

**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)> |
| <10/08/2022> | <Daniel Dávila> | <1.2> | <criação das seções 4.1.3 e 4.2.2> |
| <12/08/2022> | <Kil Mateus> | <1.3> | <atualização da seção 4 (4.2)> |
| <12/08/2022> | <Daniel Dávila> | <1.4> | <atualização da seção 4.2.2> |

**Sumário**

[**1. Introdução**](#_heading=h.2et92p0) **5**

[**2. Objetivos e Justificativa**](#_heading=h.tyjcwt) **6**

[2.1. Objetivos](#_heading=h.3dy6vkm) **6**

[2.2. Justificativa](#_heading=h.4d34og8) 6

[**3. Metodologia**](#_heading=h.2s8eyo1) **7**

[3.1. CRISP-DM](#_heading=h.17dp8vu) 7

[3.2. Ferramentas](#_heading=h.3rdcrjn) 7

[3.3. Principais técnicas empregadas](#_heading=h.26in1rg) 7

[**4. Desenvolvimento e Resultados**](#_heading=h.lnxbz9) **8**

[4.1. Compreensão do Problema](#_heading=h.35nkun2) 8

[4.1.1. Contexto da indústria](#_heading=h.1ksv4uv) 8

[4.1.2. Análise SWOT](#_heading=h.44sinio) 8

[4.1.3. Planejamento Geral da Solução](#_heading=h.2jxsxqh) 8 # Dados disponíveis

[4.1.4. Value Proposition Canvas](#_heading=h.z337ya) 8

[4.1.5. Matriz de Riscos](#_heading=h.3j2qqm3) 8

[4.1.6. Personas](#_heading=h.1y810tw) 9 #ver sobre líder de projeto

[4.1.7. Jornadas do Usuário](#_heading=h.4i7ojhp) 9

[4.2. Compreensão dos Dados](#_heading=h.2xcytpi) 10 # Dados disponíveis

[4.3. Preparação dos Dados](#_heading=h.1ci93xb) 11

[4.4. Modelagem](#_heading=h.3whwml4) 12

[4.5. Avaliação](#_heading=h.qsh70q) 13

[4.6 Comparação de Modelos](#_heading=h.be1cqj72p9wo) 14

[**5. Conclusões e Recomendações**](#_heading=h.3as4poj) **14**

[**6. Referências**](#_heading=h.1pxezwc) **15**

[**Anexos**](#_heading=h.49x2ik5) **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 deste módulo é a Everymind, uma consultoria de Salesforce. É uma empresa muito bem reconhecida pela própria SalesForce e uma das maiores nesse mercado. A Everymind fica localizada na Rua Alexandre Dumas, 1711, em Santo Amaro, São Paulo - SP, CEP 04717-004.

O foco da Everymind é vender soluções utilizando tecnologias de análise de dados da Salesforce possui, mas de forma aprimorada dentro do contexto de negócio para cada cliente, melhorando muito a eficiência dos produtos.

A Everymind não possui práticas de verificação de turnover eficientes, por isso eles têm uma rotatividade de funcionários alta.

**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.**

**2. Objetivos e Justificativa**

**2.1. Objetivos**

A partir da criação de um modelo preditivo com o intuito de prever o risco de funcionários deixarem a empresa, os objetivos são que esse modelo contribua para ações de retenção e redução da taxa de turnover da empresa. E com isso, é esperado um alinhamento dos colaboradores à cultura da empresa.

**2.2. Justificativa**

Nossa proposta de solução é a construção de um modelo preditivo (algoritmo de machine learning)que a partir da identificação de padrões dos dados disponibilizados, será possível uma ação imediata de análises sobre a saída de funcionários da empresa.

Os benefícios do nosso modelo incluem redução da rotatividade (turnover) de funcionários, maior alinhamento dos funcionários à cultura da empresa e melhora na tomada de decisão pelas lideranças da empresa.

O método usado em nossa solução se diferencia por ser um sistema ágil e flexível, que tem a capacidade de fazer análises sem intervenção humana, com potencial de processamento de grande quantidade de dados e identificação de padrões em um tempo extremamente curto, com grande precisão.

**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**

Descreva brevemente a metodologia CRISP-DM e suas etapas de processo [apagar depois]

"CRISP-DM" é uma sigla para "Cross Industry Standard [for] Data Mining". [...]

**3.2. Ferramentas**

Descreva brevemente as ferramentas utilizadas e seus papéis (Google Colaboratory)

**3.3. Principais técnicas empregadas**

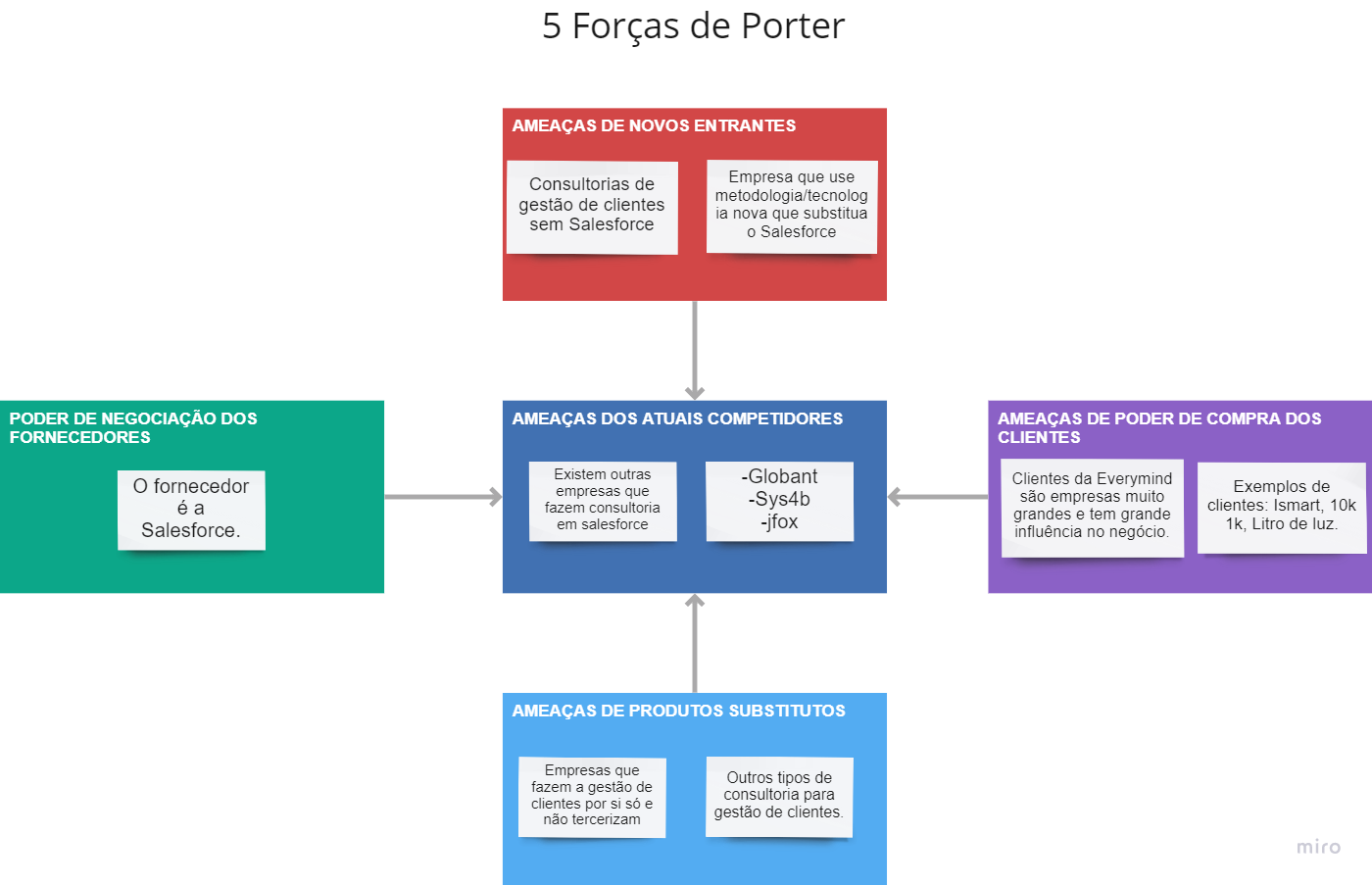
Descreva brevemente as principais técnicas empregadas, algoritmos e seus benefícios

**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**

<https://miro.com/app/board/uXjVOgCyebI=/?share_link_id=925777377534>

Os principais players da área de consultoria em Salesforce são: Everymind, Sys4b, Globant e Jfox. Por mais que a Everymind seja uma das principais empresas na área, e tenha clientes muito grandes, sempre deve haver a preocupação de se ter um diferencial.

O modelo de negócio da Everymind é vender e disponibilizar tecnologias em forma de consultoria para grandes empresas. A tecnologia em questão é a Salesforce, que pode ser usada para várias finalidades, inclusive gestão de clientes

As inteligências artificiais são usadas pelo mercado de consultorias para analisar o comportamento do cliente, como propensão a usar Salesforce, monitorar o marketplace, melhorar experiência do usuário e conseguir identificar potenciais compradores. Os ativos de TI recorrem à inteligência artificial para se antecipar aos problemas de desempenho e automatizar as correções antes que os impactos negativos sejam sentidos.

**4.1.2. Análise SWOT**



<https://miro.com/app/board/uXjVOgCyebI=/?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 descobrir 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).

**b) Dados disponíveis**

3 tabelas - “Everymind” e “Reconhecimento” - dados básicos; ”Ambiente de trabalho”

➠Spreadsheet com dados básicos sobre funcionários que saíram e que foram demitidos

➥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.

**c) Solução proposta**

Construção de algoritmo de machine learning que, após descobrir 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)**

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.

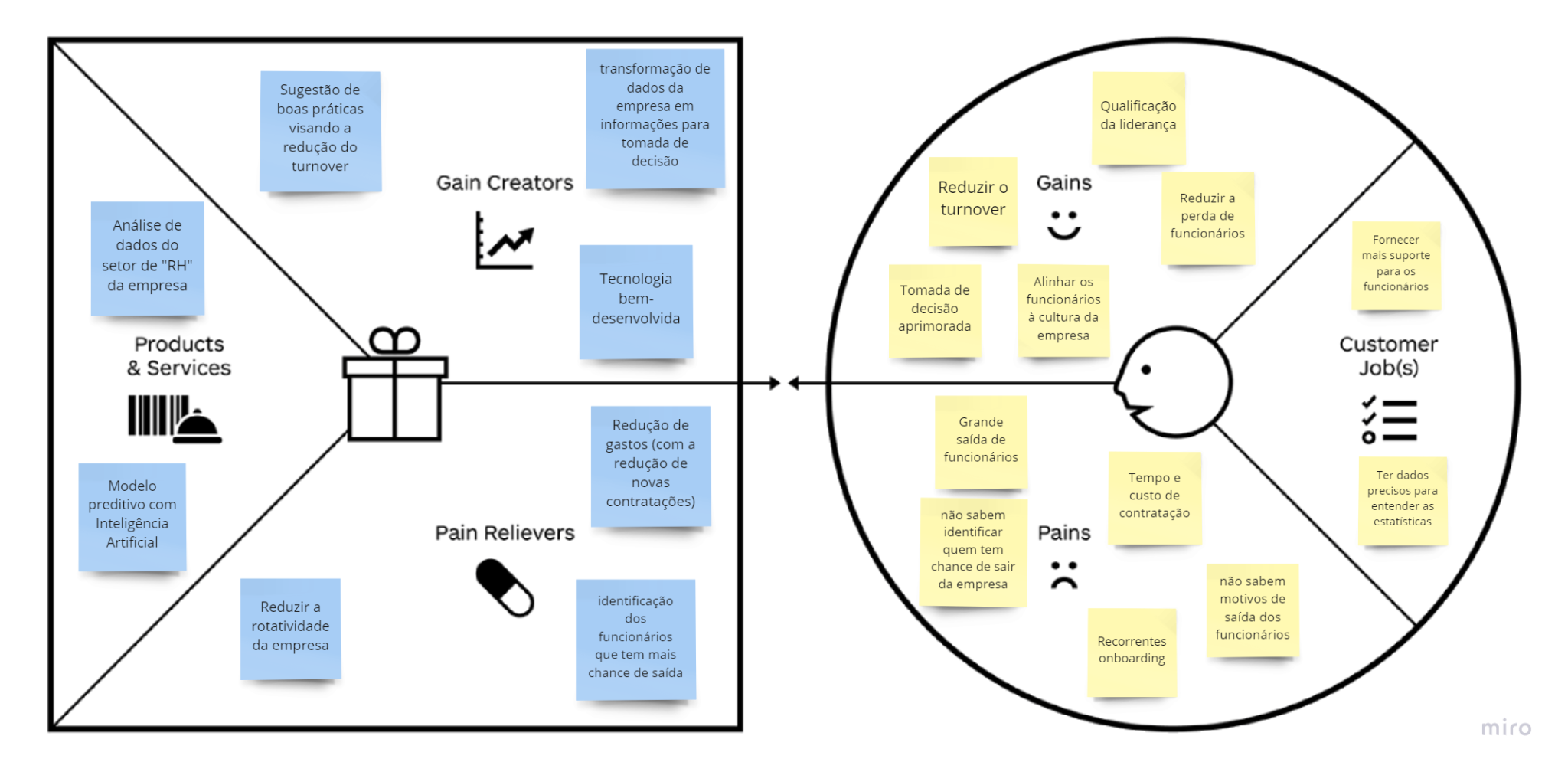
**f) Benefícios trazidos pela solução proposta**

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

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

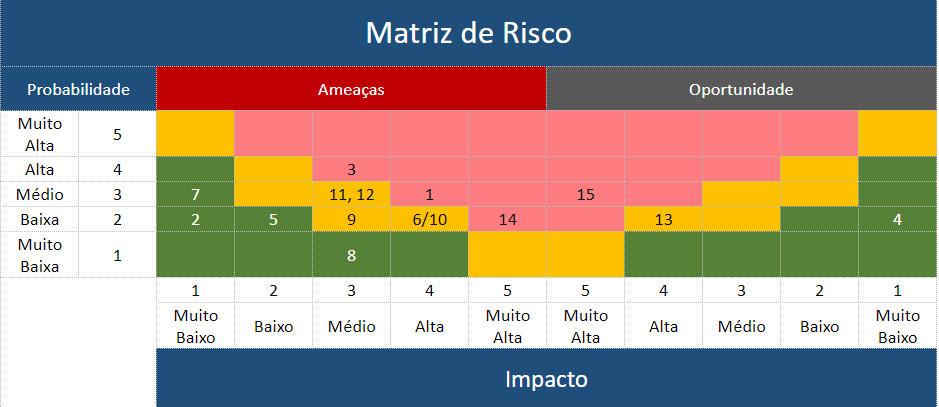
No contexto em que o índice de saída de funcionários da Everymind, atualmente alto, tornou-se baixo após a aplicação da solução que propomos, consideramos que o sucesso foi obtido.

**4.1.4. Value Proposition Canvas**



<https://miro.com/app/board/uXjVOgCyebI=/?share_link_id=925777377534>

**4.1.5. Matriz de Riscos**



<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

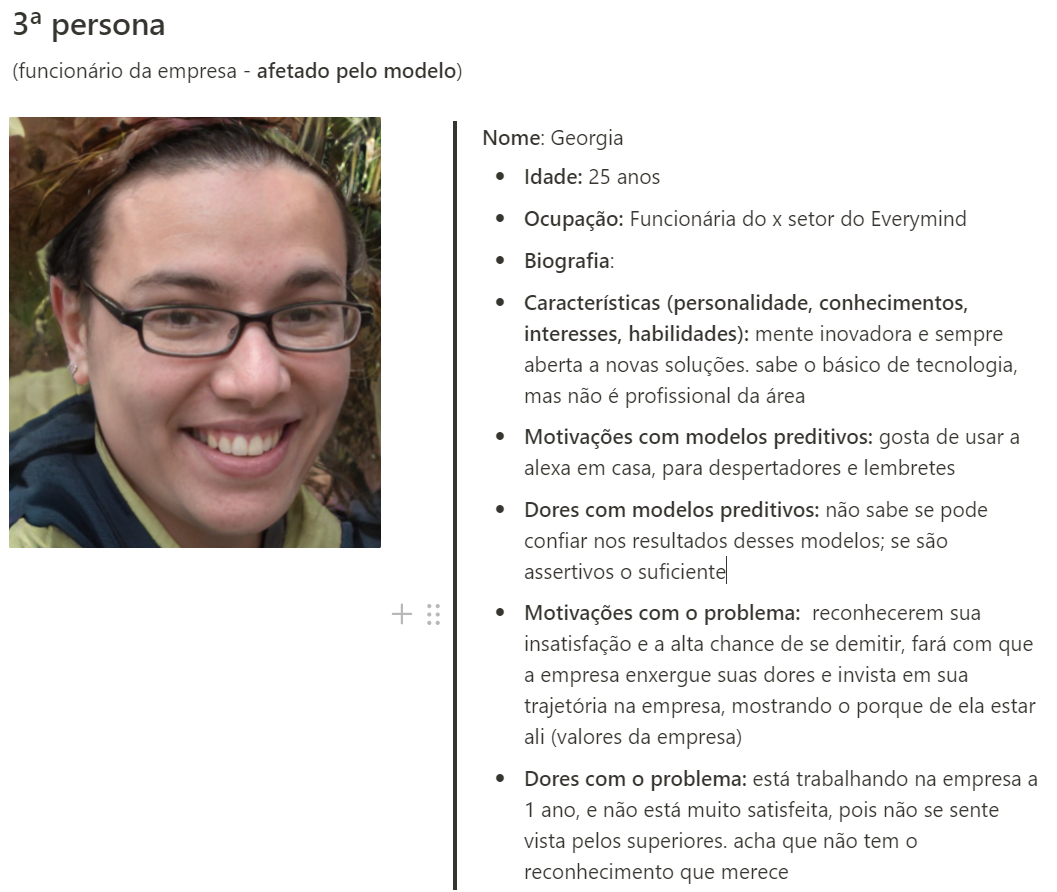
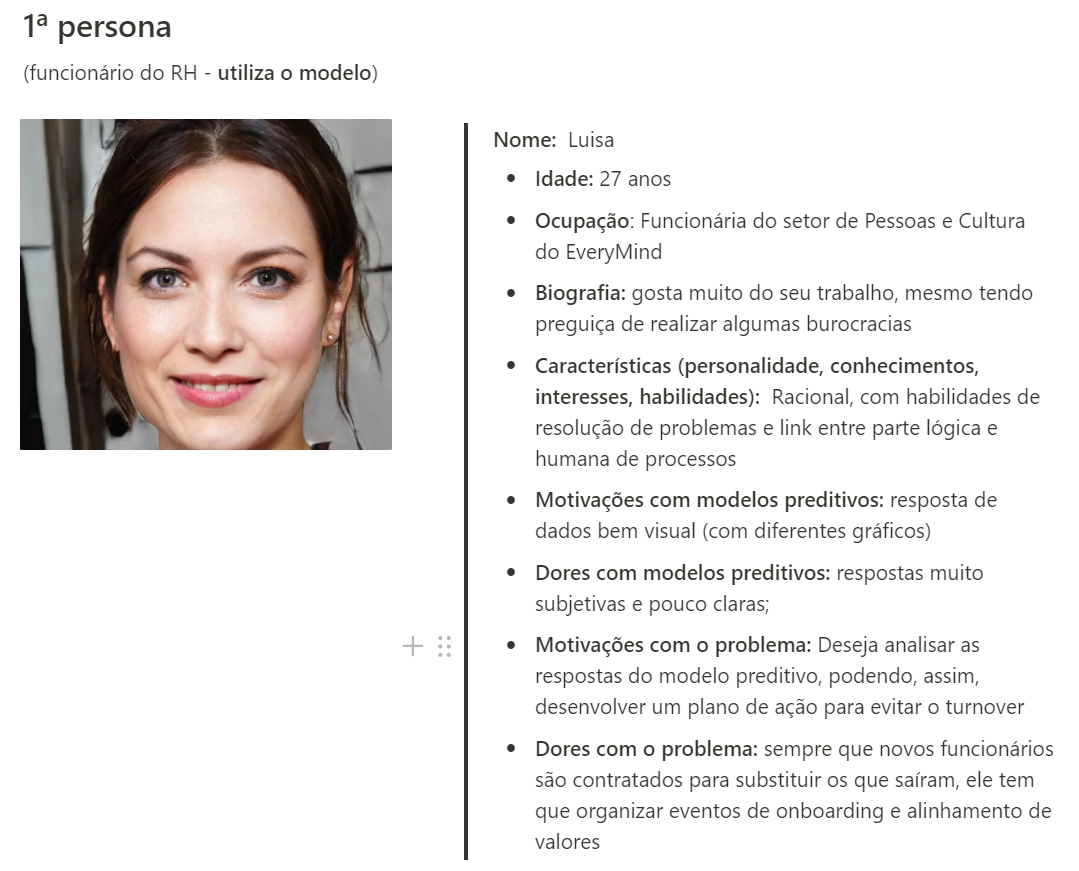
12- Má divisão de tarefas, sobrecarregando poucos

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)

**4.1.7. Jornadas do Usuário**

**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:**

➠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 são: divisão, pilar, pontuação, fator, pontuação, pergunta, pulou, muito insatisfeito, insatisfeito, neutro, satisfeito, muito satisfeito, taxa de confiabilidade.

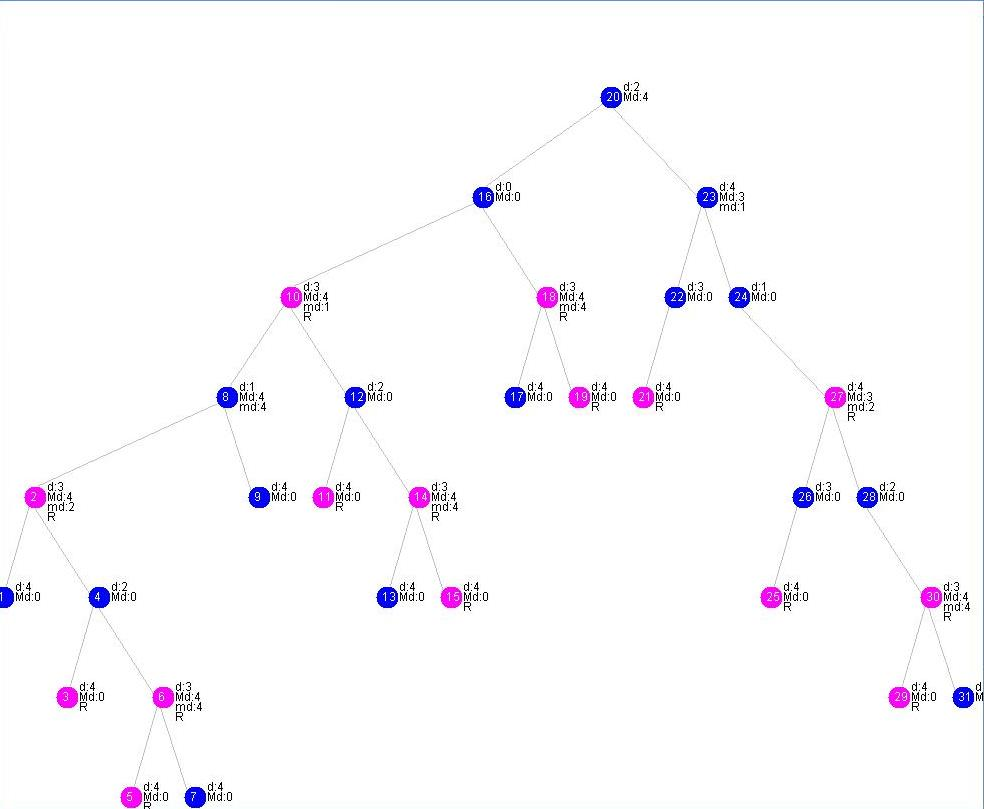
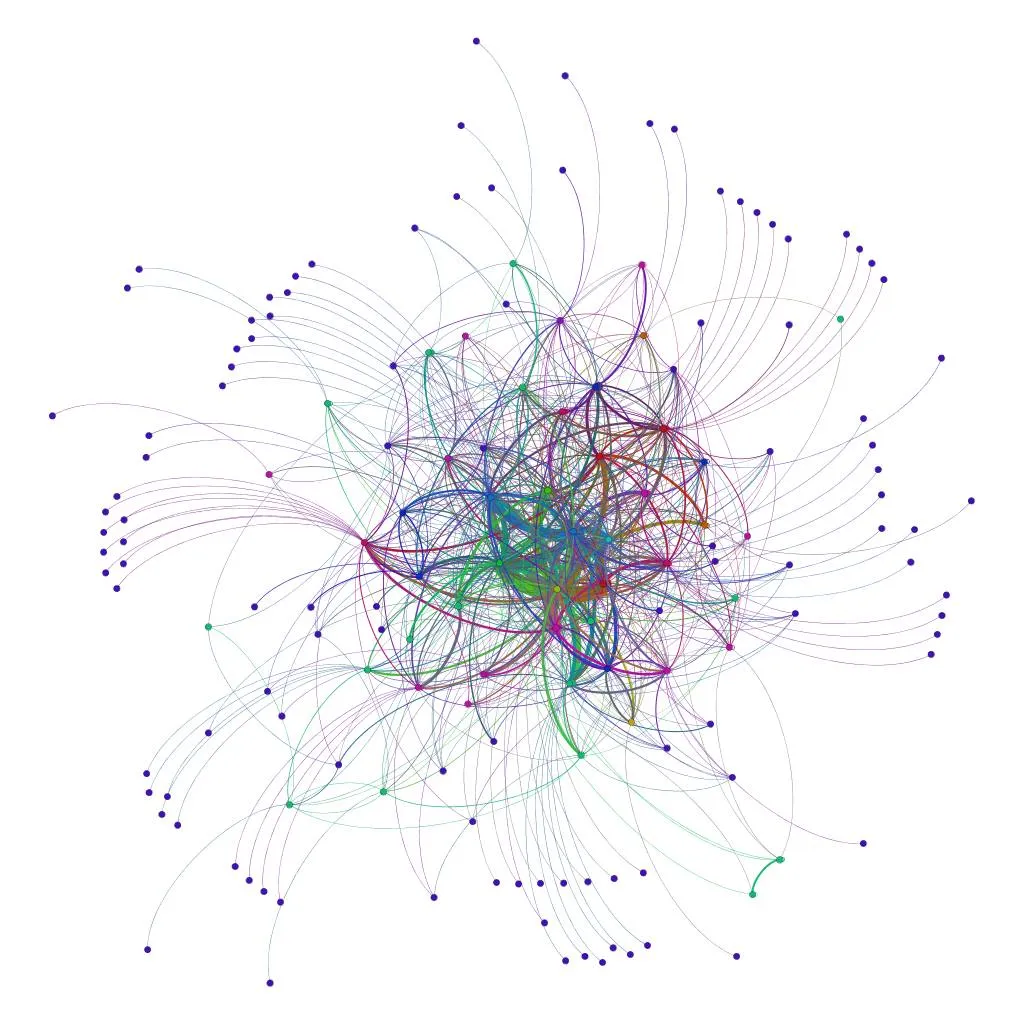
➥Spreadsheet "Chart1"

➥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 daquilo escrito no quadro, e no momento de escrita (dia 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).**

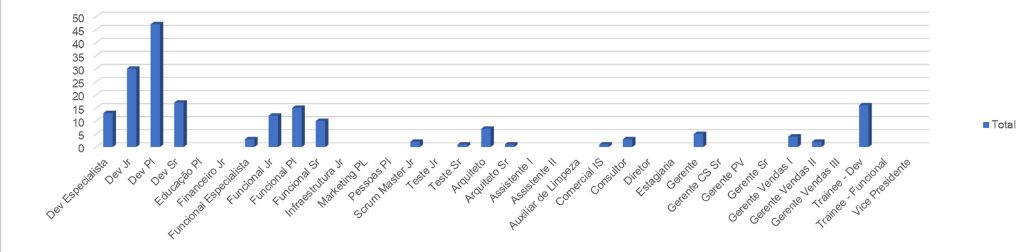
[no momento de escrita (dia 10 de agosto, 2022), não é aplicável]

**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 - como a proibição de publicá-los no GitHub.

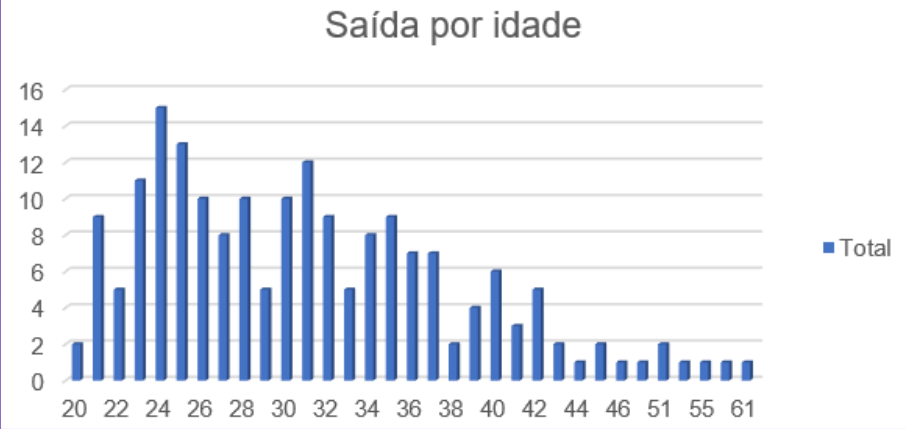
**4.2.3 Descrição estatística básica dos dados, principalmente dos atributos de interesse, com inclusão de visualizações gráficas e como essas análises embasam suas hipóteses.**

1. **Total de saídas por cargo**

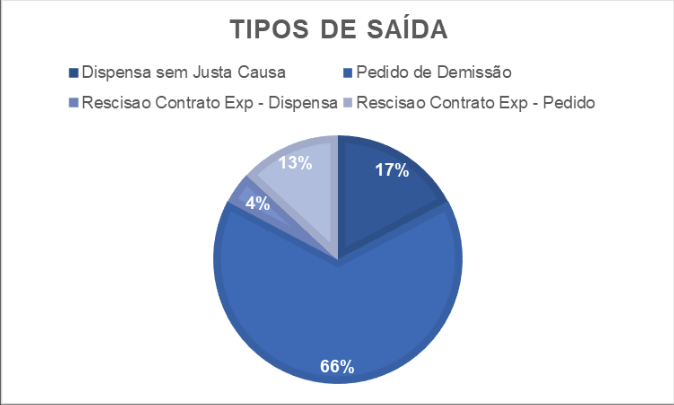
Neste gráfico temos a quantidade de pessoas que deixaram a empresa por cargo.****

1. **Saídas por idade**

Neste gráfico temos a quantidade de pessoas que saíram por idade.

****

1. **Tipos de saída**

Neste gráfico temos 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.

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

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

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

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.

* 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.