



Modelo de Predição Usp Medicina

Controle do Documento

Histórico de revisões

| Data | Autor | Versão (Sprint + Versão) | Resumo da atividade |
|------------|-------------------|--------------------------|--|
| 01/08/2022 | Dayllan Alho | 1.1 | Criação do documento |
| 11/08/2022 | Gabriel Nhoncanse | 1.2 | Inseri os modelos, que já estavam feitos, no documento, além de uma breve introdução de cada tópico. |
| 11/08/2022 | Jordan Andrade | 1.3 | Adição da introdução do documento revisão do documento com algumas alterações ortográficas |
| 12/08/2022 | Jordan Andrade | 1.4 | Revisão de alguns conceitos com base no encontro com o cliente |

Sumário

| | |
|--|-----------|
| 1. Introdução | 4 |
| 2. Objetivos e Justificativa | 5 |
| 2.1. Objetivos | 5 |
| 2.2. Proposta de Solução | 5 |
| 2.3. Justificativa | 5 |
| 3. Metodologia | 6 |
| 3.1. CRISP-DM | 6 |
| 3.2. Ferramentas | 6 |
| 3.3. Principais técnicas empregadas | 6 |
| 4. Desenvolvimento e Resultados | 7 |
| 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 | 10 |
| 4.4. Modelagem | 11 |
| 4.5. Avaliação | 12 |
| 5. Conclusões e Recomendações | 13 |
| 6. Referências | 14 |
| Anexos | 15 |

1. Introdução

O projeto apresentado tem como parceiro de negócios a Faculdade de Medicina da Universidade de São Paulo/Instituto do Câncer do Estado de São Paulo, instituição governamental voltada para atendimento de serviços médicos às comunidades assistidas pelo Sistema Único de Saúde(SUS). O hospital se encontra no endereço Av. Dr. Arnaldo, 251 - Cerqueira César - CEP 01246 903 - São Paulo - SP, e atua no tratamento e profilaxia de patologias humanas e prestação de serviços à comunidade, relacionadas à medicina, fisioterapia, fonoaudiologia e terapia ocupacional, dentro dos mais elevados preceitos éticos e morais. Atualmente é reconhecido como um dos hospitais de referência no combate ao câncer e reconhecido como o melhor hospital público do Brasil pelo World's Best Hospitals 2022.

2. Objetivos e Justificativa

2.1. Objetivos

O parceiro tem como principal objetivo a criação de um sistema preditivo usando inteligência artificial com a finalidade de obter predições sobre a sobrevivência da paciente, para assim conseguir definir o tratamento mais assertivo e menos nocivo com base nos resultados obtidos.

2.2. Proposta de Solução

O problema a ser resolvido pelo time é analisar a dificuldade de tratar o câncer de mama, devido às grandes variações de resultados em relação aos tratamentos convencionais. A partir dos dados disponibilizados podemos observar de forma um pouco mais aprofundada como a USP Medicina organiza e orquestra sua base de dados e o quais os dados indispensáveis para uma análise preditiva.

A solução proposta pelo grupo é um modelo preditivo que a partir da análise de dados clínicos definirá a situação entre normal e incomum. Isto é, será um prognóstico da condição da evolução do câncer de mama, pois a avaliação e veredito final serão ditados pelo profissional da área, que a partir de casos incomuns poderá dar seguimento com os pacientes.

O projeto executará a tarefa de classificação a partir de dados informados pelo médico durante o exame, para que a partir desses dados seja possível prever a evolução da variabilidade do câncer de mama e classificar o risco de vida do paciente. A solução poderá ser utilizada por meio de uma interface onde o profissional da saúde colocará as informações centrais do paciente e a IA analisará a evolução dos dados e do prognóstico.

Os benefícios do modelo de predição são muitos devido ao trabalho da inteligência artificial, como o ganho de tempo dos profissionais da área da saúde, que pode ser destinado ao tratamento de outros novos pacientes. Outro benefício é a precisão do prognóstico, que pode aumentar muito o número de tratamentos e de vidas salvas. Além disso, o encaminhamento de pacientes de acordo com o estágio cancerígeno e a possibilidade de evolução da doença com base na predição pode reduzir a fila de espera para o tratamento, uma vez que no cenário atual todos os pacientes devem comparecer ao hospital na mesma frequência visto que não se sabe qual será a reação do enfermo ao tratamento, o que o ocupa as vagas e sobrecarrega o hospital.

O critério de sucesso será após a conclusão do médico a partir dos dados obtidos, e então, após analisados pela IA, a forma a qual será utilizada para avaliar. Ao concluir a etapa da doença que o paciente se encontra, assim, será obtida a classificação de risco do paciente.

2.3. Justificativa

A proposta de solução é um modelo preditivo que auxilie o médico a escolher o melhor tratamento para a paciente, tendo em vista dados sobre a variabilidade do câncer de mama e sobre o próprio paciente. Um diferencial do nosso modelo é que ele funciona completamente por meio de inteligência artificial, a qual será treinada por meio de dados de casos passados e conseguirá dar uma predição mais precisa.

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

3.2. Ferramentas

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

3.3. Principais técnicas empregadas

Descreva brevemente as principais técnicas empregadas, algoritmos e seus benefícios

4. Desenvolvimento e Resultados

4.1. Compreensão do Problema

4.1.1. Contexto da indústria

Nesse mesmo setor, o Instituto do Câncer de São Paulo não possui necessariamente concorrentes, pois não disputa mercado com outros players, porém, existem outras empresas que trabalham no mesmo setor, como por exemplo: AC Camargo, hospital oncológico especializado no diagnóstico, tratamento e pesquisa de câncer em humanos, Instituto Nacional do Câncer, órgão auxiliar do Ministério da Saúde no desenvolvimento e coordenação das ações integradas para a prevenção e o controle do câncer no Brasil, e a FEMAMA, associação civil, sem fins econômicos, que busca ampliar o acesso ao diagnóstico e tratamento do câncer de mama para todas as pacientes, buscando reduzir os índices de mortalidade pela doença no Brasil.

Além disso, hoje o hospital se mantém ativo através de recursos públicos vindos majoritariamente do Ministério da Saúde, porém, podem vir de outras fontes governamentais, assim oferecendo um serviço gratuito aos pacientes.

Tendo em vista o mercado que o hospital está inserido, conclui-se que ele está constantemente buscando se aperfeiçoar e inovar, contando com investimentos públicos e privados à procura de inovações, como equipamentos novos, tratamentos mais eficazes e menos nocivos etc.

Em relação às **5 forças de porter**, concluímos que:

Rivalidade entre os concorrentes: Por ser uma instituição governamental sem fins lucrativos, o Instituto do Câncer de São Paulo não possui concorrentes de fato e sim hospitais parceiros em pesquisa e tratamento do câncer. Todavia, existem outras instituições como o Instituto Nacional do Câncer e diversos hospitais privados que realizam o tratamento de câncer e "disputam" pacientes.

Poder de negociação dos fornecedores: Por necessitar de produtos extremamente refinados e de alto valor agregado, o Instituto do Câncer de São Paulo possui fornecedores com alto poder de negociação, visto que certos medicamentos e aparelhos de pesquisa não possuem grande oferta no mercado e possuem sua produção e precificação controladas por um pequeno grupo de empresas.

Ameaça de entrada de novos concorrentes: Não há ameaça de entrada de novos concorrentes, pois as instituições da mesma área de trabalho da USP medicina não se passam

por concorrentes, e sim por parceiros. Além disso, a USP é ponto de referência na área de saúde e qualquer empresa que surgisse não conseguiria a curto prazo ser um concorrente à altura.

Ameaça de produtos substitutos: Pelo tratamento oncológico ser uma área que ainda possui muitos mistérios para a medicina e não possuir uma exatidão no tratamento e evolução dessa patologia, os riscos de surgir algum produto que substitua os métodos de tratamento convencional do câncer são praticamente nulos a curto prazo.

Poder de negociação dos clientes: Por ser um órgão público que atua predominantemente com pessoas de baixo poder aquisitivo, os clientes do Instituto do Câncer de São Paulo possuem baixo poder de negociação, uma vez que depende do Sistema Único de Saúde (SUS) para realizar o tratamento e não possuem condições de arcar com os custos de um hospital privado

4.1.2. Análise SWOT

A meta da análise SWOT é facilitar na identificação de características da empresa parceira (USP Medcina) e do mercado em que ela se encontra, assim nos ajudando no desenvolvimento do projeto. Além disso, ela facilita a potencialização de suas forças, mitigação de suas fraquezas e minimização de erros, procurando oportunidades para melhorar seus produtos ou elaborar novos protótipos. Diante disso, foi montada uma análise SWOT com base nas características do parceiro de negócios, que pode ser visualizada na imagem abaixo:



Imagem 1: Análise SWOT da USP Medicina

4.1.3. Planejamento Geral da Solução

- a) quais os dados disponíveis (fonte e conteúdo - exemplo: dados da área de Compras da empresa descrevendo seus fornecedores)
- b) qual a solução proposta (pode ser um resumo do texto da seção 2.2)
- c) qual o tipo de tarefa (regressão ou classificação)
- d) como a solução proposta deverá ser utilizada
- e) quais os benefícios trazidos pela solução proposta
- f) qual será o critério de sucesso e qual medida será utilizada para o avaliar

4.1.4. Value Proposition Canvas

É uma ferramenta desenvolvida com a meta de explorar mais profundamente o cliente e a relação dele com o nosso produto por meio de uma análise das suas dores e como o nosso software irá saná-las. Diante disso, foi elaborado um modelo de Value Proposition Canvas com base nas dores do parceiro de negócio e a solução pensada pelo time de desenvolvimento

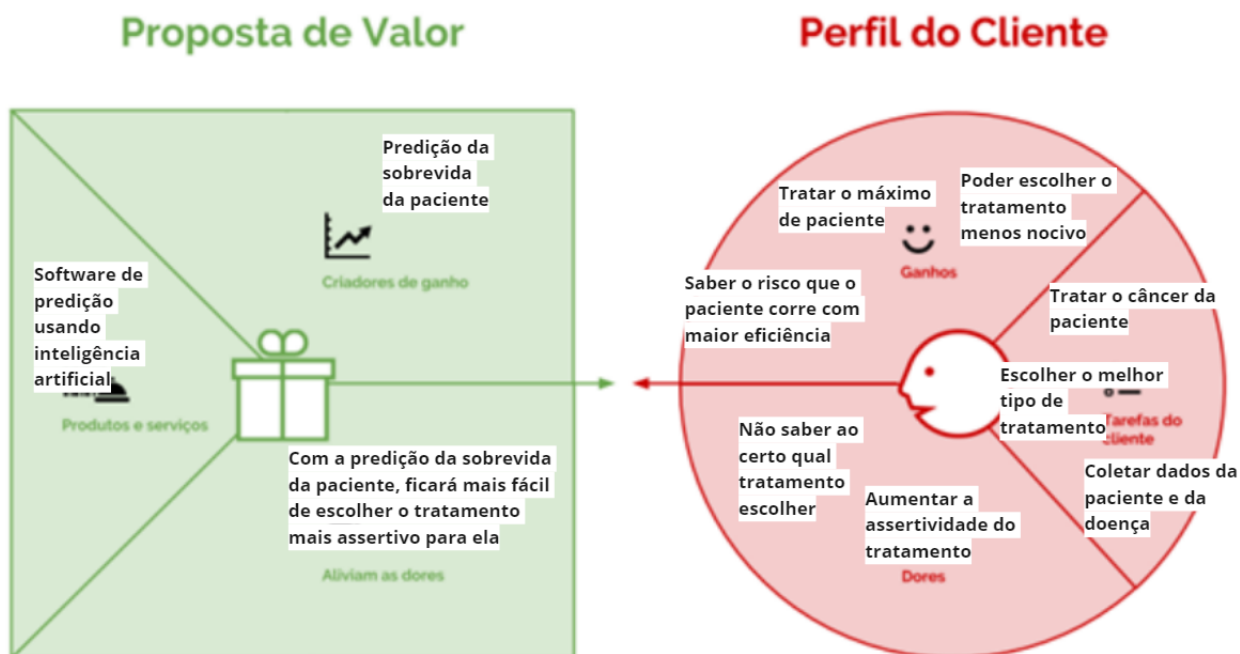


Imagem 2: Value Proposition Canvas

4.1.5. Matriz de Riscos

Também chamada de matriz de probabilidade e impacto, a matriz mapeia os riscos do projeto, sejam eles tanto riscos de ameaças quanto de oportunidades. Por ser uma ferramenta útil para gerenciar os riscos operacionais existentes em um projeto, foi elaborado uma Matriz de Riscos com base na proposta de solução elaborada pelo time de desenvolvimento, que pode ser visualizada na imagem a seguir:

| | | Ameaças | | | | | Oportunidades | | | | |
|---------------|-----|-------------|--------------------------------|--|--|------------------------------|----------------------------------|---|----------|-------|-------------|
| Probabilidade | 90% | | | | | | Análise mais assertiva dos dados | Ampliação do modelo de predição para outros tipos de análises | | | |
| | 70% | | | | Interface não intuitiva para quem for analisar | Erros ao trabalhar com dados | Redução do tempo de diagnóstico | Crescimento tecnológico | | | |
| | 50% | | | Inexperiência com modelos de predição com a Inteli | Dados faltosos ou com poucas referências | | | | | | |
| | 30% | | Incompreensão da base de dados | | | | | | | | |
| | 10% | | | | | | | | | | |
| | | Muito Baixo | Baixo | Moderado | Alto | Muito Alto | Muito Alto | Alto | Moderado | Baixo | Muito Baixo |
| Impacto | | | | | | | | | | | |

Imagem 3: Matriz de riscos

4.1.6. Personas

Persona 1 - Médico



Nome: James Wilson da Silva

Idade: 37 anos

Gênero: Masculino

Ocupação: Oncologista

“Um médico nunca desiste de seus pacientes”

Personalidade: Boa comunicação, apesar de ser introvertido; Gosta de passar seu tempo livre com a família; É uma pessoa persistente e está disposto a tudo para salvar seus pacientes.

Biografia: Pai de 3 filhos, James é chefe de Oncologia do Hospital Universitário da USP e lidera pesquisas sobre o câncer de mama. Durante sua adolescência, sua mãe foi diagnosticada com câncer de mama e, infelizmente, não sobreviveu. Esse ocorrido o motivou para se tornar um médico especializado em câncer de mama.

Motivações para usar o novo sistema: A perda de sua mãe pelo câncer de mama, fazendo com que ele procure um tratamento mais assertivo e eficaz.

Dores com o atual sistema:

- 1- Não saber ao certo qual tratamento seria melhor para determinada paciente;
- 2- Tratamentos às vezes mais agressivos do que o necessário.

Objetivos com o novo sistema: Garantir um prognóstico preciso e menos agressivo possível aos pacientes;

Persona 1 - Paciente



Nome: Lisa Andressa dos Santos

Idade: 45 anos

Gênero: Feminino

Ocupação: Professora do Fundamental

“Se há um impossível, meu criador não o conhece”

Personalidade: Muito inteligente e paciente; Sonhadora e esperançosa; Mulher de fé; Mãe de 2 filhos; Adora crianças.

Biografia: Nascida e criada na Zona Leste de São Paulo, Lisa sempre foi apaixonada pela educação, paixão essa que a motivou a se tornar professora e, por meio do ensino e do afeto com seus alunos, criar uma sociedade melhor alterando as bases de ensino. Infelizmente, foi diagnosticada com câncer de mama há cerca de 2 anos, o que a levou a ser afastada do cargo de professora. Entretanto, Lisa é uma mulher sonhadora e cheia de fé, e acredita na progressão positiva do tratamento para poder voltar a dar aulas e ver seus filhos e alunos crescerem e terem um futuro brilhante.

Motivações para usar o novo sistema: Ver o crescimento dos filhos; Voltar a dar aula; Busca por um tratamento eficaz.

Objetivos com o novo sistema: Ser curada; Ter um retorno rápido sobre a situação da evolução do câncer; Conseguir um tratamento menos agressivo.

4.1.7. Jornadas do Usuário

Posicione aqui seus mapas de jornadas do usuário que utiliza o modelo

4.2. Compreensão dos Dados

Tendo como princípio os dados fornecidos pelo Instituto do Câncer de São Paulo, coletados de prontuários eletrônicos de pacientes diagnosticados com câncer de mama em diferentes estágios, foram disponibilizados no formato csv e xml.

Com o recebimento de dados, foi realizado uma triagem da relevância das informações, com um aprofundamento no conhecimento das informações relativas ao câncer, e a posteriori levantadas possibilidades de teorias que a predição poderia realizar a partir de novas informações.

Ademais, com uma análise da base disponibilizada, foi possível observar uma certa instabilidade na base, ausência de dados relevantes e a confusão na identificação dos itens das colunas, colaboram em ampliar a dificuldade da predição.

- Última informação do paciente: status se está vivo ou morto, para avaliar a perspectiva de vida conforme determinados parâmetros;
- consumo de álcool: procurar relação com o progresso do câncer;
- consumo de tabaco: procurar relação com o progresso do câncer
- Possui histórico familiar: procurar relação hereditária do câncer
- Recidiva: verificar casos em que houve reincidência do câncer para determinado tratamento.
- Amamentou na primeira gestação? Validar a correlação da existência ou não do câncer pós amamentação.
- Por quanto tempo amamentou: Validar a correlação do câncer com o período de amamentação.

Idade da primeira menstruação: avaliar relações do câncer e períodos destrutivos. Os parâmetros foram obtidos a partir do arquivo csv "BDIPMamaV11 - DATA LABELS".

Sendo os dados usados para fins estudantis do inteli, foi mapeado algumas colunas definidas como principais para obtenção de predições que trazem clareza para as perspectivas do médico e do paciente.

Portanto, utilizando de todos os artefatos de dados científicos, o algoritmo será direcionado a buscar um melhor mapeamento do câncer e seu progresso, conforme os parâmetros estabelecidos para calcular uma perspectiva de vida ou tratamento.

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.

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:

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.