



NOME DO PROJETO
Nome do Parceiro

Controle do Documento

Histórico de revisões

Data	Autor	Versão	Resumo da atividade
01/08/2022	Sophia Mello Dias	1.1	Criação do Documento
02/08/2022	Rafael Alves Cabral	1.2	Preenchimento da seção 2
02/08/2022	Yves Levi Paixão Lapa	1.3	Contexto da Indústria e 5 Forças de Potter
02/08/2002	Sophia Mello Dias	1.4	Preenchimento da introdução

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 stakeholder Banco Pan atua na área de mercado financeiro com destaque para as áreas de cartões de crédito, crédito consignado, financiamento de veículos, investimentos de renda fixa e banco digital. O banco tem foco nas classes C, D e E da população e conta com mais de 17 milhões de clientes.

Atualmente o Banco Pan possui índices que devem ser melhorados acerca da relação com o cliente, pois está em 3º lugar no índice BACEN do Ranking de Reclamações e conta com diversas reclamações no Procon. Ademais, o banco após se digitalizar aumentou sua cartela de produtos à oferecer e busca atingir um maior número de usuários.

2. Objetivos e Justificativa

2.1. Objetivos

O Banco Pan tem a necessidade de melhorar a sua reputação com o mercado e para isso ele precisa de uma solução que classifique melhor os clientes e não clientes do banco junto da sua necessidade de atendimento.

Sua nota no Reclame Aqui é de 7,5 e ele é o terceiro pior banco segundo o BaCen, sendo assim, o banco precisa classificar os seus clientes entre atritado, engajado e novo cliente para dar o procedimento correto ao atendimento.

2.2. Proposta de Solução

O grupo PanDevs se propõe a desenvolver um software com o uso de Aprendizado de Máquina e Inteligência Artificial que seja capaz de analisar o banco de dados de clientes, segmentar grupos (atritado, engajado e novo cliente) e identificar qual a necessidade do atendimento de cada cliente ou não cliente e com isso retornar ao Banco Pan (...perguntas sexta)

2.3. Justificativa

A computação tem algumas vantagens na hora de analisar dados em grandes escalas, como é o caso do nosso cliente, pois ela consegue processar as informações de forma sistemática em grande escala em uma velocidade que nenhum ser humano conseguiria replicar. Além disso, o algoritmo consegue armazenar e comparar uma quantidade simultânea de dados que garante uma maior confiabilidade dos resultados obtidos.

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 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

Em uma perspectiva macroeconômica, é importante frisar que o país passa por um momento de retomada das atividades habituais e trabalhos presenciais após um período em que foram concedidos diversos auxílios. Portanto, em uma análise top-down é possível de se observar o constante aumento da política de retração da oferta monetária no mercado, ou seja, com o aumento da taxa de juros – definida pelo Bacen – que chega a 13,75% ao ano, o dinheiro acaba sendo mais custoso, afetando empréstimos, tanto para novas empresas que querem captação para a abertura e expansão de novos negócios, mas também de pessoas físicas, podendo aumentar o risco de inadimplência. Porém, com a retomada da economia, o desemprego no Brasil já recua para 9,3%, sendo a menor taxa para o segundo trimestre desde 2015, de acordo com o jornal Folha de São Paulo. Apontando que, apesar do aumento do custo do dinheiro, as pessoas, principalmente de classes sociais mais baixas – C, D, E – conseguem captar algum dinheiro mediante a volta dos seus empregos.

Tendo em vista os drivers do setor bancário, estes são, em resumo: as taxas de desemprego, o ciclo de crédito e o custo do dinheiro mediante a inflação. Desse modo, a indústria apresenta como principais players o Itaú, o BTG Pactual, o Santander, Bradesco, Caixa Econômica Federal, Nubank, Inter, sendo os primeiros mais voltados ao nicho de alto varejo, enquanto as novas fintechs foram no crédito barato para as classes mais baixas financeiramente.

Sob uma perspectiva administrativa desenvolvida por Michael Porter, tal setor apresenta, contemporaneamente, baixíssimas barreiras de entrada, dado o desenvolvimento das tecnologias e a facilidade de se abrir uma fintech atualmente, pois um dos fatores que confere desenvolvimento ao banco é justamente o seu menor custo de captação com depósitos varejistas, maior receita com juros, facilidade de acesso no meio digital, e, ainda, a presença de agências físicas, ainda mais no ambiente C, D, E em que atua o Banco Pan, o que aumenta a ameaça de produtos substitutos. Assim, aumenta-se a rivalidade com concorrentes que acabam focando em uma estratégia de isenção de taxas, nichamento em poucos produtos financeiros que operam bem e melhor experiência do usuário.

Sob a perspectiva do poder de barganha com os clientes, tende-se a observar um alto custo de aquisição do cliente e fraca aderência ao produto, tendo em vista a diversidade de fintechs no mercado com os mesmos atrativos e mesmas funções. Ao passo que, numa relação de poder de compra com fornecedores, o setor bancário - e o Banco Pan em si - estaria mais focado em negociar com credores o financiamento mais barato para a sua expansão e aquisição de mercado, mesmo com os sucessivos aumentos das taxas de juros, o que dificulta esse movimento. Quanto ao modelo de negócios do Banco Pan, assim como dos seus principais concorrentes, estes focam em desburocratizar os serviços financeiros e facilitar a vida do consumidor por meio da tecnologia por meio de um melhor atendimento, isto é, personalizado, e inovar nos produtos fornecidos. Por fim, dentro do modelo de negócios bancário atual é evidente o uso de modelos preditivos, para definirem o tipo do cliente, o seu perfil de investimento, o local onde estão mais propensos a serem adquiridos, entre outros fatores em que a inteligência artificial consegue ser atribuída.

4.1.2. Análise SWOT

Pontos Fortes <ul style="list-style-type: none"> • Banco acessível • Crédito para pessoas baixa renda • App do banco • Diversidade de serviços para classes C, D e E 	Pontos Fracos <ul style="list-style-type: none"> • Baixa nota no Bacen • Muitas reclamações no procon • Atendimento insatisfatório
Oportunidades <ul style="list-style-type: none"> • Atrair mais clientes baixa renda • Interação nas redes sociais • Open Banking 	Ameaças <ul style="list-style-type: none"> • Inadimplência • Má reputação • Concorrência de novas fintechs

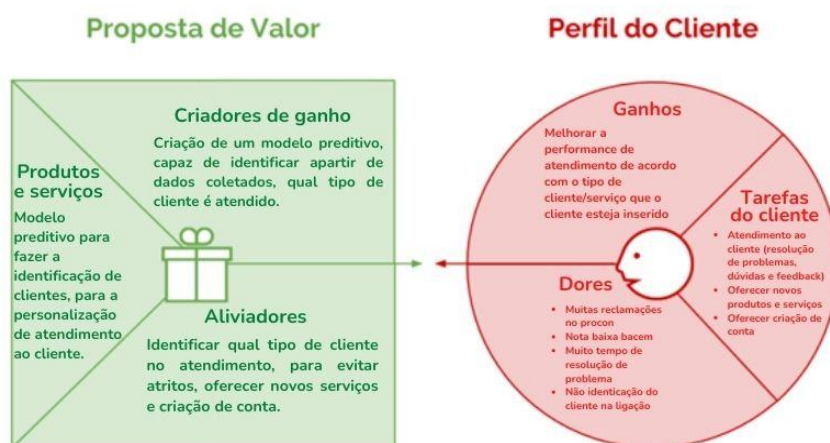
4.1.3. Planejamento Geral da Solução

- 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



4.1.5. Matriz de Riscos

Matriz de Risco						
Probabilidade	Ameaça			Oportunidade		
Alta	Conhecimento limitado do assunto			Automatização no atendimento	Numero alto de usuários	
Médio		Código oferecer direcionamento errado para o atendimento	Dataset incompleto	Fidelização do usuário	Muitas reclamações / clientes atritados	
Baixa	Faltar com ética no código	Vazamento de dados	Não concluir o projeto	Índice Bacem piorar		
	Baixo	Médio	Alta	Alta	Médio	Baixo
	Impacto					

4.1.6. Personas

a) Afetadas pelo modelo:

Sophia

Sophia quer resolver seus problemas com o banco.



Descrição

Sophia tem 23 anos e abriu a sua conta há 2 anos, ela ainda mora com os seus pais e sempre gostou da sua experiência com o Pan, mas nos últimos 3 meses ela está com dificuldade em mudar a sua senha do cartão.

Maiores objetivos

- Conseguir trocar a senha do seu cartão de crédito
- Comprar o seu carro
- Estabilidade financeira
- Ser efetivada no seu estágio

Personalidade

- Gosta de Funk
- Muito paciente
- Adora conhecer bares novos

Interesses

- Academia
- Livros
- Podcasts
- TikTok

Apps usados pela Sophia

- Banco Pan
- Uber
- Instagram
- WhatsApp

Ganhos

- Facilidade de atendimento via app

Dores

- Muito tempo gasto para solucionar problemas

Sophia tem 23 anos e sua primeira conta em uma instituição financeira foi com o Banco Pan. Infelizmente ela está com problemas em mudar a sua senha do cartão, as ligações demoram muito e não resolvem seu problema, que já persiste a 3 meses.

Ivone

Ivone quer adquirir produtos com o banco.



Brief description

Ivone tem 63 anos, está trabalhando na área de engenharia civil e em processo de aposentadoria e precisa de um banco para guardar sua renda de forma segura.

Maiores objetivos

- Ivone quer se aposentar e se mudar para o interior.

Personality

- Introvertida
- Gosta de gatos

Interests

- Tecnologia
- Conhecimento
- Natureza
- Animais
- Cozinhar

Apps used by Ivone

- Petlove
- Petz
- Facebook
- Pinterest

Ganhos

- Conta rende mais que a poupança
- Segurança contra roubos na rua
- Confiabilidade no nome do banco


Dores

- Insegurança em investir o seu dinheiro
- Precisa de um banco de forma rápida
- Não tem muito tempo para sair de casa e ir resolver as coisas do banco

Ivone está em processo de aposentadoria e já é cliente do Banco Pan, por isso deseja contratar novos produtos que condizem com sua atual condição financeira e de vida, atendendo de forma fácil suas necessidades.

Lucca

Lucca é um possível novo cliente.



Brief description

Lucca é um jovem entusiasmado com a tendência atual do mundo tech, então ele quer abrir uma conta num banco digital que supra suas principais necessidades

Maiores objetivos

- Lucca tem como principal objetivo se tornar CEO de uma empresa.

Personality

- Geek
- Extrovertido
- Curioso

Interests

- Tecnologia
- Festas
- Sair com os amigos
- Redes sociais

Apps used by Lucca

- Discord
- Instagram
- Twitter
- Reddit

Ganhos

- Facilidade em encontrar suas soluções bancárias
- Rapidez em abertura de conta

Dores

- Não tem cartão de crédito ainda
- Foi negado nos demais bancos

Lucca já tem contas em outros bancos, mas não está satisfeito com nenhum deles. Como é jovem, quer um banco que esteja conectado com o mundo atual, seja tecnológico e tenha tudo o que precisa na palma de sua mão.

4.1.7. Jornadas do Usuário

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

4.2. Compreensão dos Dados

1. Conforme enviado pelo stakeholder (Banco Pan), os dados utilizados são no formato CSV, com mais de 12 milhões de linhas (1.424 GB) contendo informações sobre: Valor de Crédito, Valor de Saldo, Número de Atendimentos Atrasados, Valor de Score, Número de Produtos, Número de Produtos, Número de Atendimentos, Número de CPF, Quantidade de Operações, Quantidade de Reclamações, Quantidade de Restritivos, Valor de Renda e Cod Rating. Desses dados primeiramente fornecidos, os dados referentes ao Valor de Crédito, Valor de Score, Valor de Renda e Cod Rating referem-se ao mercado, e não restritivamente ao Banco Pan.
 - a. O Banco Pan forneceu apenas um conjunto de dados.
 - b. Apesar de possuímos uma quantidade considerável de dados para serem analisados e contribuir para a inteligência a ser criada (mais de 12 milhões de linhas), algumas das colunas possuem um número extremamente elevado de linhas com valor nulo. São elas, o número de atendimentos atrasados ("num_atend_atrs"), o número de atendimentos totais ("num_atend") e a quantidade de reclamações ("qtd_reclm"). Precisamos alinhar com o cliente se essas linhas com valor nulo podem ser consideradas pela nossa inteligência com um valor padrão (exemplo: "0"), ou realmente devem ser completamente desconsideradas. Também serão necessários dados ainda não providenciados pelo nosso cliente e são de extrema importância para o modelo a ser produzido posteriormente.
 - c. Como dito anteriormente, precisamos alinhar com o nosso cliente se os dados que estão como nulos podem ser considerados pela nossa inteligência com um valor default. Nesse momento inicial (11/08/2022), estamos analisando a estruturação da base de dados que recebemos e tentando realmente compreender como poderemos utilizar os valores de cada coluna para a criação do nosso sistema de predição. Além disso, estamos definindo quais outras informações poderiam ser utilizadas para o nosso modelo de inteligência para que possamos fazer uma requisição de novos dados ao nosso cliente (Banco Pan).
 - d. Por se tratar dos dados de um banco, é necessário que nós tenhamos um cuidado ainda mais elevado para manter a base de dados o mais privada possível. Dito isso, não podemos, de forma alguma, copiar essas informações para alguma forma de armazenamento que seja externa ao ambiente do Inteli.

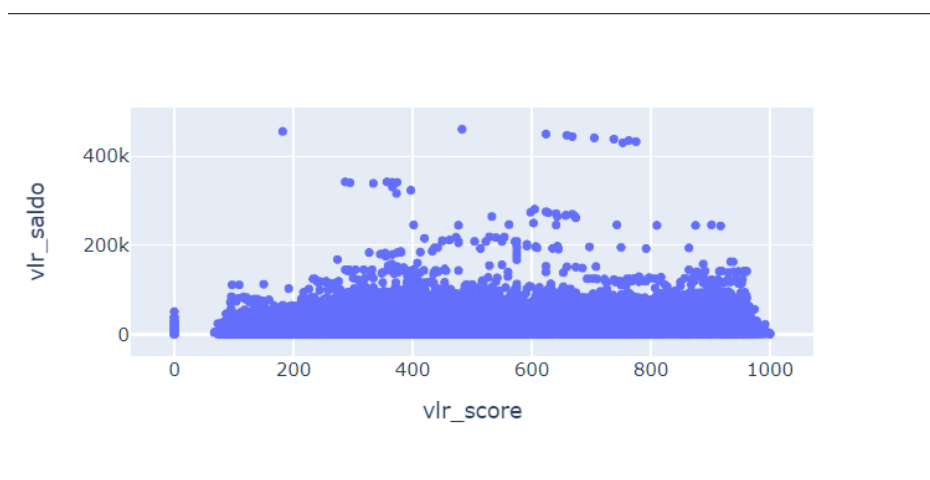
2. Abaixo há a descrição estatística básica dos dados, priorizando os atributos de interesse. Segue uma imagem de uma tabela contendo a descrição estatística básica de cada coluna em nossa base de dados:

	anomes	vlr_credito	vlr_saldo	num_atend_atrs	vlr_score	num_produtos	num_atend	qtd_oper	qtd_reclm	qtd_restr
count	1.250529e+07	7.032474e+06	6.600003e+06	6604.000000	8.738902e+06	6.688796e+06	26545.000000	7.032474e+06	1364.000000	8.330173e+06
mean	2.021353e+05	3.206503e+04	5.864501e+03	1.117656	4.667822e+02	1.619372e+00	1.373517	1.198031e+01	1.000733	2.857061e+00
std	4.285810e+01	6.567294e+04	2.855815e+04	0.389080	2.074592e+02	9.849524e-01	0.697494	1.027429e+01	0.027077	3.561156e+00
min	2.021040e+05	0.000000e+00	1.000000e-02	1.000000	0.000000e+00	1.000000e+00	1.000000	0.000000e+00	1.000000	1.000000e+00
25%	2.021070e+05	2.974323e+03	9.946900e+02	1.000000	3.290000e+02	1.000000e+00	1.000000	5.000000e+00	1.000000	1.000000e+00
50%	2.021100e+05	1.424500e+04	2.358240e+03	1.000000	4.290000e+02	1.000000e+00	1.000000	1.000000e+01	1.000000	2.000000e+00
75%	2.022010e+05	3.395993e+04	6.748950e+03	1.000000	5.800000e+02	2.000000e+00	2.000000	1.600000e+01	1.000000	3.000000e+00
max	2.022040e+05	1.034811e+07	3.210277e+07	7.000000	1.000000e+03	1.500000e+01	17.000000	3.060000e+02	2.000000	4.130000e+02

a.

b. Hipótese 1:

- i. Clientes com o score mais alto possuem um valor de crédito no banco pan maior.



ii.

iii. Hipótese não comprovada.

c. Hipótese 2:

i. Banco pan possui mais clientes com classificação A do que H.

```
df_para_plot.cod_rating.describe()
```

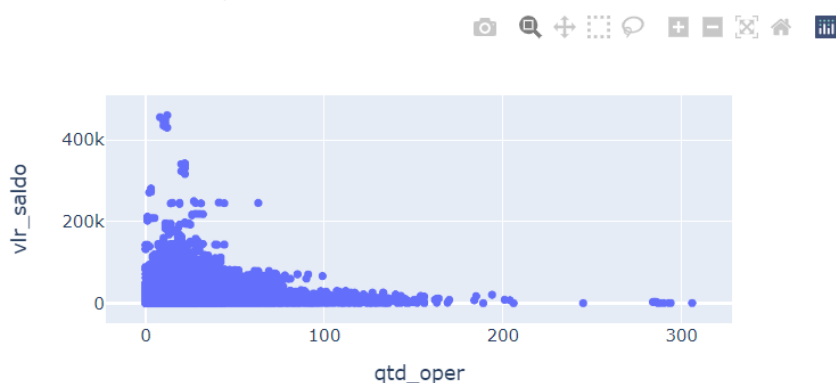
count	6600003
unique	10
top	A
freq	5088485
Name: cod_rating, dtype: object	

ii.

iii. Comprovada.

d. Hipótese 3:

i. Como o banco atende a clientes de baixa renda, acreditamos que eles contratem vários produtos com baixos valores.



ii.

iii. Comprovada.

3. O modelo de predição tem como objetivo("target"), determinar se o cliente que está entrando em contato com o banco está pendendo para **Atritado**, **Engajado** ou **Novo cliente**. A natureza dessa predição será discreta, pois os dados que utilizaremos para o modelo também serão discretos. Além disso, utilizar a predição discreta se encaixa melhor na solução implementada, pois a predição irá categorizar o cliente.

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.