



# PROJETO AGATHA FACULDADE DE MEDICINA DA USP

## Controle do Documento

### Histórico de revisões

Data	Autor	Versão	Resumo da atividade
30/01/2023	Yago Araújo	1.1	Criação do Documento e personalização conforme o grupo 3
02/02/2023	Enya Oliveira Arruda	1.2	Preenchimento da seção 1
02/02/2023	Enya Oliveira Arruda	1.3	Preenchimento da seção 2
02/02/2023	Yago Araújo	1.4	Preenchimento da seção 4.1.4
03/02/2023	Luis Miranda	1.5	Preenchimento da seção 4.1.1
04/02/2023	Marcelo Maia	1.6	Preenchimento da seção 4.1.1.2
05/02/2023	Luis Miranda	1.7	Preenchimento da seção, 4.1.2
05/02/2023	Luis Miranda	1.8	Preenchimento da seção 4.1.5
05/02/2023	Isabela Rocha Marcelo Maia	1.9	Preenchimento da seção 4.1.6
05/02/2023	Enya Oliveira, Thomaz Barboza	1.9.1	Preenchimento da seção 4.1.7
06/02/2023	Enya Oliveira	1.9.2	Preenchimento da seção, 4.1.3
10/02/2023	Fabio Piemonte	2.0	Agrupando e adicionando correções

# Sumário

<b>1. Introdução</b>	<b>4</b>
<b>2. Objetivos e Justificativa</b>	<b>5</b>
2.1. Objetivos	5
2.2. Proposta de Solução	5
2.3. Justificativa	5
<b>3. Metodologia</b>	<b>6</b>
<b>4. Desenvolvimento e Resultados</b>	<b>7</b>
4.1. Compreensão do Problema	7
4.1.1. Contexto da indústria	7
4.1.2. Análise SWOT	7
4.1.3. Planejamento Geral da Solução	7
4.1.4. Value Proposition Canvas	7
4.1.5. Matriz de Riscos	7
4.1.6. Personas	8
4.1.7. Jornadas do Usuário	8
4.2. Compreensão dos Dados	9
4.3. Preparação dos Dados e Modelagem	10
4.4. Comparação de Modelos	11
4.5. Avaliação	12
<b>5. Conclusões e Recomendações</b>	<b>13</b>
<b>6. Referências</b>	<b>14</b>
<b>Anexos</b>	<b>15</b>

# 1. Introdução

A faculdade de medicina da USP é a instituição de ensino médico mais renomada do Brasil, fundada em 1913, é uma das instituições mais antigas do país. Oferece cursos de graduação, pós-graduação, residência médica e especialização, possui um hospital universitário, Hospital das clínicas da FMUSP, que alcançou uma das primeiras posições em relação a qualidade, sendo um centro de atendimento muito importante na cidade de São Paulo e é também um centro de pesquisa para estudantes e médicos.

A FMUSP tem uma forte presença na pesquisa médica e tem colaborado com instituições nacionais e internacionais no avanço do conhecimento médico, o estudo para a eficiência do tratamento do câncer de mama é um tópico muito discutido devido a quantidade de variáveis para que a decisão aconteça e a incerteza do raciocínio humano que não consegue analisar centenas de dados e estatísticas para realizar a escolha.

## 2. Objetivos e Justificativa

### 2.1. Objetivos

A Faculdade de medicina da USP é reconhecida pelo seu papel na área das pesquisas médicas, e no diagnóstico e tratamento de doenças. O tratamento para o câncer de mama é um dos mais procurados na instituição, e devido a ampla variedade de variáveis ao designar o tratamento mais viável para as pacientes, sendo eles coadjuvante ou adjuvante, o objetivo do parceiro é determinar uma ferramenta que auxilie com a prescrição do diagnóstico.

### 2.2. Proposta de Solução

Estamos desenvolvendo um modelo preditivo proposto para definir um tratamento de câncer baseado em dados clínicos e de saúde do paciente, incluindo informações sobre o tipo e estágio do câncer, idade, histórico médico e outros fatores relevantes.

O objetivo é fornecer uma recomendação de tratamento personalizada e baseada em evidências para cada paciente, que leve em consideração todos os fatores relevantes e ajuda a equipe médica a tomar uma decisão informada sobre o melhor curso de ação, existem duas opções de tratamento, a coadjuvante e a neoadjuvante e a ferramenta auxiliará o mastologista e sua equipe na precisão da prescrição.

### 2.3. Justificativa

O câncer é uma doença complexa e pode afetar diferentes pessoas de maneiras diferentes. Um modelo preditivo permite uma abordagem mais personalizada ao tratamento, levando em consideração as características individuais do paciente. O modelo preditivo é alimentado por dados clínicos e de saúde dos pacientes, como por informações sobre as últimas descobertas científicas e as melhores práticas médicas disponíveis para o tratamento do câncer. Isso significa que a recomendação de tratamento é baseada em evidências sólidas e atualizadas. ajudar a minimizar o risco de erro humano na definição do tratamento. Além disso, o modelo permite a integração de uma grande quantidade de dados e informações, o que é difícil de ser feito de forma manual.

### 3. Metodologia

*Descreva as etapas da metodologia CRISP-DM 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.*

## 4. Desenvolvimento e Resultados

### 4.1. Compreensão do Problema

#### 4.1.1. Contexto da indústria

O Hospital das Clínicas de São Paulo é uma instituição de destaque na área da saúde no Brasil. Fundado em 1940, este complexo hospitalar universitário localizado na cidade de São Paulo é considerado um dos principais centros médicos da América Latina, com uma ampla gama de serviços de saúde prestados.

Com uma infraestrutura moderna e tecnológica, o Hospital das Clínicas se destaca como uma instituição de referência em diversas áreas médicas, com foco para cardiologia, oncologia e neurologia. Além de prestar atendimento de qualidade com a saúde pública, o hospital também é responsável pela formação de novos profissionais da saúde, como médicos e enfermeiros.

Sua administração está sob a responsabilidade da Universidade de São Paulo (USP), uma das maiores e mais prestigiadas instituições de ensino superior do Brasil e da América Latina, garantindo assim a qualidade e excelência nos serviços prestados aos pacientes. O hospital também conta com colaborações com diversas empresas e outras instituições de ensino, como o Inteli, o que possibilita a condução de pesquisas e o avanço de novas tecnologias para a área da saúde.

No cenário em geral, a indústria de saúde no Brasil tem experimentado um crescimento constante ao longo dos anos, e a presença do Hospital das Clínicas é fundamental para fortalecer esse setor. Com uma equipe altamente qualificada e dedicada à saúde da população, o hospital é visto como referência por outras instituições e contribui para melhorar a qualidade de vida da população.

Em síntese, a indústria do Hospital das Clínicas é caracterizada pela excelência na prestação de serviços da saúde pública, pela formação de novos profissionais, pela realização de pesquisas e pelo avanço de novas tecnologias. Tudo isso confere ao hospital uma posição de destaque na indústria da saúde do Brasil.

### 4.1.1.2 As 5 forças de porter

#### As 5 forças de porter:

##### **-Rivalidade entre concorrentes:**

Considerando que o ICESP (instituto do câncer do estado de são paulo) é uma instituição estatal e sem fins lucrativos, não possui concorrentes diretos. Contudo, é necessário atentar-se aos métodos e tratamentos oferecidos pelos principais hospitais privados especializados em tratamento de câncer no Brasil, para que assim o instituto mantenha-se atualizado e sendo um centro de tratamento com muita qualidade e seja capaz de oferecer serviços equiparáveis ao setor privado.

##### **-Poder de barganha dos Fornecedores:**

Novamente, por ser uma instituição governamental o capital disponível é limitado de acordo com quanto o Estado irá destinar para a área da saúde. Além disso, tendo em vista que os produtos necessários para a manutenção e avanço do instituto são altamente especializados e de alto valor, assim como são indispensáveis para o funcionamento do ICESP. Dessa forma, os fornecedores possuem um alto poder de barganha, principalmente devido ao seus bens fornecidos serem de um nicho muito específico e de alto valor, mas também por não ser um mercado tão diluído.

Já em relação ao modelo preditivo, o poder de barganha é baixo, sabendo que os fornecedores de dados também são os clientes/parceiros, dessa maneira é de interesse mútuo a disponibilização dos dados.

##### **-Poder de barganha dos compradores:**

Em relação ao ICESP, o poder de barganha é baixo, pois os clientes/pacientes estão utilizando-se de um serviço público, que é prestado de maneira gratuita, logo uma possível barganha de preço não aplica-se a situação.

Por outro lado, considerando o poder de barganha dos clientes em relação ao software desenvolvido é altíssimo, tendo em vista que de acordo com a LGPD (Lei Geral de Proteção de Dados) os clientes podem negar o fornecimento de dados, assim como recusar a qualquer a utilização de seus dados, ainda que a permissão tenha sido dada anteriormente.

##### **-Ameaça de novos entrantes:**

Com o avanço cada vez maior das instituições de pesquisa, o risco da ameaça de novos entrantes é alto, pois a iniciativa privada tem mais capital, além disso o risco do surgimento de instituições privadas com soluções mais adequadas e viáveis é alto e deve ser levado em consideração pelo ICESP.

##### **-Ameaça de produtos substitutos:**

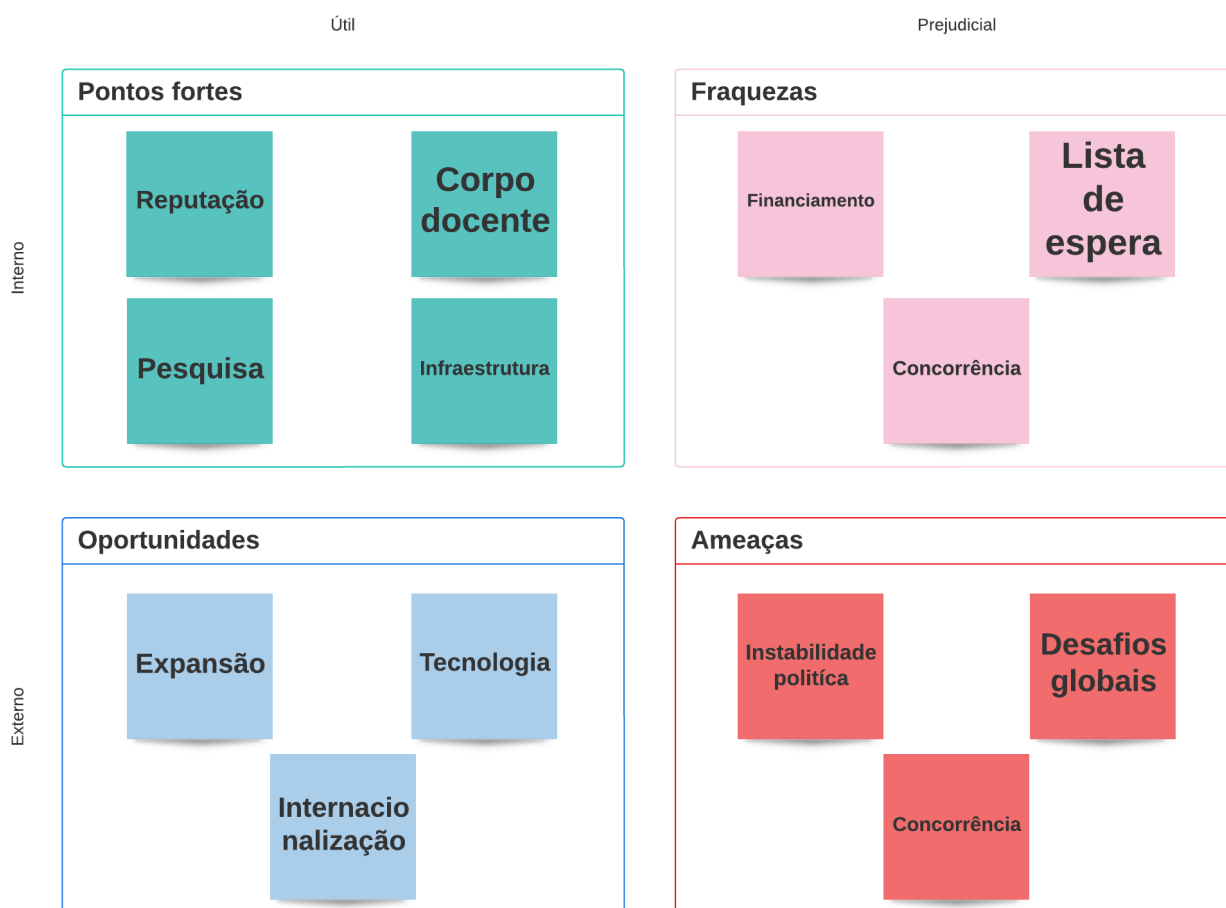
O crescimento exponencial do uso de IA na área da saúde representa um alto risco para o modelo preditivo desenvolvido, tendo em vista que já existem modelos que atualmente são utilizados para a prescrição de tratamento para outras doenças, mas poderiam ser facilmente convertidos para a utilização para o câncer de mama.

Além disso, o desenvolvimento de modificação de genes baseado na sequenciação do genoma está sendo bastante pesquisado e, caso concretizado, não haveria a necessidade do modelo preditivo desenvolvido.



## 4.1.2. Análise SWOT

Na análise SWOT, buscamos definir os pontos fortes e fracos, oportunidades e ameaças do Hospital das Clínicas, com influência de forças externas (algo que o hospital não possui controle) e internas (que podem ser controladas pelo hospital). Em seguida da matriz, disponibilizamos a legenda de cada ponto levantado



**Fonte:** elaboração dos autores

### Pontos Fortes:

- **Reputação:** O Hospital das Clínicas possui um amplo reconhecimento por sua excelência em pesquisa e atendimento médico público, fazendo com que possua pacientes de todo o país.
- **Corpo docente:** O hospital possui uma equipe altamente qualificada que trabalha juntos para oferecer o melhor atendimento aos pacientes.
- **Pesquisa:** O hospital é líder em pesquisa médica, tanto no Brasil quanto no mundo.
- **Infraestrutura:** O hospital possui instalações modernas e tecnológicas, incluindo laboratórios, equipamentos médicos e clínicas de última geração.

### Fraquezas:

- **Financiamento:** Apesar de ter uma equipe altamente qualificada, o hospital pode sofrer com a falta de recursos financeiros para investir em pesquisas, infraestrutura e tecnologia.
- **Lista de espera:** Pelo fato de ter uma equipe altamente qualificada e com serviços gratuitos para a população, o hospital acaba sendo muito procurado, o que pode resultar em listas de espera longas para procedimentos e consultas.
- **Concorrência:** O hospital enfrenta uma intensa concorrência de outros hospitais escola de medicina com infraestrutura de ponta e ensino de qualidade no Brasil e no mundo.

#### Oportunidades:

- **Expansão:** Há potencial para o hospital expandir seu alcance de ensino e serviços por meio de parcerias com outras universidades, instituições e atendimento médico remoto, o que pode melhorar a acessibilidade aos pacientes.
- **Tecnologia:** O hospital pode aproveitar as vantagens de tecnologias novas e emergentes para melhorar o atendimento ao paciente, investindo em inteligência artificial para melhorar a eficiência e a qualidade do atendimento aos pacientes.
- **Internacionalização:** O hospital pode aumentar seu impacto global atraindo estudantes, professores e pesquisadores de todo o mundo.

#### Ameaças:

- **Instabilidade política:** A instabilidade política no Brasil pode afetar o financiamento e a estabilidade do hospital e de seu corpo docente, fazendo com que a qualidade de atendimento caia.
- **Desafios globais de saúde:** O hospital pode enfrentar desafios para responder a crises emergentes de saúde global, como pandemias.
- **Concorrência:** O hospital enfrenta a concorrência de novos hospitais escola de medicina, como o Albert Einstein e a Santa Casa, o que pode dificultar a atração de alunos e a manutenção de sua reputação.

### 4.1.3. Planejamento Geral da Solução

**a) Qual é o problema a ser resolvido?**

A evolução do câncer de mama e sua resposta a tratamentos convencionais é muito variável, portanto, é necessário encontrar padrões e relações entre os tratamentos já realizados em inúmeros pacientes para entender qual o melhor tratamento deve ser indicado para um indivíduo.

**b) Qual a solução proposta (visão de negócios)?**

Um modelo preditivo que encontra padrões e relações entre os dados de inúmeros pacientes durante o tratamento do câncer de mama, levando em consideração dados demográficos, histopatológicos, registros dos tumores, como também peso e altura.

**c) Como a solução proposta deverá ser utilizada?**

A solução poderá ser usada a partir de um simples site onde o médico poderá colocar os dados do paciente para então receber uma recomendação de tratamento da doença levando em consideração a melhor opção para o seu paciente.

IDEIA - Fornecer a solução via API, possibilitando a implementação dela em softwares já usados por laboratórios/médicos/hospitais

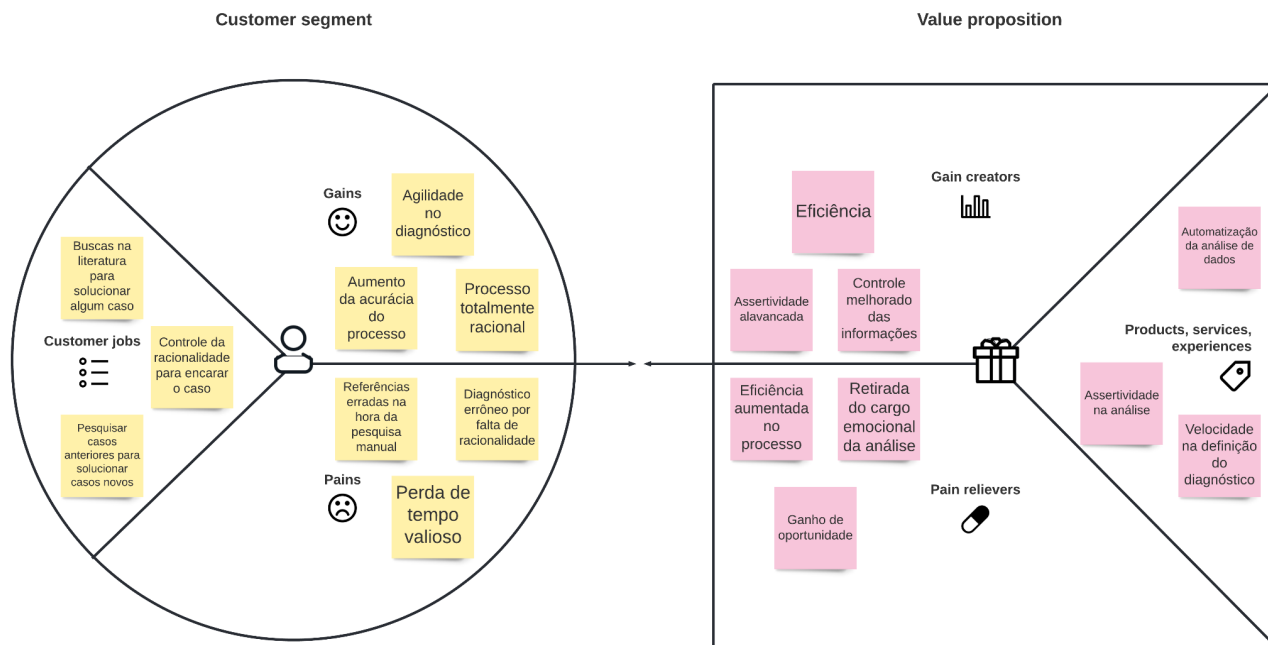
**d) Quais os benefícios trazidos pela solução proposta?**

Auxílio estatístico para a tomada de decisão do tratamento do câncer de mama, levando em consideração os dados do indivíduo e o relacionando com os dados de outros milhares de pacientes resultando na melhor opção a ser tomada.

**e) Qual será o critério de sucesso e qual medida será utilizada para o avaliar?**

O critério de sucesso para o nosso modelo será baseado nas métricas de avaliação para machine learning relacionando a taxa de sobrevida do indivíduo após o tratamento com o tipo de câncer que ele possui.

## 4.1.4. Value Proposition Canvas



Fonte: elaboração dos autores

### Customer segment

#### Customer jobs:

- Buscas na literatura para solucionar algum caso
- Controle da racionalidade para encarar o caso
- Pesquisar casos anteriores para solucionar casos novos

#### Gains:

- Aumento da acurácia do processo
- Processo totalmente racional

#### Pains:

- Referências erradas na hora da pesquisa manual
- Diagnóstico errôneo por falta de racionalidade
- Tempo gasto tentando encontrar o melhor tipo de tratamento para o paciente

### Value proposition

#### Gain creators:

- assertividade alavancada
- Controle melhorado das informações

#### Pain relievers:

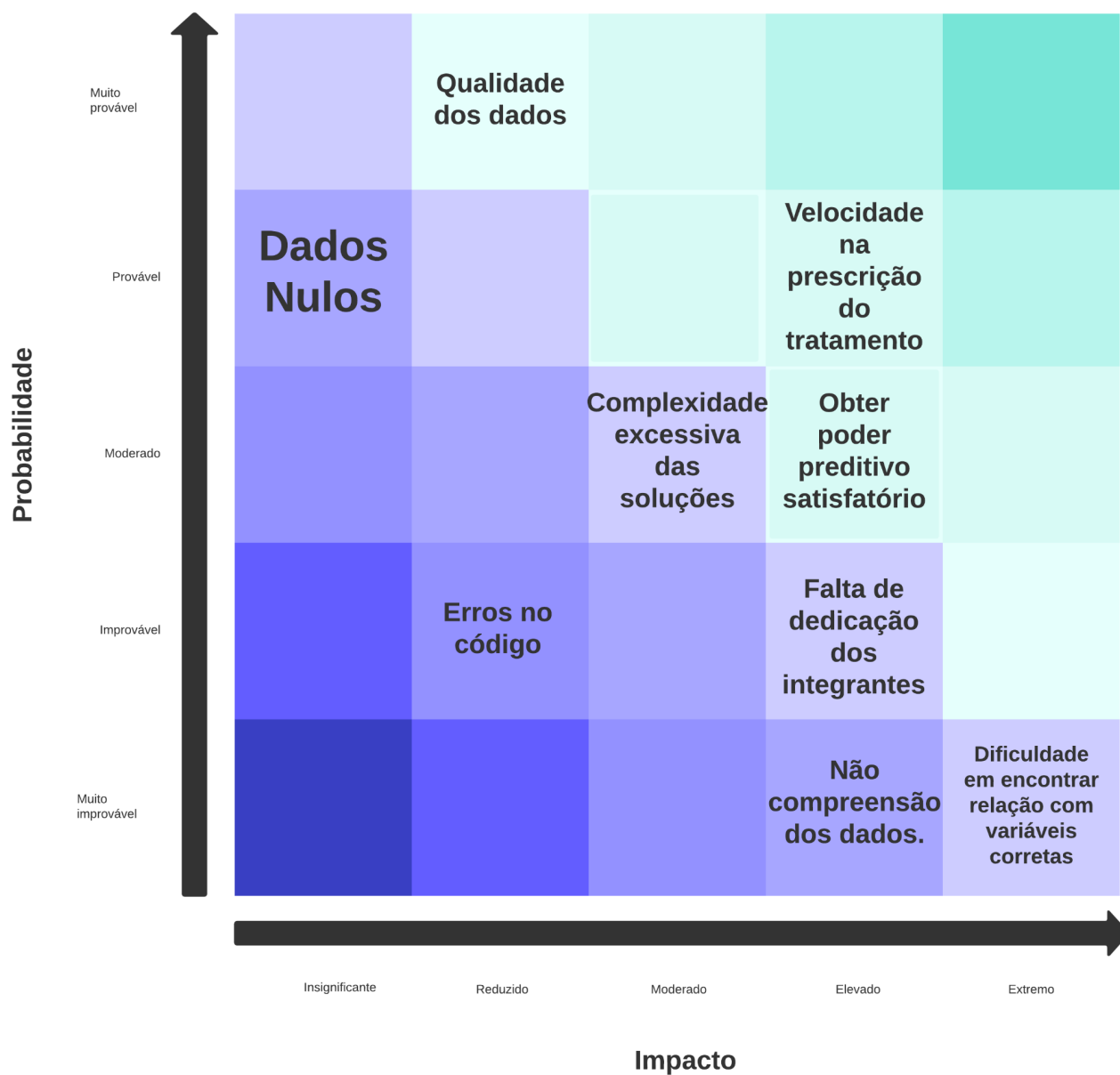
- Eficiência aumentada no processo

- Retirada do cargo emocional da análise

Products, services, experiences:

- Automatização da análise de dados
- Velocidade na definição do diagnóstico

## 4.1.5. Matriz de Riscos




**Fonte:** elaboração dos autores

## 4.1.6. Personas

### ● Persona 1

Médica Oncologista utilizadora do modelo preditivo para prescrição do tratamento.

<p><b>Marcela Arruda Rocha</b></p>  <p>AGE 34 EDUCATION Médica STATUS Casada OCCUPATION Oncologista LOCATION Belo Horizonte - MG TECH LITERATE Ensino Superior Completo</p>	<p><b>Bio</b></p> <p>Ingressou na UFMG aos 20 anos, sempre foi muito dedicada e ocupou as primeiras posições de sua sala. Durante o curso conheceu seu marido, que também é médico, casou-se e teve um filho logo após o término da faculdade. Marcela preza bastante pelo contato interpessoal, sendo assim gosta de atender e acompanhar o tratamento de seus pacientes, logo querendo impactar e transformar a vida de outros indivíduos optou por especializar-se em oncologia, após isso iniciou sua carreira nessa área no Instituto do câncer do estado de São Paulo.</p>	<p><b>Necessidades</b></p> <p>Oferecer uma prescrição mais assertiva para seus pacientes;</p> <p>Plataforma acessível para auxiliar na prescrição de um tratamento;</p> <p>Mais tipos de tratamento disponíveis para seus pacientes no ICESP.</p>
<p><b>Interesses</b></p> <p>Pratica yoga; Gosta de séries de médicos; Leitora voraz de livros sobre filosofia; Ama viajar; Gosta de estudar tecnologia por diversão</p>	<p><b>Motivações:</b></p> <p>Aliar o conhecimento medicinal com a tecnologia;</p> <p>Mais precisão diagnósticos e principalmente na escolha dos tratamentos;</p> <p>Deseja tornar acessível para todos o conhecimento e o tratamento oncológico;</p> <p>Espalhar o conhecimento necessário para que as mulheres em sua maioria sejam capazes de realizar o diagnóstico oncológico precoce.</p>	<p><b>Dores</b></p> <p>Grandes filas para os seus pacientes serem atendidos;</p> <p>Os tratamentos especializados são oferecidos apenas nos grandes centros;</p> <p>Incerteza da precisão de sua prescrição;</p> <p>Sente falta de tempo de qualidade com seus familiares.</p>

**Fonte:** elaboração dos autores.

## • Persona 2

Paciente impactada pelo modelo preditivo na prescrição do tratamento.

<p><b>Maria Miranda Piemonte Barbosa</b></p>  <p>AGE 65</p> <p>EDUCATION Administração</p> <p>STATUS Casada</p> <p>OCCUPATION Administradora</p> <p>LOCATION Recife - PE</p> <p>TECH LITERATE Ensino Superior Completo</p>	<p><b>Bio</b></p> <ul style="list-style-type: none"> <li>Seus pais são cearenses, porém ela nasceu na capital de Pernambuco, passou no vestibular aos 18 anos, quando conheceu seu marido e após dois anos casou-se. Junto com seu marido criou uma loja de produtos regionais. Infelizmente aos 64 anos, foi diagnosticada com câncer de mama estágio 3, devido a isso mudou-se para são paulo com seu marido para a realização do tratamento no ICESP.</li> </ul>
<p><b>Interesses</b></p> <p>Cozinhar comidas típicas do nordeste; Passar tempo com a família; Assistir orquestras; Torcar violino clássico. Viagem anual para a Europa</p>	<p><b>Motivações:</b></p> <ul style="list-style-type: none"> <li>Curar-se do câncer</li> <li>Comemorar a vitória contra a doença viajando para a Europa</li> <li>Tempo de qualidade com a família</li> </ul>
	<p><b>Necessidades</b></p> <ul style="list-style-type: none"> <li>Ter segurança na prescrição de seu tratamento</li> <li>Apoio de seus familiares</li> <li>Tratamento humanizado</li> <li>Hábitos saudáveis</li> <li>Tratamento com baixas chances de recidivas</li> </ul>
	<p><b>Dores</b></p> <ul style="list-style-type: none"> <li>Não realizar suas metas</li> <li>Insegurança com o tratamento</li> <li>Possíveis Efeitos colaterais do tratamento</li> <li>Saudades de sua casa.</li> </ul>

**Fonte:** elaboração dos autores.



## 4.1.7. Jornadas do Usuário

Este mapa fornece a jornada que o usuário terá unicamente com o uso da solução fornecida pela Inteligência Artificial.



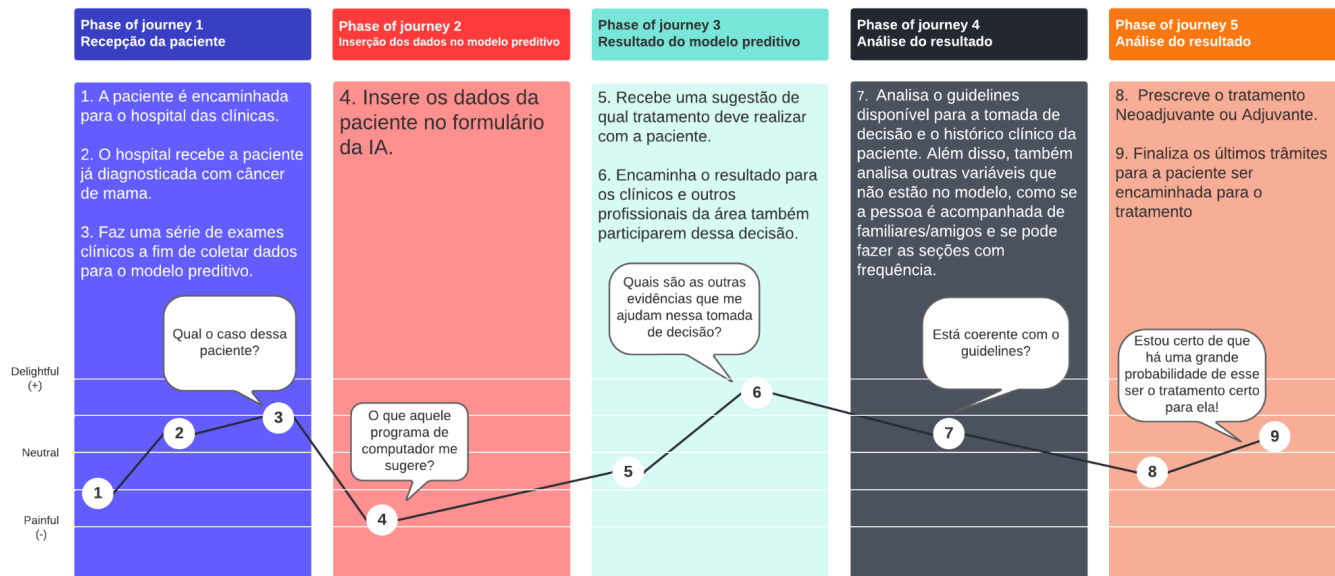
**Marcela Arruda**

### Expectations

- Receber uma resposta objetiva e com alta precisão.
- Design da plataforma intuitiva.
- Receber insights sobre o caso da paciente.

### Scenario description

Dra. Marcela Arruda necessita de ajuda na escolha do tratamento ideal para o câncer de mama de uma paciente. É necessário utilizar um modelo preditivo de alta precisão para auxiliá-la na tomada de decisão.



## 4.2. Compreensão dos Dados

### 1. Exploração de dados:

*Apresentar a estatística descritiva básica de cada coluna, identificação se a coluna é numérica ou categórica e pelo menos 3 gráficos para visualizar a relação entre variáveis escolhidas pelo grupo.*

- a) Cite quais são as colunas numéricas e categóricas.
- b) Estatística descritiva das colunas.

### 2. Pré-processamento dos dados:

*Realizar a limpeza (tratamento de missings e remoção de outliers) e transformação (normalização e codificação) das colunas.*

- a) Cite quais são os outliers e qual correção será aplicada.

### 3. Hipóteses:

*Três hipóteses sobre a relação dos dados e o problema.*

- a) Levantamento das três hipóteses com justificativa.

## 4.3. Preparação dos Dados e Modelagem

Caso seu projeto seja:

1. Modelo supervisionado:

- a) Modelagem para o problema (proposta de features com a explicação completa da linha de raciocínio).
- b) Métricas relacionadas ao modelo (conjunto de testes, pelo menos 3).
- c) Apresentar o primeiro modelo candidato, e uma discussão sobre os resultados deste modelo (discussão sobre as métricas para esse modelo candidato).

Caso seu projeto seja:

1. Modelo não-supervisionado:

- a) Modelagem para o problema (proposta de features com a explicação completa da linha de raciocínio).
- b) Primeiro modelo candidato para o problema.
- c) Justificativa para a definição do K do modelo.
- d) Escolha de um tipo de sistema de recomendação e a justificativa para essa escolha.

## 4.4. Comparação de Modelos

- Escolha da métrica do modelo baseado no que é mais importante para o problema ao se medir a qualidade do modelo;
- Pelo menos três modelos candidatos com tuning de hiperparâmetros e suas respectivas métricas;
- Definição do modelo escolhido e justificativa.

a) Escolha da métrica e justificativa.

b) Modelos otimizados.

- Apresentar três modelos e suas métricas.

- Os modelos apresentados foram otimizados utilizando algum algoritmo de otimização para os hiperparâmetros? Ex. Grid Search e Random Search.

c) Definição do modelo escolhido e justificativa.

## 4.5. Avaliação

*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. Descreva também um plano de contingência para os casos em que o modelo falhar em suas previsões.*

*Além disso, discuta sobre a explicabilidade do modelo e realize a verificação de aceitação ou refutação das hipóteses.*

*Se aplicável, utilize equações, tabelas e gráficos de visualização de dados para melhor ilustrar seus argumentos.*

## 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

*Incluir as principais referências de seu projeto, para que seu parceiro possa consultar caso ele se interessar em aprofundar.*

*Um exemplo de referência de livro:*

*LUCK, Heloisa. **Liderança em gestão escolar**. 4. ed. Petrópolis: Vozes, 2010.*

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

# Anexos

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