore de decisão e a da direita tem acurácia de 74,8% é de regressão logística, apesar de ambas possuírem a mesma acurácia

a matriz da esquerda é considerada melhor, pois possui a menor taxa de erro no quadrante inferior esquerdo, que é a métrica mais importante para o projeto, significa menor erro para dizer quando o colaborador tem pretensão de sair da empresa. E o erro do canto superior esquerdo fala que o colaborador vai sair e ele não vai.



A matriz de confusão do lado esquerdo tem 72,2% de acurácia e a direita tem 69% de acurácia, ambas de regressão logística e possuem quase as mesmas taxas de erro, no entanto as duas combinações de variáveis que formaram essas matrizes foram descartadas para regressão logística pois possuem uma acurácia menor do que as demais combinações para o modelo de regressão.



As duas matrizes são de árvore de decisão, a primeira possui acurácia de 76% e a segunda de 77%, as duas possuem a mesma taxa de erro na métrica mais importante, do quadrante inferior

esquerdo, então portanto olhamos para o superior direito, e nessa métrica, a de 77% possui uma menor taxa de erro, e por conta disso é o melhor modelo encontrado até o atual momento.

analise pelas matrizes de confusao

conclusoes finais sobre a combinação das analises, explicando porque escolhemos continuar explorando esses 2 modelos

print da melhor árvore de decisão

5. Conclusões e Recomendações

oq tem que fazer ? to terminando a avaliacao pela acuracia, e tem que fazer melas matrizs de confusao a matriz dos que estão na tabela ??isso, da tabrla 2 todos ou alguma já tem ?? as que tem ta no notionn tem que analisar elas, as vezes nao precisa te todas, so as 4 melhores acuraciad, ouentao faz dessas todas da tabela 2, mostrando pq vai descartar as outras

posso

Escreva, de forma resumida, sobre os principais resultados do seu projeto e faça recomendações formais ao seu parceiro de negócios em relação ao uso desse modelo. Você pode aproveitar este espaço para comentar sobre possíveis materiais extras, como um manual de usuário mais detalhado na seção “Anexos”.

Não se esqueça também das pessoas que serão potencialmente afetadas pelas decisões do modelo preditivo, e elabore recomendações que ajudem seu parceiro a tratá-las de maneira estratégica e ética.

6. Referências

Nesta seção você deve incluir as principais referências de seu projeto, para que seu parceiro possa consultar caso ele se interessar em aprofundar.

Utilize a norma ABNT NBR 6023 para regras específicas de referências. Um exemplo de referência de livro:

SOBRENOME, Nome. **Título do livro**: subtítulo do livro. Edição. Cidade de publicação: Nome da editora, Ano de publicação.

- CHAPMAN, Pete; CLINTON, Julian; KERBER, Randy; KHABAZA, Thomas; REINARTZ Thomas; SHEARER, Colin; WIRTH, Rüdiger. CRISP-DM 1.0: Step-by-step Data Mining Guide. SPSS, 2000
- Imagem arborescente por Do not want - Own work, CC BY 3.0, <https://commons.wikimedia.org/w/index.php?curid=14947263>
- DELEUZE, Gilles; GUATTARI, Felix. (1987) [1980]. A Thousand Plateaus. Translated by Massumi, Brian. University of Minnesota Press.
- LAND, Nick. Fanged Noumena: Collected Writings 1987-2007, ed. Robin Mackay and Ray Brassier (Urbanomic, 2011). ISBN 978-0955308789
- <https://blog.swile.com.br/rotatividade-de-funcionarios-fatores-que-influenciam-a-saude-da-empresa/>
-

Anexos

Utilize esta seção para anexar materiais como manuais de usuário, documentos complementares que ficaram grandes e não couberam no corpo do texto etc.