



Predição com aprendizagem supervisionada

Banco Pan

Histórico de revisões

[illegible]

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

Nosso parceiro de negócios é o Banco PAN, que tem sua sede em São Paulo (SP) e atua ajudando pessoas das classes C, D e E, transformando desafios em conquistas. Quando a questão é seu porte e posicionamento no mercado, o Banco PAN ultrapassou 15 milhões de clientes no 3º trimestre de 2021.

A problemática envolve o atendimento deles, que, atualmente, não é personalizado para os possíveis propósitos do cliente e do banco. Ele é feito sem muito preparo dos atendentes quanto à situação atual de seus clientes: se estão atritados com o banco, buscam novos produtos ou são potenciais clientes novos. Este processo é um dos responsáveis por um grande número de reclamações, as quais fizeram com que a instituição figurasse como a segunda maior no ranking de reclamações de bancos do Brasil.

2. Objetivos e Justificativa

2.1. Objetivos

Descreva resumidamente os objetivos gerais e específicos do seu parceiro de negócios

2.2. Proposta de Solução

A partir dos objetivos do nosso parceiro de negócios, a nossa solução visa mitigar as dificuldades enfrentadas pela empresa quando o assunto é o relacionamento com clientes do banco. Assim, criaremos uma predição com aprendizagem supervisionada capaz de classificar o status do cliente a partir dos dados oferecidos pelo banco, fazendo com que ele se enquadre em três principais grupos: aqueles que têm potenciais atritos com o banco, aqueles que são novos clientes ou os que estão dispostos a adquirir novos produtos. O foco da nossa implementação deve ser atender melhor os clientes com atritos, a fim de melhorar essa relação entre os indivíduos e a empresa e facilitar o trabalho dos atendentes que precisam lidar com essas pessoas.

2.3. Justificativa

Faça uma breve defesa de sua proposta de solução, escreva sobre seus potenciais, seus benefícios e como ela se diferencia.

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

Análises preditivas são de fundamental importância para as instituições financeiras e possuem uma série de aplicações em Business Intelligence. No contexto em que se insere o nosso cliente — o Banco PAN — um modelo preditivo capaz de tornar mais eficiente o processo de atendimento ao cliente seria de grande utilidade para que o relacionamento entre os usuários e o banco fosse mais satisfatório.

Com a maior adesão da inteligência artificial em ferramentas de predição para grandes empresas nos últimos anos, notou-se um ganho de valor significativo entre as corporações. Nesse sentido, compreende-se que um atrativo de novos clientes se refere a um bom atendimento por parte da instituição, o qual pode ser potencializado pelo uso da IA, já consolidada como uma das principais tendências de mercado atualmente. Um exemplo da aplicação dos modelos preditivos para atendimento ao cliente, nesse contexto, são os chatbots utilizados por muitos bancos em plataformas digitais. Destacam-se, nesse cenário, os seguintes players de mercado:

Bank of America: um dos maiores bancos dos Estados Unidos, o Bank of America, lançou em 2018 uma assistente virtual que ajuda os clientes a tomarem decisões. Ela já atendeu a mais de 150 mil chamados desde seu lançamento, referentes a sugestões de investimentos, pagamento de contas e emissão de notificações.

Bradesco: constitui um dos cases mais conhecidos de bots que utilizam IA no Brasil. A Bia — assistente virtual do banco — nasceu como uma forma de automatização para o back-office, utilizada apenas por funcionários. Atualmente, atua sobre 91 serviços e produtos do banco.

Banco Original: o chatbot de IA foi criado com o objetivo de ampliar o nível de resolução dos atendimentos, a fim de aumentar a retenção de clientes. A IA atua na resolução de dúvidas e na efetivação de transações bancárias, sendo responsável por cerca de 70% de todos os

atendimentos ao cliente. Desde sua implementação, a taxa de retenção de clientes subiu de 60% para 90%.

Royal Bank of Canada: nos últimos anos, vem utilizando um bot de inteligência artificial que aprende com as solicitações e atividades bancárias dos clientes. Conforme ele vai cruzando os dados de atividades financeiras recentes, torna-se capaz de realizar análises preditivas e antecipar questões e problemas, oferecendo sugestões personalizadas.

AS CINCO FORÇAS DE PORTER

Michael E. Porter, no artigo “How Competitive Forces Shape Strategy”, ressalta a importância de se analisar não somente a estrutura interna do negócio, como também as forças competitivas que o cercam. Considerando-se as circunstâncias sob as quais o Banco PAN se encontra, visualizar a intensidade dessas forças é imprescindível para uma melhor compreensão de sua situação no mercado ante a solução a ser desenvolvida.

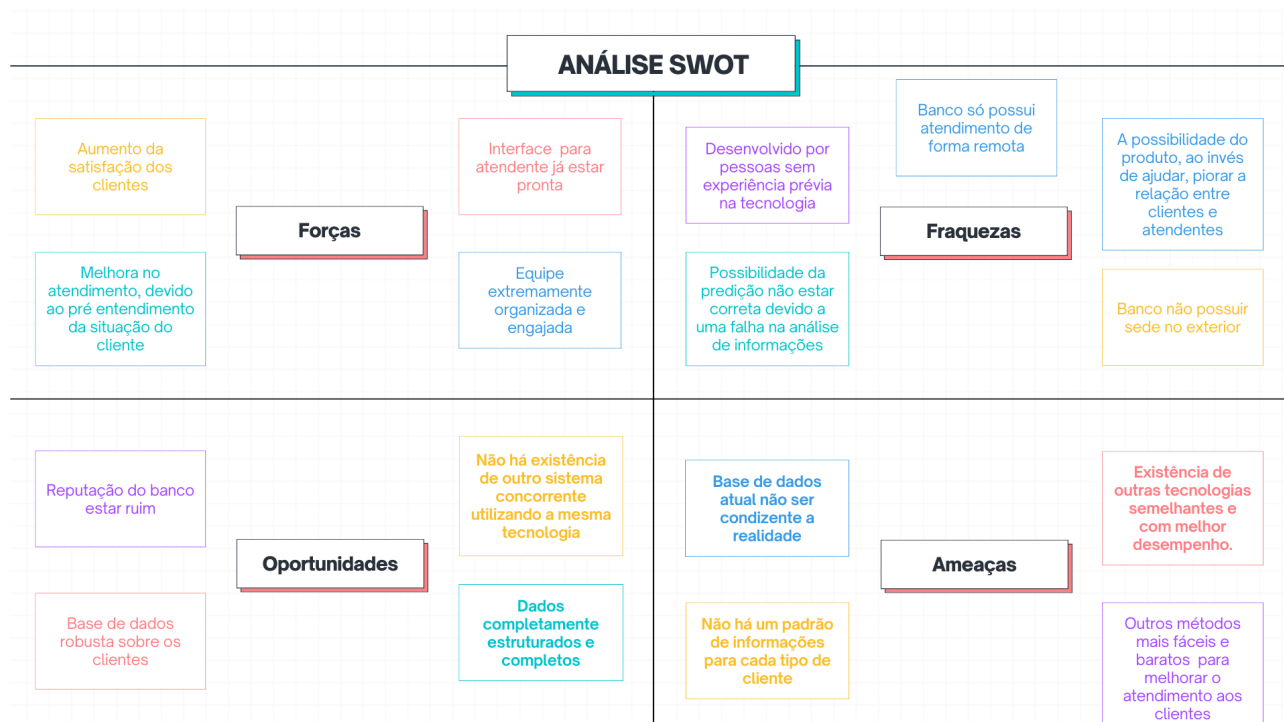
Quando se trata da rivalidade entre concorrentes, é fundamental pontuar que os clientes de um banco, naturalmente, terão preferência por manter uma conta bancária onde houver um melhor sistema de atendimento. Assim, considerando-se o potencial dos sistemas das demais instituições do mesmo segmento do Banco PAN, este encontra-se mais suscetível à força competitiva da concorrência.

Existem, ainda, outras forças competitivas que devem ser consideradas ao se realizar uma análise de negócios. Uma delas se refere ao poder de barganha dos fornecedores, que, nesse contexto, não é prioritário, visto que a equipe será responsável por fornecer ao cliente a solução esperada. O poder de barganha dos compradores, por outro lado, constitui uma ameaça significativa, visto que os clientes do banco, em busca de uma boa prestação de serviços – incluindo um atendimento eficiente –, podem pressioná-lo no sentido de considerar o encerramento de suas contas na instituição.

A ameaça de novos entrantes também constitui um fator de preocupação, haja vista a crescente tendência da adoção de Inteligência Artificial por parte das empresas a fim de oferecer uma boa experiência para o cliente. Por fim, a ameaça de produtos ou serviços substitutos, os quais poderiam configurar alternativas ao modelo preditivo que será construído, não possui relevância significativa, visto que a opção mais favorável na indústria, atualmente, contempla o uso de tecnologias como a que será entregue para o Banco PAN.

Espera-se que a solução desenvolvida acompanhe as atuais tendências do mercado, atingindo as expectativas do Banco PAN quanto ao objetivo de potencializar a qualidade do atendimento que propõe. Desse modo, há a possibilidade de que a proposta, além de beneficiar a instituição, a eleve ao patamar de player de mercado, assim como os outros casos mencionados.

4.1.2. Análise SWOT



4.1.3. Planejamento Geral da Solução

a) Dados disponíveis:

Fonte: dados captados pelo Banco PAN

Conteúdo:

Dados **Descrição**

“vlr_credito”: Crédito no mercado

“vlr_saldo”: Quanto o cliente tem de crédito no PAN (diferentemente do valor que possui na conta, este é quanto ele deve)

“num_atend_atrs”: Número de atendimentos que o cliente tem e estão em atraso (atraso por parte do PAN em não cumprir os prazos)

“vlr_score”: Score do cliente (de 0 a 1000) no mercado

“num_produtos”: Quantidade de produtos que o cliente adquiriu no banco.

“num_atend”: Número de atendimentos, dentro e fora do prazo (é considerado atendimento se gerou um protocolo)

“num_cpf”: CPF do cliente

“qtd_oper”: Quantidade de operações de um cliente referentes a um determinado produto

“qtd_reclm”: Quantidade de reclamações por cliente

“qtd_restr”: Restritivo de mercado, se tem alguma pendência ou não.

“vlr_renda”: Valor de renda do mercado. É uma medida preditiva

“cod_rating”: Rating mensal de risco do cliente, que varia de A (zero risco) a HH (prejuízo absoluto). É uma avaliação interna do banco, não tendo relação com a situação do cliente no mercado. Tem um viés de risco (entender quem é o cliente, como ele se comporta no mercado e no Banco PAN). É uma variável fundamental para medir o atrito dele com o banco.

b) Solução proposta:

A solução que estamos propondo pretende diminuir as dificuldades enfrentadas pela empresa quando o assunto é o relacionamento com clientes do banco, visto que, atualmente, existem muitas reclamações, em diferentes veículos de avaliação, a respeito de atendimento que não estão direcionados para o real problema do atendido.

c) Tipo de tarefa: (regressão ou classificação)

O tipo de tarefa que iremos fazer será de classificar, isso devido ao fato de que o objetivo da nossa IA será de analisar, levando em consideração o comportamento dos clientes do banco no passado, um cliente, assim, em que esse efetuar um contato com o banco seja por qualquer um dos meios hoje disponíveis por eles e assim com base em suas informações classificá-lo como cliente atritado, cliente que busca novos produtos ou cliente novo, passando tal informação para o atendente que efetuará o atendimento deste cliente, isso com a intenção de oferecer diferentes tipos de atendimento, sendo esse correspondente a situação de cada cliente.

d) Utilização da solução proposta:

A solução proposta deverá ser utilizada no sistema do Banco PAN pelos atendentes do canal de telefone, que irão conseguir visualizar qual categoria os clientes que estão em contato se consideram, sendo elas: cliente atritado, cliente novo, clientes que querem adquirir novos produtos.

e) Benefícios trazidos pela solução proposta:

A problemática trazida pelo parceiro indica a falta de informações sobre o cliente como um grande dificultador na comunicação. Isso ocorre porque, nas situações em que ele se encontra a buscar ajuda devido aos problemas que enfrentou diretamente com o banco, acaba sendo recebido com outras indicações de produtos, situação que afeta negativamente a efetividade do sistema de atendimento dos clientes e agrega reclamações constantes ao banco. Situado o ponto fraco da comunicação entre cliente e empresa, a nossa solução busca prever a personalidade do cliente em relação ao banco, indicando, assim, se busca novos produtos ou se está com problemas e dúvidas sobre funcionalidades da empresa. Logo, os

benefícios da aplicação serão reduzir as reclamações em aberto e melhorar a qualidade de atendimento ao cliente.

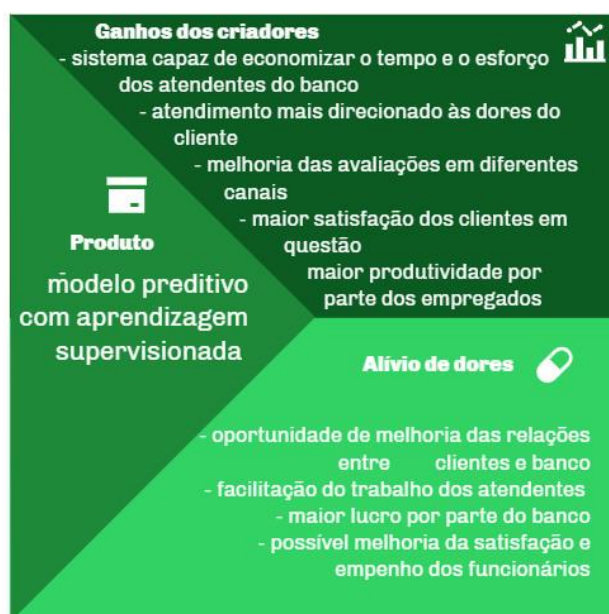
f) Critério de sucesso e a medida utilizada para o avaliar:

Contando que o principal objetivo do banco parceiro seja prever os futuros clientes que podem ter problemas com a empresa de acordo com os dados coletados do mesmo, o principal critério para declarar que o projeto teve êxito é uma boa taxa de acerto em previsões, já que é a medida principal onde podemos avaliar o nosso produto. Dentro das taxas de acerto, será possível analisar também quais os principais aspectos do cliente que alteram os resultados.

4.1.4. Value Proposition Canvas

Canvas da proposta de valor

Proposta de valor



Perfil do cliente



4.1.5. Matriz de Riscos

Matriz de Risco						
Probabilidade	Ameaça			Oportunidade		
Alta	Atraso na entrega dos entregáveis de cada sprints.				Facilitar o trabalho dos atendentes.	
Médio		Não cumprir o propósito de melhorar a relação entre banco e clientela.	Criar um sistema cuja atribuição seja errônea.	Melhorar o padrão de relacionamento do banco PAN com seus clientes.		
Baixa			Vazamento de dados.	O banco PAN colocar a solução desenvolvida em uso.		
	Baixo	Médio	Alta	Alta	Médio	Baixo
	Impacto					

4.1.6. Personas



NOME: Aparecida Silva de Jesus

IDADE: 62 anos

GÊNERO: Feminino

OCUPAÇÃO: Aposentada

"Sou uma mulher guerreira que faz de tudo pelos filhos". Aparecida é moradora de uma simples casa na região do Brás em São Paulo - SP. Ela é mãe de 3 filhos: Werlison, José e Iara, que a ajudam financeiramente. Ela escolheu o Banco PAN para receber sua aposentadoria.

Considerações biográficas e comportamentais

Pouco
conhecimento
tecnológico

Gosta de
simplicidade

Não tem
muita
paciência
no dia a dia

Dores/Motivações atuais com o problema:

Escolheu o
PAN por
produtos
como o saúde
pan

Está enfrentando
alguns problemas
para concluir o
processo de
aposentadoria

Objetivos/necessidades específicas em relação ao problema:

Conseguir
ser uma
cliente nova
do banco

Conseguir
colocar a
aposentadoria
no banco

Obter o
Saúde
PAN



NOME: Claudio Santos Bezerra

IDADE: 28 anos

GÊNERO: Masculino

OCUPAÇÃO: Estudante de mestrado/Professor

"Apenas a educação transforma". Claudio é morador de um apartamento na Consolação em São Paulo - SP. Ele está fazendo mestrado em "Estudos Literários" na USP. Ele está com um score muito baixo no SERASA, porém deseja um empréstimo consignado no Banco PAN.

Considerações biográficas e comportamentais

Deseja investir em um comércio para ajudar na renda

É otimista e esperançoso

Pavio curto

Dores/Motivações atuais com o problema:

Tem insistido para o seu banco disponibilizar o empréstimo, apesar do score baixo.

Liga semanalmente para o SAC do banco PAN para fazer reclamações

Tem tido problemas para pagar o seu financiamento de veículos feito pelo PAN

Objetivos/necessidades específicas em relação ao problema:

Conseguir o empréstimo para o seu comércio

Resolver sua dívida com o financiamento

4.1.7. Jornadas do Usuário

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

4.2. Compreensão dos Dados

1. Descreva os dados a serem utilizados (disponibilizados pelo cliente e outros se tiverem sido incluídos), detalhando a fonte, o formato (CSV, XLSX, banco de dados, etc.), o conteúdo e o tamanho.
- a) Se houver mais de um conjunto de dados, descrição de como serão agregados/mesclados.

Não se aplica, visto que o conjunto de dados é único. A base de dados fornecida pelo Banco PAN consiste em uma tabela que une todos os campos que serão utilizados para análises.

b) Descrição dos riscos e contingências relacionados a esses dados (qualidade, cobertura/diversidade e acesso).

A manipulação dos dados que serão tratados para a construção do modelo preditivo é bastante delicada. Visando extrair da melhor forma possível as informações necessárias, prezando pelo cuidado e segurança do conjunto de materiais que foi confiado à equipe, é fundamental supervisionar os riscos e estabelecer contingências que procurem mitigar os seus impactos.

Os maiores riscos estão associados ao vazamento de dados e à insuficiência de dados para a construção de um modelo preditivo adequado. Todavia, cabe ressaltar que os dados foram fornecidos pelo cliente de forma anonimizada através do uso da tecnologia de criptografia hash para que, caso haja vazamento de tais dados, os prejuízos sejam mínimos.

A linha do tempo dos dados em que o modelo irá se basear compreende o período de tempo do mês de abril do ano 2021 até abril do ano seguinte, sendo os clientes identificados pelo campo CPF, que está mascarado e representa o número do documento do cliente.

Todos os riscos podem ter consequências, logo são importantes e devem ser gerenciados. Por isso, visando minimizar o risco de exposição de informações privadas e sigilosas dos usuários do Banco PAN, cabe ao grupo manipular os dados de maneira cautelosa, contemplando a restrição solicitada de não incluir a base de dados no repositório do GitHub, devendo esta ser administrada de forma independente.

a. Se aplicável: descrição de como será selecionado o subconjunto para análises iniciais:

Não se aplica, visto que não haverá subconjunto para análises iniciais dos dados.

b. Se houver: descrição das restrições de segurança.

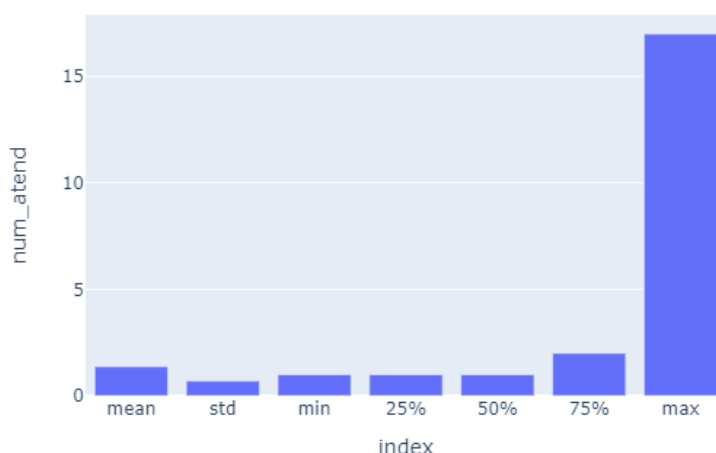
Apesar dos dados já terem sido disponibilizados após uma criptografia daqueles que são sensíveis, a base de dados não deve ser incluída no repositório do GitHub ou em qualquer via de acesso público, sendo gerenciada de maneira independente.

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

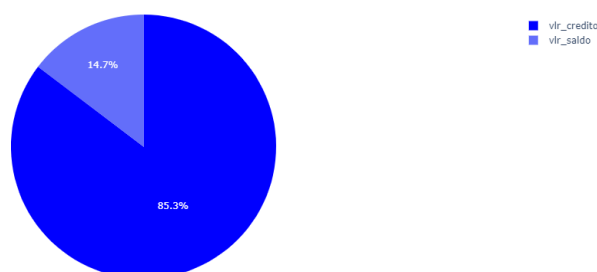
A abordagem do modelo preditivo a ser desenvolvido prioriza as previsões relacionadas aos clientes atritados com o banco. Desse modo, para uma melhor visualização dos dados que embasariam esse objetivo, foram definidos os campos de maior relevância da base de dados utilizada pela equipe: quantidade de reclamações, número de atendimentos atrasados, número total de atendimentos, valor do Score, quantidade de restrições, rating, valor de crédito e valor

de renda. Com base nesse conjunto de informações de cada cliente, será possível desenvolver um modelo capaz de indicar se existem atritos em sua relação com a instituição financeira. Alguns desses campos serão analisados em conjunto. São eles: rating e score, e valor de renda e valor de crédito, de modo que seja possível obter informações mais detalhadas através do cruzamento de dados.

Abaixo, encontram-se as visualizações gráficas realizadas a partir dos cálculos estatísticos dos campos estudados.

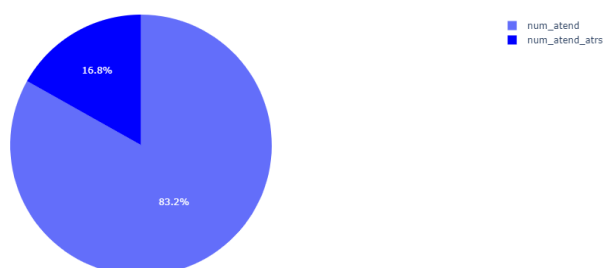


Número de atendimentos por cliente: média, desvio-padrão, mínimo, máximo e quantis

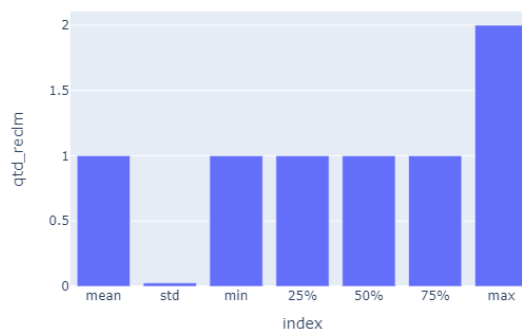


Comparação entre o valor de crédito e o valor de saldo)

Comparação entre o número de atendimentos e atendimentos atrasados:



Comparação entre o número de atendimentos e o número de atendimentos atrasados



Quantidade de reclamações: média, desvio-padrão, mínimo, máximo e quantis

3. Descrição da predição desejada (“target”), identificando sua natureza (binária, contínua, etc.)

As colunas que serão usadas como “target” serão de natureza binária, sendo elas:

ind_atritado	Indicador de cliente atritado, 1 para Sim e 0 para Não
ind_engajado	Indicador de cliente engajado, 1 para Sim e 0 para Não
ind_novo_cliente	Indicador de potencial novo cliente, 1 para Sim e 0 para Não

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.

CACAU, Camila. CHATBOT para bancos: 11 cases globais para conhecer. **TIVIT Labs**, 12 mar. 2021. Disponível em: <https://labs.tivit.com/ivirtualemployee/cases-chatbot-para-bancos/>. Acesso em: 08 ago. 2022.

STRATEGY, How Competitive Forces Shape. by Michael E. Porter. **Harvard Business Review**, 1979

Anexos

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