



# 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 da seção 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) >
< >	< >	< >	< >

# Sumário

<b>1. Introdução</b>	<b>5</b>
<b>2. Objetivos e Justificativa</b>	<b>6</b>
2.1. Objetivos	6
2.2. Justificativa	6
<b>3. Metodologia</b>	<b>7</b>
3.1. CRISP-DM	7
3.2. Ferramentas	7
3.3. Principais técnicas empregadas	7
<b>4. Desenvolvimento e Resultados</b>	<b>8</b>
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
<b>5. Conclusões e Recomendações</b>	<b>14</b>
<b>6. Referências</b>	<b>15</b>
<b>Anexos</b>	<b>16</b>

# 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 sede da Everymind localiza-se na Rua Alexandre Dumas, 1711, em Santo Amaro, São Paulo - SP, CEP 04717-004.

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.

**Descreva resumidamente o problema a ser resolvido (sem ainda mencionar a solução).**

**Caso utilize citações ao longo desse documento, consulte a norma ABNT NBR 10520. Sugerimos o uso do sistema autor-data para citações.**

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] visitar 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 diminuiriam 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 inerentemente 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
- 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

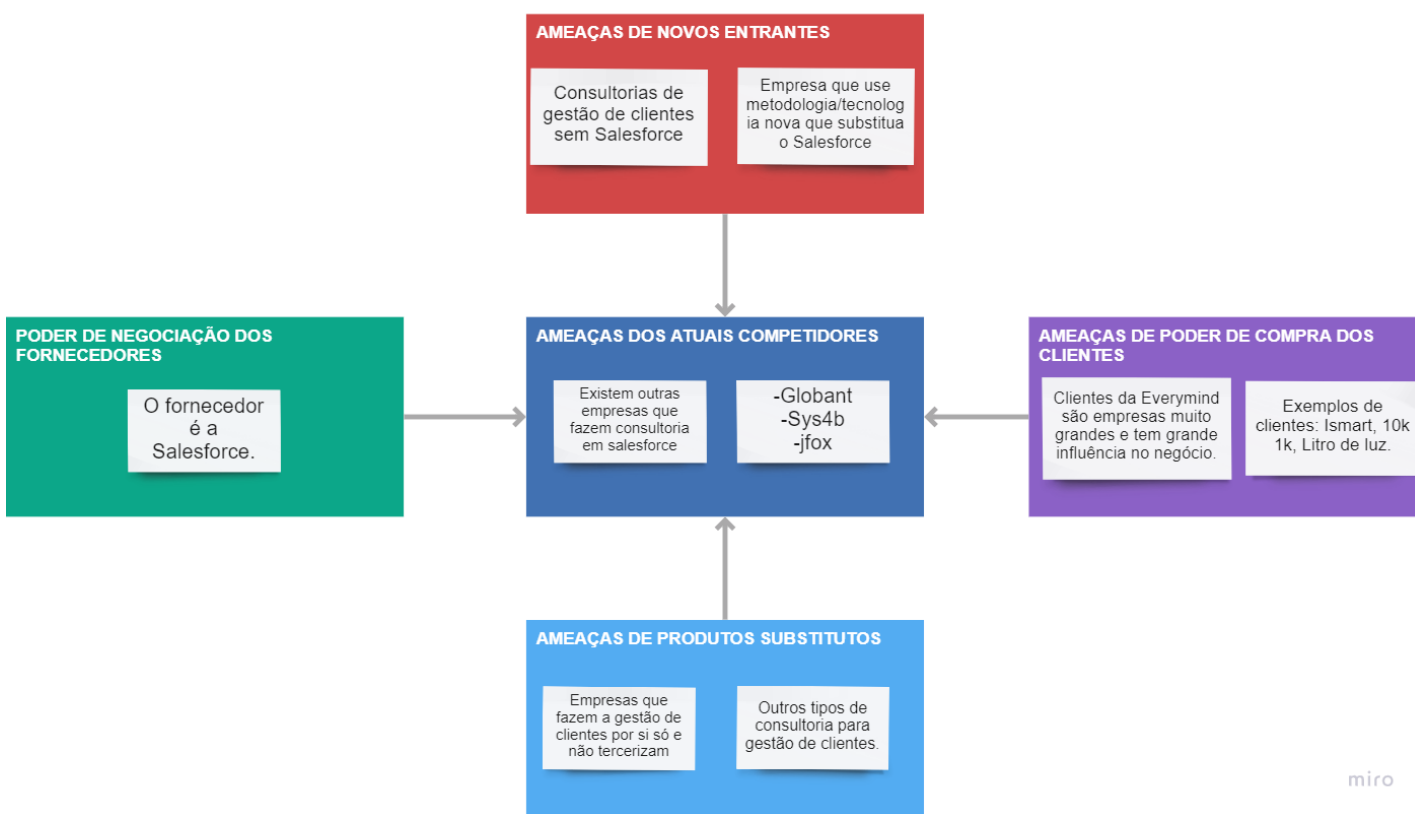
## 4. Desenvolvimento e Resultados

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/uXjVOgCyebl=?share\\_link\\_id=925777377534](https://miro.com/app/board/uXjVOgCyebl=?share_link_id=925777377534)

miro

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.

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 detrimntosos à performance.



### 4.1.2. Análise SWOT



[https://miro.com/app/board/uXjVOgCyebl=/?share\\_link\\_id=925777377534](https://miro.com/app/board/uXjVOgCyebl=/?share_link_id=925777377534)

### 4.1.3. Planejamento Geral da Solução

#### a) Problema a ser resolvido

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] visitar os demais processos de carreira e [de] desenvolvimento" (descrição oficial da demanda).

## 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] visitar os demais processos de carreira e [de] desenvolvimento" (descrição oficial da demanda).

## d) Tipo de tarefa (regressão ou classificação)

Tendo em mente os fatos de que em árvores de classificação as variáveis dependentes são categóricas, e de que em árvores de regressão as variáveis numéricas são dependentes, e que os dados sobre os quais será construído o algoritmo ML originam-se principalmente de spreadsheets, pode-se concluir que mostra-se mais apropriado utilizar o método regressivo para o desenvolvimento da AI requisitada.

## e) Como a solução proposta deverá ser utilizada

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

- 1- AI é construída;
- 2- Preparação dos dados (organização e análise);
- 3- Dados são minerados;
- 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.

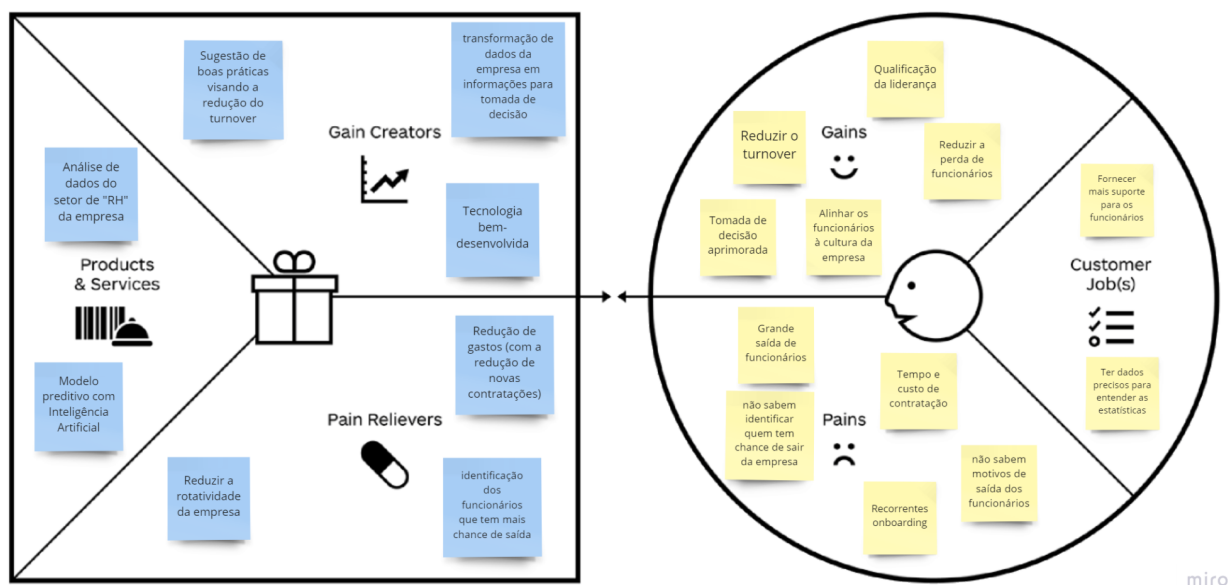
#### **f) Benefícios trazidos pela solução proposta**

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/uXjVOgCyebI=/?share\\_link\\_id=925777377534](https://miro.com/app/board/uXjVOgCyebI=/?share_link_id=925777377534)

## 4.1.5. Matriz de Riscos

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

[https://miro.com/app/board/uXjVOgCyebI=/?share\\_link\\_id=925777377534](https://miro.com/app/board/uXjVOgCyebI=/?share_link_id=925777377534)

### Lista de riscos:

1- Variáveis pouco claras

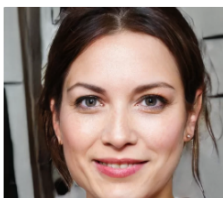
- 2- Falta de dados necessários
- 3- Resposta pouco específica/subjetiva
- 4- Falta de experiência do time ao utilizar as ferramentas novas
- 5- Não alcançar expectativas do cliente
- 6- Falta de organização e gestão de tempo
- 7- Mau entendimento sobre o contexto da indústria de Sales force
- 8- Tecnologias 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

#### **4.1.6. Personas**

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

## 1ª 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

## 2ª 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

### 3ª 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

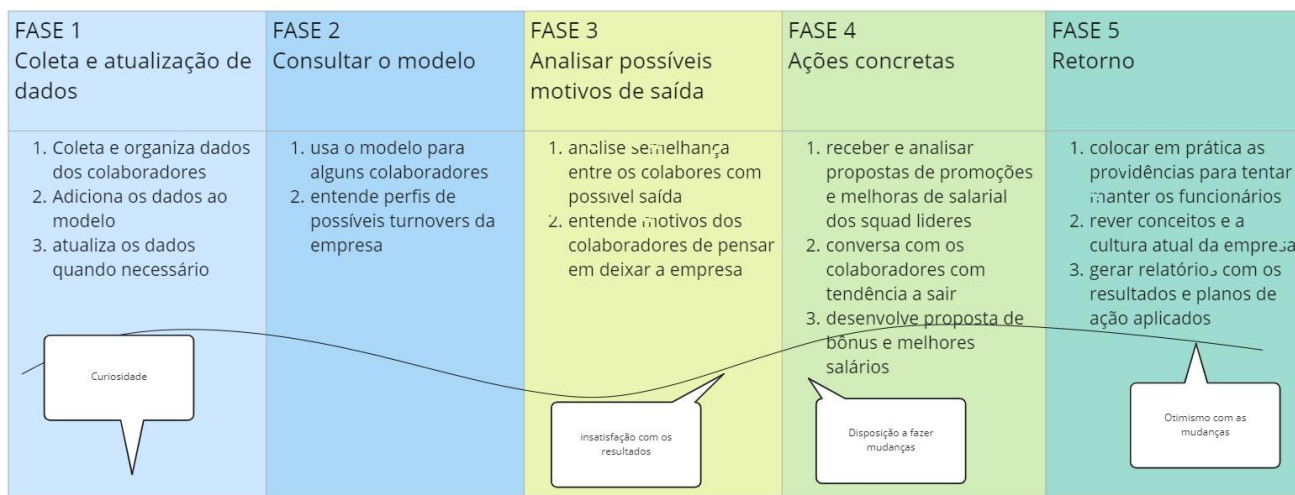


**Luisa**

**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



### 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.



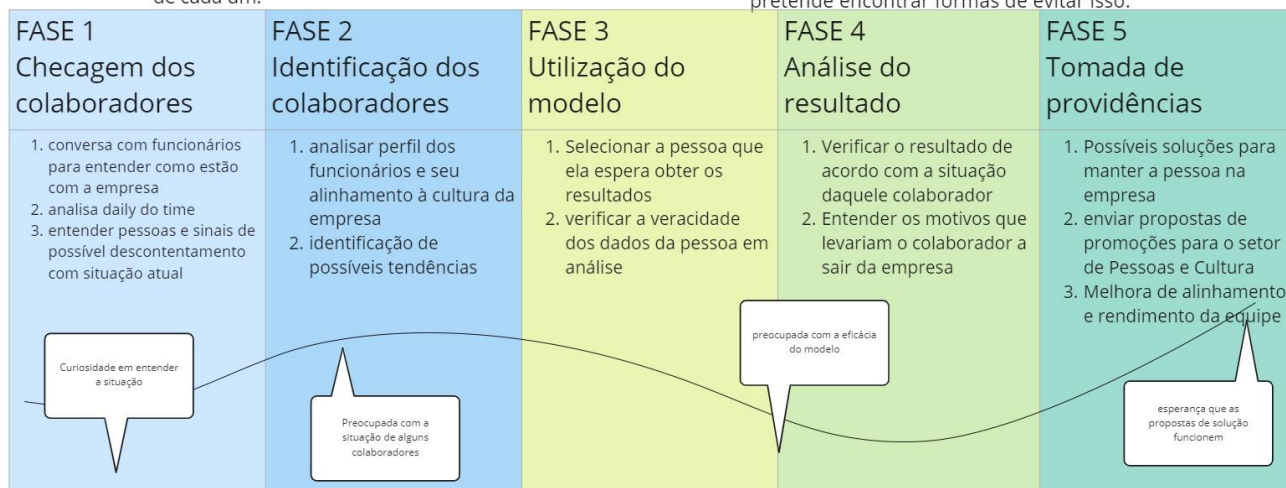


### Janice

**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.

### 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.



### 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.

### Responsabilidades

Analisar 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

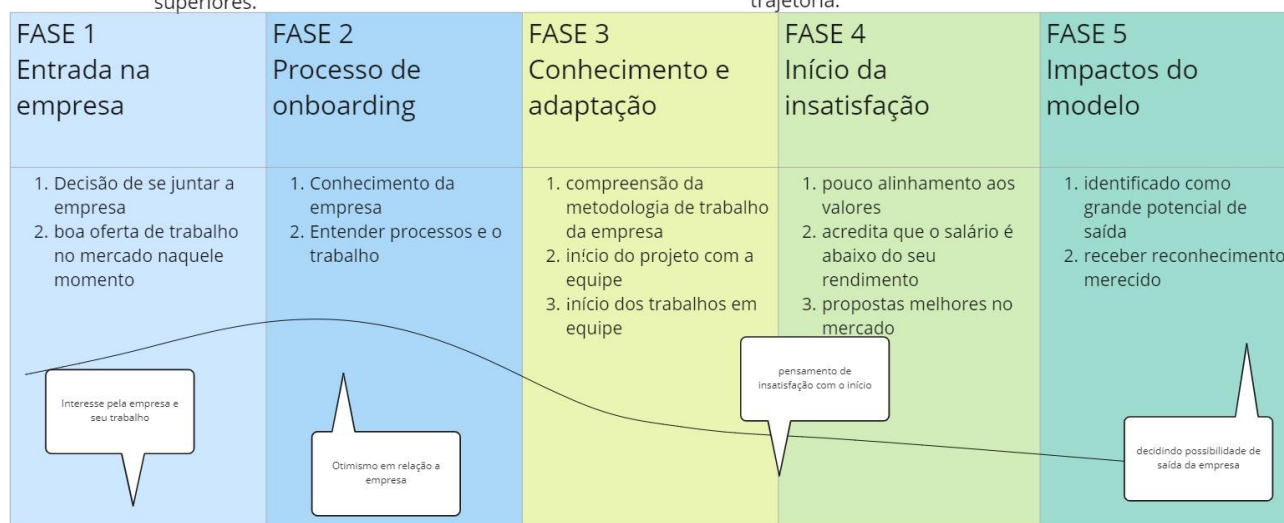


### Jonas

**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.

### Expectativas

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



### 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.

### 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.

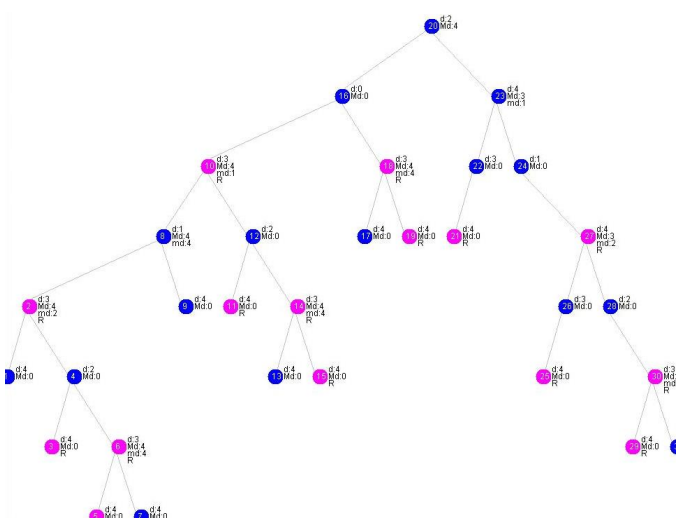
➡ Dados referentes às perguntas: divisão, pilar, pontuação, fator, pontuação, pergunta, pulou, muito insatisfeito, insatisfeito, neutro, satisfeito, muito satisfeito, taxa de confiabilidade.

➡ Spreadsheet "Chart1"

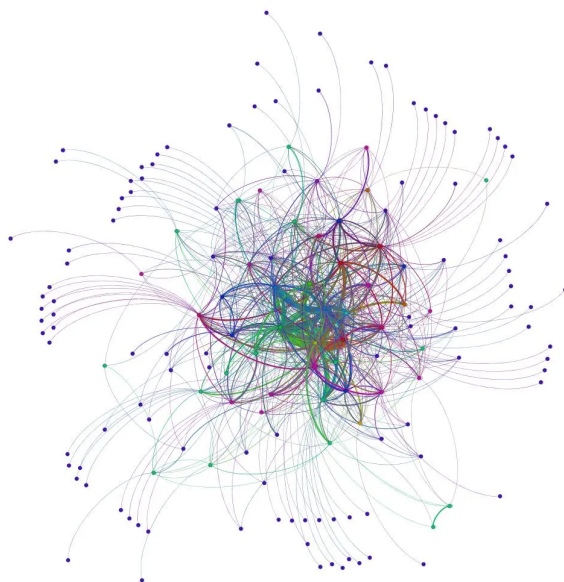
➡ Dado é: gráfico demonstrando trendline para respectivas quantidades de "muito satisfeito" (vide seção Dados do Spreadsheet "Ambiente de Trabalho 27.07").

### 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 significativa: 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]

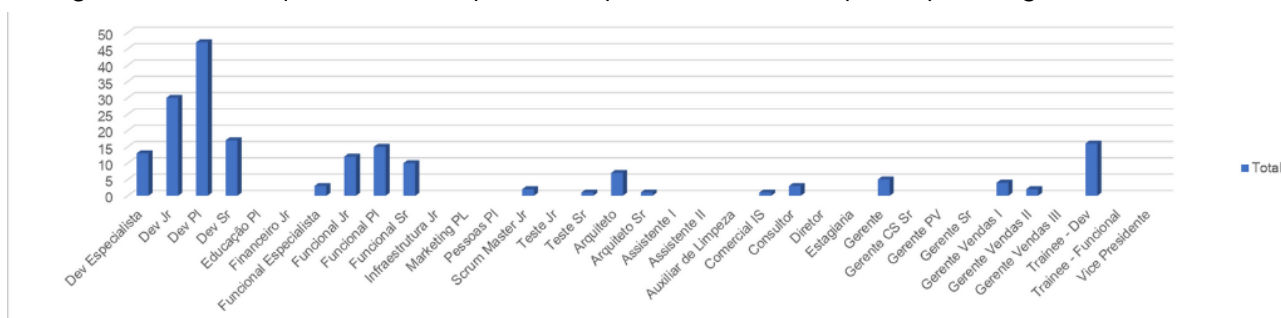
**d) Se houver, descrição das restrições de segurança.**

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.



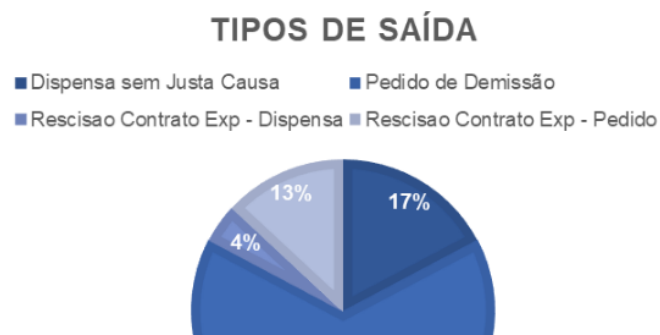
#### 2. Saídas por idade

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



#### 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 Saida’, 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 Saida'.  
#.isnull() Ela pega todos os valores Null  
Salario_ativos = planilha[planilha['Tipo Saida'].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 Saida’, e filtramos usando o comando ‘dropna()’ que retorna apenas o 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    4486.965
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.

## 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.

```
#Função pega a data de admissão do colaborador e a data do seu desligamento, e encontra o período entre elas.
#.dropna() Ela dropa todos os dados que possuem valores Na...
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 da 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**

Para a Sprint 3, você deve descrever aqui os experimentos realizados com os modelos (treinamentos e testes) até o momento. Não deixe de usar equações, tabelas e gráficos de visualização de dados para melhor ilustrar seus experimentos e resultados.

Para a Sprint 4, você deve realizar a descrição final dos experimentos realizados (treinamentos e testes), comparando modelos. Não deixe de usar equações, tabelas e gráficos de visualização de dados para melhor ilustrar seus experimentos e resultados.

## 4.5. Avaliação

Nesta seção, descreva a solução final 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.

## 4.6 Comparação de Modelos

# 5. Conclusões e Recomendações

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
- 

## Anexos

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