

2023M3T7-Inteli / grupo5 ⚡

Type to search | [+ ↗](#) | [⟳](#) | [☰](#) | [✉️](#) | [👤](#)

[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Security](#) [Insights](#)

[Files](#)

[main](#) [+](#) [🔍](#)

[Go to file](#)

> [assets](#)
< [documentos](#)
> [outros](#)

[documentacao.md](#)

> [notebooks](#)
[.gitattributes](#)
[.gitignore](#)
[README.md](#)

[grupo5 / documentos / documentacao.md](#)

 [isabellasaldanha](#) [Update documentacao.md](#) 2b0dd1a · 3 hours ago [History](#)

[Preview](#) [Code](#) [Blame](#) 800 lines (487 loc) · 78 KB  Your organization can pay for GitHub Copilot

[Raw](#) [🔗](#) [💾](#) [✍️](#) [⋮](#)

Documentação Modelo Preditivo - Inteli

PrevIA

PrevIA



Integrantes

Drielly Farias, Isabella Fernandes Saldanha, Leonardo Kalid Guene, Luigi Otávio Macedo, Patrick Savoia, Ricardo Novaes

Sumário

- [1. Introdução](#)
- [2. Objetivos e Justificativa](#)
- [3. Metodologia](#)
- [4. Desenvolvimento e Resultados](#)
- [5. Conclusões e Recomendações](#)
- [6. Referências](#)

[Anexos](#)

1. Introdução

A empresa Fortum é uma grande fornecedora de energia, atuando na Finlândia, Suécia, Noruega, Dinamarca, Polônia e Alemanha. Fornece energia elétrica usando fontes renováveis como energia eólica, solar e hidrelétrica. A fim de manter o nível de clientes constante, a Fortum precisa de melhores estratégias para lidar melhor com o churn, ou seja, clientes que deixam de usar o serviço da empresa.

2. Objetivos e Justificativa

2.1 Objetivos

O objetivo é construir um modelo preditivo de churn, capaz de antecipar com precisão a saída de clientes da corporação. Nosso modelo está fundamentado na utilização da linguagem de programação Python, com suporte das bibliotecas pandas e numpy. Além disso, a concepção e implementação deste modelo possui um valor estratégico de grande escala, contribuindo no fortalecimento das estratégias de gestão da mediana. O modelo auxilia a empresa a tomar decisões mais informadas e eficientes para lidar com o churn.

estratégias do setor de marketing. O modelo pretende contribuir a equipe de marketing e desenvolvimento analise eficazes que ajudem a promover campanhas publicitárias direcionadas, almejando prolongar a relação desses clientes com empresa.

2.2 Proposta de solução

A solução para a questão apresentada é o desenvolvimento e a implementação de um modelo preditivo usando Inteligência Artificial (IA) como estratégia de retenção de clientes. A solução será desenvolvida em Python e objetiva prever o churn e auxiliar na definição dos descontos adequados para evitar a perda de clientes. Entre as técnicas que são utilizadas no campo da ciência de dados, a que mais se encaixa no problema proposto é a regressão logística, um método estatístico que calcula a probabilidade de um cliente deixar a empresa, baseando-se em variáveis fornecidas no banco de dados, como histórico de consumo.

Entre as técnicas que pretendemos usar, a Regressão Logística é um método estatístico que calcula a probabilidade de um cliente deixar a empresa, baseando-se em variáveis como idade e histórico de compras. Também utilizaremos o método chamado Random forests (Florestas Aleatórias), compostas por várias árvores de decisão trabalhando juntas, proporcionam uma visão mais abrangente e precisa na identificação e classificação de clientes em grupos de risco. As Florestas Aleatórias operam combinando as previsões de várias árvores de decisão. Cada árvore de decisão é um modelo simples que faz previsões com base em algumas variáveis. Ao combinar o resultado de várias árvores, a Floresta Aleatória consegue uma visão mais precisa e robusta.

2.3 Justificativa

A finalidade do modelo preditivo é a previsão da evasão de clientes da empresa, alinhada também a missão de cooperar ativamente com o time de marketing no impulsionamento de campanhas destinadas a fortalecer a interação dos clientes com os produtos oferecidos. Nesse contexto, é fundamental que a inteligência artificial esteja em paralelo com as equipes de marketing e desenvolvimento, visando um cenário em que a primeira estará responsável por conceber estratégias de engajamento voltadas para a maximização dos ganhos e diminuição do churn, enquanto a segunda se dedica a explorar tecnologias e soluções capazes de automatizar processos e proporcionar dinamismo aos serviços prestados.

3. Metodologia

A metodologia CRISP DM (cross industry standard process for data mining) é um processo amplamente utilizado na mineração e análise de dados. Ela oferece uma estrutura para guiar projetos analíticos desde o início até a implementação, ajudando as equipes a obterem insights valiosos a partir de seus conjuntos de dados. Essa metodologia foi desenvolvida no final dos anos 1990 e é amplamente reconhecida como uma abordagem padrão na indústria de análise de dados. (AFFONSO, 2021). Essa metodologia é dividida em seis fases:

Compreensão do Negócio: fase onde a equipe tenta entender os critérios de negócio da empresa para relacionarem com os dados de trabalho.

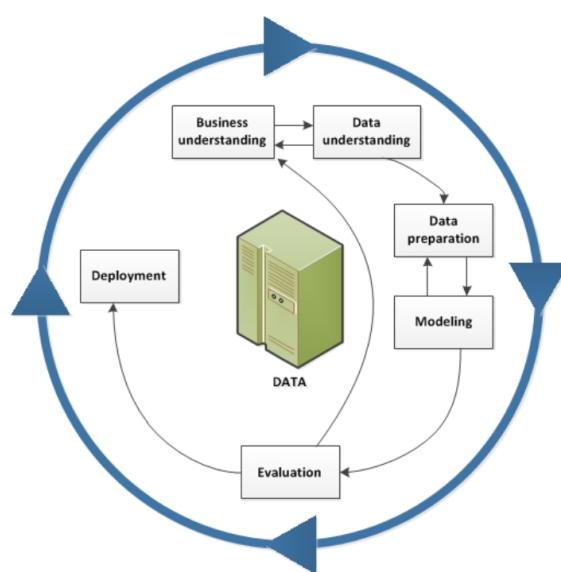
Compreensão dos Dados: analisar e capturar o maior número de features e variáveis que possam ajudar no desenvolvimento e análise do projeto.

Preparação dos dados e modelagem: fases onde há a escalação e criação de hipóteses que podem ajudar no desenvolvimento e treinamento da IA ou de uma análise estatística.

Avaliação: aqui ocorre a finalização do conjunto e o início da análise de feedbacks e reavaliações que podem resultar em novas tecnologias de suporte.

Implantação: sendo a última fase, aqui ocorre a instalação da ferramenta e disposição estratégica com outra equipe de trabalho, tornando a ferramenta imersível e dinâmica no uso cotidiano.

Figura 1 - CRISP-DM



Fonte: ibm.com

Em nosso projeto, entendemos o projeto e a parte de negócios na primeira sprint, entendemos mais sobre os dados na segunda sprint, modelamos na terceira, melhoramos o modelo na quarta sprint e refinamos na quinta. Por ser um processo iterativo, revisitamos algumas etapas várias vezes, considerando os erros e acertos aprendidos ao longo do caminho.

4. Desenvolvimento e Resultados

4.1. Compreensão do Problema

Para entender melhor o problema proposto, usamos os frameworks: 5 forças de Porter, análise SWOT, planejamento geral da solução, value proposition canvas, matriz de risco, personas, jornadas do usuário e política de privacidade.

4.1.1. Contexto da indústria

A Fortum é uma empresa finlandesa de energia elétrica com foco em sustentabilidade e redução da emissão de carbono. Suas principais operações nos países nórdicos compreendem geração de energia eficiente e livre de CO₂, bem como fornecimento confiável de eletricidade e aquecimento distrital para clientes particulares e empresariais.

Principais competidores

EDF (Electricité de France): Uma das maiores empresas de energia da Europa, com presença em geração, distribuição e comercialização de eletricidade. Ocupa a posição de maior empresa do setor de energia elétrica na Europa e é a principal produtora global de energia nuclear, operando 58 usinas nucleares exclusivamente na França.

Vattenfall: Uma empresa de energia sueca de capital aberto que opera em várias regiões, incluindo geração de energia, distribuição e soluções de energia sustentável. Seu compromisso reside na criação livre de combustíveis fósseis. Junto a parceiros, a empresa assume a responsabilidade de desenvolver soluções sustentáveis para eletrificação do transporte, indústrias e sistemas de aquecimento.

Enel Green Power: Fundada em 2008 como parte do Grupo Enel, é uma das maiores empresas de energia da Europa, atuando em geração, distribuição e comercialização de eletricidade, com foco crescente em fontes de energia renovável. Destaca-se sua operação com mais de 1.200 centrais em todos os 5 continentes.

Modelos de negócio

A Fortum possui um modelo de negócios que abrange desde a geração eficiente e livre de CO₂ até o fornecimento de eletricidade. Além disso, a empresa está envolvida na descarbonização das indústrias nórdicas, por meio de parcerias estratégicas e investimentos em energias limpas e hidrelétrica.

Produção de energia sustentável: sendo um dos produtores de energia renovável mais notáveis da Europa, a Fortum foca em gerar energia de uma forma eficiente, buscando fontes renováveis como eólica, solar e hidrelétrica.

Tratamento de resíduos perigosos: a empresa oferece soluções para a gestão de resíduos, incluindo purificação de líquidos radioativos, solidificação de resíduos e armazenamento temporário de combustível usado, com o objetivo de produzir energia limpa e garantir um descarte seguro.

Produção de energia do futuro: a empresa está dedicada em desenvolver soluções avançadas de geração de energia, por conta disso está envolvida na construção da produção de usinas termelétricas em todo o mundo.

Tendências de mercado

Atualmente o mercado de energia elétrica está evoluindo e aumentando sua preocupação ambiental. Por isso, pode-se destacar que as tendências mais observadas são:

Uso eficiente de energia e armazenamento: é essencial buscar melhorar a produção e armazenamento de energia para lidar com a natureza intermitente das fontes de energia renovável. Pois as fontes de energia renovável, como a energia solar e eólica, nem sempre estão disponíveis de forma constante. Sendo assim, é necessário ter maneiras de lidar com essa variação.

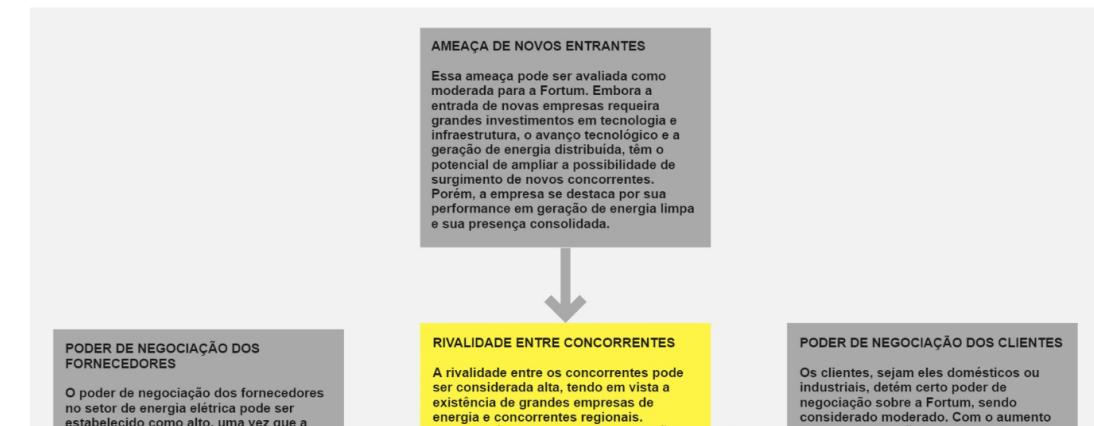
Digitalização e uso de tecnologias avançadas: a utilização de tecnologias avançadas como a análise de dados traz uma transformação significativa na maneira como a energia é produzida, distribuída e utilizada. Por conta disso, a Fortum pode aproveitar essas tecnologias para aprimorar suas operações e proporcionar soluções mais inteligentes aos seus clientes.

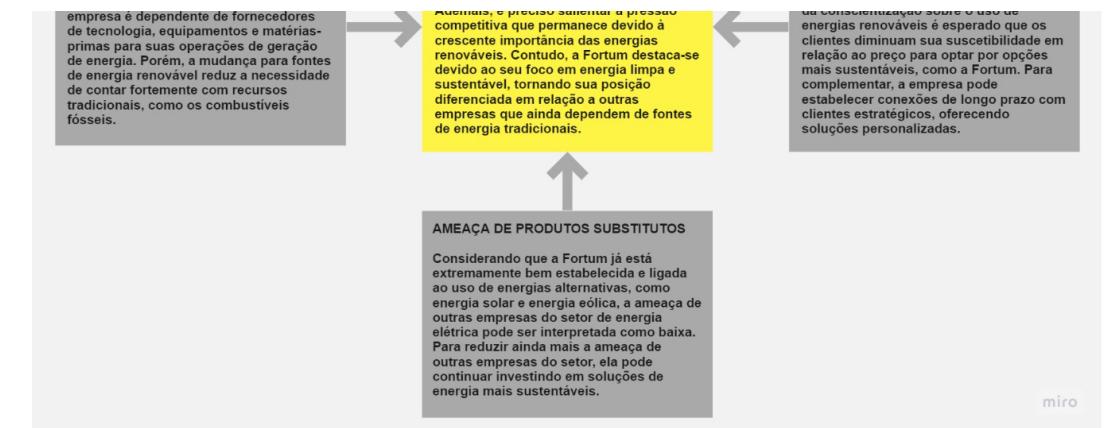
Colaborações com outras empresas: as parcerias estratégicas podem desempenhar um papel importante na promoção da economia circular. Ao trabalhar em conjunto, a Fortum pode explorar oportunidades para reciclar resíduos, reutilizar subprodutos e desenvolver processos que reduzem o desperdício e a pegada de carbono.

5 Forças de Porter

As 5 Forças de Porter representam uma ferramenta essencial para avaliar o ambiente competitivo de um projeto. Esse conceito foi introduzido pelo renomado economista Michael Porter em 1979 e tem sido amplamente utilizado em estratégias empresariais e no planejamento de investimentos. A análise das Cinco Forças permite avaliar o potencial de lucro de uma indústria e a posição estratégica de uma empresa dentro desse contexto. (MAGRETTA, 2019). Ao aplicar essa ferramenta para analisar a Fortum, conforme a figura 1, obtivemos um maior entendimento de como a empresa se encontra no mercado atual e as principais influências:

Figura 2 - 5 Forças de Porter





Fonte: Autoria própria

Produtos substitutos: Considerando que a Fortum já está extremamente bem estabelecida e ligada ao uso de energias alternativas, como energia solar e energia eólica, a ameaça de outras empresas do setor de energia elétrica pode ser interpretada como baixa. Para reduzir ainda mais a ameaça de outras empresas do setor, ela pode continuar investindo em soluções de energia mais sustentáveis. Nesse sentido, os produtos substitutos em questão são: energia solar, eólica, maremotriz e geotérmica. Com o avanço da tecnologia e sociedade como um todo, questões éticas e morais começaram a ser o fator primordial na escolha de um produto. Logo, os produtos substitutos podem estar totalmente ligados com esses fatores, além da vantagem que cada empresa pode proporcionar em seus serviços.

Novos entrantes: Essa ameaça pode ser avaliada como moderada para a Fortum. Embora a entrada de novas empresas requeira grandes investimentos em tecnologia e infraestrutura, o avanço tecnológico e a geração de energia distribuída, têm o potencial de ampliar a possibilidade de surgimento de novos concorrentes. Porém, a empresa se destaca por sua performance em geração de energia renovável e sua presença consolidada.

Clientes: Os clientes, sejam eles domésticos ou industriais, detêm certo poder de negociação sobre a Fortum, sendo considerado moderado. Com o aumento da conscientização sobre o uso de energias renováveis é esperado que os clientes diminuam sua suscetibilidade em relação ao preço para optar por opções mais sustentáveis, trazendo uma certa vantagem para a empresa. Para complementar, a empresa pode estabelecer conexões de longo prazo com clientes estratégicos, oferecendo soluções personalizadas.

Poder de negociação dos fornecedores: O poder de negociação dos fornecedores no setor de energia elétrica pode ser estabelecido como alto, uma vez que a empresa é dependente de fornecedores de tecnologia, equipamentos e matérias-primas para suas operações de geração de energia. Porém, a mudança para fontes de energia renovável reduz a necessidade de contar fortemente com recursos tradicionais, como os combustíveis fósseis.

Concorrentes: A rivalidade entre os concorrentes pode ser considerada alta, tendo em vista a existência de grandes empresas de energia e concorrentes regionais. Ademais, é preciso salientar a pressão competitiva que permanece devido à crescente importância das energias renováveis. Contudo, a Fortum destaca-se devido ao seu foco em energia limpa e sustentável, tornando sua posição diferenciada em relação a outras empresas que ainda dependem de fontes de energia tradicionais.

4.1.2. Análise SWOT

A Análise SWOT, referida em português como Análise FOFA, é uma abordagem utilizada no planejamento estratégico para ajudar tanto indivíduos quanto organizações a reconhecer os pontos fortes, fraquezas, oportunidades e ameaças que afetam suas atividades comerciais ou a concepção de projetos.(Osterwalder, 2011). Nesse contexto, afim de entender mais sobre a empresa, apresentamos a análise referente à Fortum:

Figura 3 - Análise SWOT



Fonte: Autoria própria

Forças (Strengths):

- Presença Geográfica: A Fortum opera em vários países europeus, tendo uma presença significativa e diversificada. Investimento em Energias Renováveis: Focada na energia limpa, a Fortum é reconhecida por seus investimentos em fontes renováveis, como a energia hidrelétrica, eólica, solar, e nuclear.
- Infraestrutura Robusta: A empresa possui uma infraestrutura sólida com plantas, redes, e tecnologia de última geração.
- Parcerias e Alianças Estratégicas: Colaboração com governos, instituições, e outras empresas para promover a sustentabilidade e o crescimento. Reputação Sólida: Marca forte e bem estabelecida no setor de energia.
- Transição para Energia Limpa: Com a crescente demanda por energia limpa, a Fortum tem uma oportunidade única de liderar essa transição na Europa.
- Inovação Tecnológica: Investir em inovação e tecnologia para melhorar a eficiência e a sustentabilidade.

Fraquezas (Weaknesses):

- Dependência de Mercados Regulados: Algumas operações da Fortum podem ser afetadas por regulamentações governamentais rígidas. Possível Exposição a Riscos de Energia Nuclear: Embora a energia nuclear seja uma parte da estratégia de energia limpa da empresa, ela vem com riscos e preocupações ambientais.
- Desafios na Integração de Aquisições: A empresa pode enfrentar desafios na integração efetiva de novas aquisições ou fusões. Concorrência Intensa: O setor de energia europeu é altamente competitivo.

Oportunidades (Opportunities):

- Incentivos Governamentais: Os governos europeus estão apoiando a energia renovável, o que pode se traduzir em incentivos e subsídios favoráveis.

Ameaças (Threats):

- Mudanças Regulatórias: As alterações nas políticas e regulamentações governamentais podem afetar adversamente as operações da empresa.
- Volatilidade dos Preços de Energia: A flutuação nos preços da energia pode impactar as margens de lucro.
- Concorrência de Novos Entrantes: Novas empresas no mercado podem aumentar a competição e pressionar os preços. Questões Ambientais e de Segurança: Possíveis acidentes ou falhas, especialmente relacionados à energia nuclear, podem resultar em danos significativos à reputação e operações da empresa.

4.1.3. Planejamento Geral da Solução

O principal problema identificado é o risco de Churn de clientes da Fortum, o que implica em perda de receita e custos associados à reconquista desses clientes. Além disso, para reter os clientes, é preciso identificar o desconto ótimo no preço da energia que não apenas evita o Churn mas também maximiza a receita da empresa. Uma abordagem para esse desafio é a identificação do desconto ótimo no preço da energia. Isso não envolve apenas a prevenção do Churn, mas também a maximização da receita da empresa. Encontrar o equilíbrio certo entre oferecer descontos atraentes para manter os clientes e garantir que esses descontos não afetem negativamente a lucratividade da empresa é um aspecto chave da estratégia de retenção de clientes.

A solução proposta é um modelo que prevê a probabilidade de Churn para cada cliente nos próximos meses e busca fornecer análises que ajudem a otimizar a estratégia de precificação para reter clientes e maximizar a receita da empresa por meio de:

4.1.3.1 Estimativa da Probabilidade de Churn:

- O modelo utilizará um conjunto de dados do histórico fornecido, incluindo informações como consumo de energia, padrões de pagamento, reclamações e outros comportamentos relevantes dos clientes.
- Por meio de algoritmos de aprendizado de máquina, como regressão logística, o modelo será treinado para identificar padrões e correlações entre esses dados e o Churn.
- Com base nas características atuais de cada cliente, o modelo fornecerá uma estimativa da probabilidade de Churn nos próximos meses. Isso permitirá à empresa identificar os clientes mais propensos a deixar o serviço.

4.1.3.1 Cálculo do Desconto Ótimo no Preço da Energia:

- O modelo também considerará a sensibilidade ao preço de cada cliente, avaliando como diferentes descontos afetam suas decisões de permanecer ou sair.
- Usando métodos de otimização, o modelo determinará o desconto ótimo no preço da energia para cada cliente. Esse desconto será aquele que equilibra a retenção do cliente com a maximização da receita da empresa.
- A abordagem garantirá que os descontos não sejam excessivos, o que poderia impactar negativamente a receita, enquanto mantém os clientes satisfeitos e engajados.
- Essa solução integrada permitirá à Fortum tomar decisões informadas e estratégicas para reduzir o Churn, reter clientes e otimizar sua receita. Além disso, o uso de técnicas avançadas de Data Science garante que a empresa esteja baseando suas decisões em análises precisas e cientificamente embasadas, fortalecendo sua posição competitiva no mercado.

A solução proposta será implementada de forma sistemática e integrada à operação da Fortum, visando a maximização da retenção de clientes e da receita. O processo de utilização envolverá diversas etapas, garantindo uma abordagem abrangente e eficaz para lidar com o risco de Churn e otimizar a precificação. Algumas etapas que a equipe deverá trabalhar:

- Coleta e Atualização de Dados: a equipe de tecnologia e análise de dados será responsável por coletar, integrar e processar os dados relevantes dos clientes. Isso inclui informações de consumo, histórico de pagamentos, reclamações e outras métricas-chave. A coleta de dados será realizada de forma contínua e atualizada periodicamente para garantir a precisão das previsões e das estratégias de retenção.
- Treinamento e Validação do Modelo: A equipe de cientistas de dados treinará os modelos preditivos utilizando algoritmos de aprendizado de máquina. Os modelos serão ajustados e validados para garantir sua eficácia na previsão do Churn.
- Geração de Listas e Recomendações: os modelos gerarão listas periódicas de clientes em risco de Churn, acompanhadas das probabilidades de Churn e dos descontos sugeridos. Essas listas serão disponibilizadas para a equipe de vendas em um formato acessível e compreensível.
- Intervenções Personalizadas: com base nas listas geradas, a equipe de vendas e atendimento poderá identificar os clientes mais propensos ao Churn e planejar intervenções personalizadas. Isso pode incluir a oferta de descontos específicos, comunicações direcionadas para conscientização sobre os benefícios da permanência ou participação em campanhas de retenção.
- Acompanhamento e Análise de Resultados: a eficácia das intervenções será monitorada ao longo do tempo, comparando a retenção de clientes que receberam as ofertas personalizadas com aqueles que não receberam. A equipe de análise de dados avaliará regularmente o desempenho do modelo e fará ajustes conforme necessário.

Assim, a solução será integrada ao fluxo de trabalho existente da equipe de vendas e atendimento, permitindo a tomada de decisões estratégicas para retenção de clientes. A combinação de previsões precisas, descontos otimizados e intervenções direcionadas garantirá uma abordagem eficiente na gestão do Churn e na maximização da receita da Fortum.

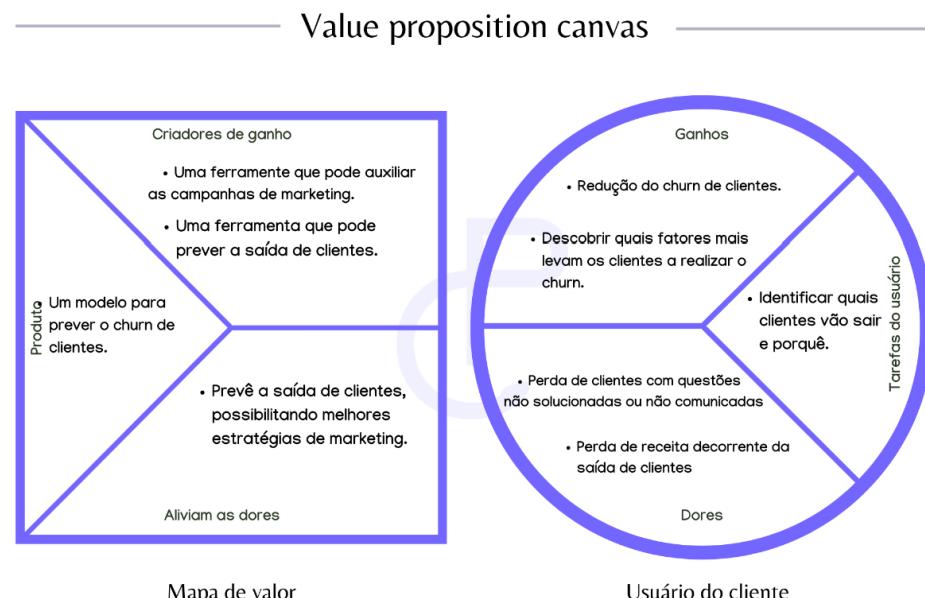
O critério de sucesso será baseado na assertividade da taxa de Churn em comparação com um conjunto de dados que será usado para testar esse quesito. Além disso, a receita retida (ou ganha) devido aos descontos otimizados, em comparação com um cenário em que descontos arbitrários são oferecidos, servirá também como medida de sucesso. Especificamente, as métricas a serem monitoradas incluem:

Taxa de Churn: Uma redução na taxa de Churn após a implementação da solução em comparação com a taxa anterior. **Receita Retida:** A diferença entre a receita potencial perdida devido ao Churn e a receita realmente perdida após a implementação do modelo e das ações de retenção. **ROI (Retorno sobre Investimento):** Avaliar o retorno sobre o investimento feito na implementação do modelo e nas campanhas de retenção versus o valor retido através da redução do Churn e da receita maximizada.

4.1.4. Value Proposition Canvas

O Value Proposition Canva (Canvas de Proposta de Valor) é um framework que permite entender qual valor do produto ou negócio oferece para os clientes. É dividido em duas partes: proposta de valor e segmento de cliente. A primeira analisa o produto, os criadores de ganho e o alívio de dores e a segunda visa entender quais são os trabalhos do cliente, as dores e os ganhos. Ao relacionar esses pontos, é possível obter um maior entendimento do produto e seu valor. (Osterwalder, 2014) Assim, nossa solução será um modelo preditivo utilizando regressão logística, a fim de identificar os clientes em risco e, consequentemente, oferecer descontos e promoções personalizados para promover a retenção eficaz.

Figura 4 - Value Proposition Canvas



Fonte: Autoria própria

Mapa de valor

Criadores de ganho:

- Uma ferramenta que pode auxiliar as campanhas de marketing.
- Uma ferramenta que pode prever a saída de clientes.

Produto:

- Um modelo para prever o churn dos clientes. **Aliviam as dores:**
- Prevê a saída de clientes, possibilitando campanhas de marketing.

Usuário do cliente

Ganhos:

- Redução do churn de clientes.
- Descobrir quais fatores mais levam os clientes a realizar o churn.
- Dores:**
 - Perda de clientes com questões não solucionadas ou previstas.
 - Perda de receita decorrente da saída de clientes
- Tarefas do usuário:**
 - Identificar quais clientes vão sair e o porquê.

4.1.5. Matriz de Riscos

A utilização da matriz de risco desempenha um papel significativo na avaliação e controle dos riscos associados a projetos ou implementações. Uma matriz cuidadosamente construída proporciona à equipe a capacidade de hierarquizar os riscos de maior relevância e de implementar ações preventivas e corretivas apropriadas, reduzindo assim os potenciais efeitos adversos. (NAPOLEÃO, 2019) No contexto deste projeto particular, foi formulada a seguinte análise de riscos, levando em consideração as características específicas da equipe envolvida e das nuances inerentes ao próprio projeto.

Figura 5 - Matriz de Risco

	Riscos Negativos			Riscos Positivos			
90%	Dados faltantes na base de dados	Os dados não serem suficientes para treinar	Falha na comunicação do grupo	A IA conseguir prever o Churn dos clientes em questão			90%
60%	Divergências com o cliente	Não atingir a melhor acurácia que nosso modelo pode proporcionar.	A IA performar comportamentos antiéticos	A IA atingir um alto nível de acurácia	A IA ser implementada de verdade		60%
30%	Algum integrante do grupo faltar muito	Escolher as features erradas para o treinamento da IA	Não conseguir finalizar a IA e prever o Churn	O modelo desempenhar uma performance excelente nos primeiros testes	Incrementar novas funcionalidades e tecnologias à IA		30%
P/I	Baixo	Médio	Alto	Alto	Médio	Baixo	

Com o objetivo de aprimorar a legibilidade da imagem, foi feito a transcrição dos textos presentes e o plano de ação para os riscos e oportunidades:

Fonte: Autoria própria

Riscos potenciais:

-Algum integrante do grupo faltar muito -A IA performar comportamentos antiéticos -Não conseguir finalizar a IA nem prever o Churn -Divergências com o cliente

Riscos reais:

-Escolher as features erradas para o treinamento da IA -Não atingir a acurácia que nosso modelo pode proporcionar -Dados faltantes na base de dados -Os dados não serem suficientes para treinar -Falha na comunicação do grupo

Plano de ação e resposta para todos os riscos:

Potenciais:

-Para evitar que um membro da equipe falte muito é necessário ocorrer uma interação e acompanhamento do grupo regularmente, especialmente durante a dailys e retrospectivas buscando entender o contexto de cada um. -Construir o modelo preditivo com base na ética, buscando entender os resultados dos testes antes para implementá-los com responsabilidade. -Criar um planejamento muito criterioso e dividir bem as tarefas do grupo para analisar o banco de dados de uma forma ágil, otimizando o tempo para que o grupo consiga trabalhar no modelo preditivo de forma satisfatória e dentro do prazo. -Buscar entender todas as sugestões do cliente, elencando as melhores formas de implementação, sempre respeitando opiniões divergentes e prezando pela comunicação não-violenta.

Reais:

-Fazer a etapa de entendimento de negócio e entendimento dos dados com atenção, a fim de ter informações o suficiente para julgar quais são as features adequadas. -Estabelecer boas métricas de sucesso para conseguir identificar os níveis atingidos e as possíveis melhorias. -Analizar, tratar e limpar os dados antes de começar a treinar o modelo, a fim de não influenciar o modelo com dados incompletos. -Analizar e entender bem os dados, a fim de fazer o melhor uso possível das informações fornecidas -Ter atenção com as dailys, sprint reviews e planejamentos, momentos que facilitam a comunicação e evitam conflitos entre os integrantes.

Oportunidades potenciais:

-A IA ser implementada de verdade -O modelo desempenhar uma performance excelente nos primeiros testes -Incrementar novas funcionalidades e tecnologias a IA

Oportunidades reais:

-A IA atingir um alto nível de acurácia -A IA conseguir prever o Churn dos clientes em questão

Plano de ação e resposta para todos as oportunidades:

Potenciais:

-Construir um modelo preditivo com uma acurácia alta e que consiga prever o Churn dos clientes com êxito pode levar a empresa a querer dar continuidade no projeto e implementá-lo de verdade. -Desenvolver um modelo que consiga ter uma ótima performance mesmo nos teste iniciais e consiga dar uma vantagem para o andamento do projeto. -Com o êxito e finalização do escopo do projeto, iniciar a implementação de melhorias e novas tecnologias alinhadas ao modelo desenvolvido buscando aprimorar ainda mais a IA.

Reais: -Manejar para que a IA tenha uma boa implementação e o alto nível de acurácia não elimine as respostas fora do padrão treinado. -A IA ser usada na estratégia de marketing da empresa.

4.1.6. Personas

Uma persona é um perfil fictício de usuário. Ela representa um cliente idealizado e ajuda as equipes de desenvolvimento a compreenderem melhor quem são seus principais públicos-alvo, suas necessidades, desafios, hábitos e comportamentos, o que colabora para a criação de estratégias e produtos direcionados.(HARLEY, 2015).

Para esse projeto, escolhemos duas personas. Começando com a Amanda Nunes, gerente de contas. Ela é fundamental na criação de estratégias alinhadas com o time de marketing e desenvolvimento por ter acesso integrado a toda base lógica do problema em questão. Vale ressaltar que, o modelo preditivo irá funcionar como um item mediador em seu trabalho, e não uma ferramenta com uso direto.

A seguir, apresentamos algumas informações relevantes das nossas personas, como: considerações biográficas e comportamentais, dores/motivações atuais com o problema e seus objetivos gerais.

Figura 6 - Persona 1



NOME: Amanda Nunes

IDADE: 38

GÊNERO: Feminino

OCUPAÇÃO: Gerente de contas

Considerações biográficas e comportamentais

Amanda tem mais de 8 anos de experiência na gestão de contas dentro da indústria de energia. Ela é conhecida por sua habilidade de comunicação, liderança e pensamento estratégico.

Amanda é altamente focada em metas e trabalha com sua equipe para alcançar e muitas vezes superar as metas de vendas e retenção de clientes.

Ela valoriza relacionamentos de longo prazo com os clientes e trabalha arduamente para entender suas necessidades e expectativas.

Dores/Motivações atuais com o problema:

A retenção de clientes em um mercado competitivo é uma preocupação constante. Amanda busca maneiras inovadoras de manter os clientes satisfeitos sem comprometer a lucratividade da empresa.

Ela muitas vezes enfrenta desafios na comunicação entre o departamento de vendas e os técnicos que desenvolvem as soluções, o que pode levar a mal-entendidos ou atrasos.

Amanda está sempre atenta às mudanças no mercado e às necessidades dos clientes, o que requer adaptações rápidas e eficientes.

Objetivos/necessidades específicas em relação

ao problema:

Amanda quer oferecer descontos e pacotes que sejam personalizados para cada cliente, aumentando a satisfação e a lealdade do cliente.

Elá precisa de uma colaboração eficaz com a equipe técnica (incluindo o Cientista de Dados) para garantir que as soluções propostas sejam viáveis e alinhadas com os objetivos da empresa.

Amanda deseja um sistema que permita o acompanhamento em tempo real do desempenho das contas, para que ela possa tomar decisões informadas e responder rapidamente às necessidades dos clientes.

miro

Fonte: Autoria própria

Já nossa segunda persona é Anderson Silva, um profissional com mais de 10 anos de experiência na área de DataScience. Anderson trabalhará em conjunto com Amanda, mas, especificamente, com o time de marketing e publicidade. No seu trabalho haverá mais do que análises de dados do treinamento da IA, haverá integração de novas tecnologias - com base no feedback dos clientes e reavaliações - e disposição de estratégias para evitar o churn de futuros clientes.

Figura 7 - Persona 2



NOME: Anderson Silva

IDADE: 40

GÊNERO: Masculino

OCUPAÇÃO: Cientista de dados

Considerações biográficas e comportamentais

Anderson possui um Ph.D. em Ciência da Computação e mais de 10 anos de experiência na indústria de energia. Ele é metódico, detalhista e altamente analítico.

Ele está comprometido com o uso responsável de dados e tecnologia, sempre considerando o impacto ambiental e social de suas decisões.

Anderson valoriza o trabalho em equipe e a comunicação aberta com colegas de diferentes áreas. Ele acredita que uma abordagem interdisciplinar leva a soluções mais eficazes.

Dores/Motivações atuais com o problema:

A vasta quantidade e variedade de dados do setor de energia tornam a tarefa de filtragem e análise um desafio constante. Anderson sente a pressão para encontrar soluções eficientes e precisas.

Ele está preocupado em encontrar o equilíbrio certo entre maximizar o lucro da empresa e fornecer um desconto justo e atraente para o cliente.

Anderson sabe que o mercado é altamente competitivo e está motivado a desenvolver uma solução rápida, mas sem comprometer a qualidade e precisão.

Objetivos/necessidades específicas em relação ao problema:

Ele quer criar uma

Anderson precisa que a solução se integre

Ele está empenhado em estabelecer métricas e

solução que seja precisa, escalável e capaz de se adaptar a mudanças no mercado.

Solução se integra perfeitamente com os sistemas e processos atuais da empresa, garantindo uma transição suave.

processos de monitoramento que permitam melhorias contínuas no modelo, para garantir que a solução permaneça relevante e eficaz.

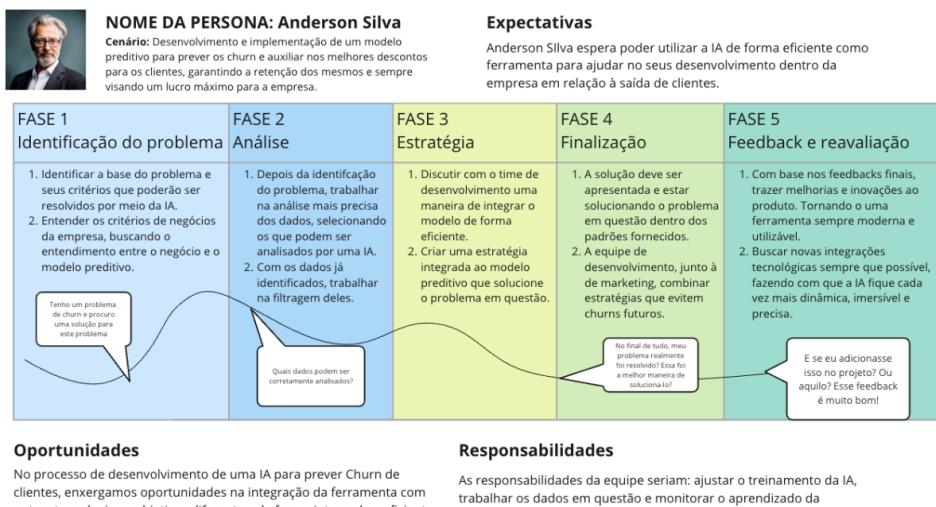
miro

Ambos são pessoas experientes em suas respectivas áreas de atuação e se beneficiariam do uso do modelo preditivo, observando impactos na equipe de dados e de negócios.

4.1.7. Jornadas do Usuário ↗

A jornada do usuário, de acordo com Nielsen, é um modelo que descreve a experiência de um usuário ao interagir com um produto ou serviço ao longo do tempo. Ela envolve várias etapas, como a descoberta do produto, a aprendizagem sobre seu uso, a realização de tarefas, a solução de problemas e a avaliação contínua da experiência. A análise da jornada do usuário proporciona insights para otimizar a usabilidade e a satisfação do usuário com o produto ou serviço. Afim de entender melhor o contexto de uso e as interações entre nossas pessoas, Amanda e Anderson, e o modelo preditivo, desenvolvemos as jornadas representadas nas figura 8 e figura 9.

Figura 8 - Jornada de Usuário 1



Fonte: Autoria própria

Nome:Anderson Silva

Cenário: Desenvolvimento e implementação de um modelo preditivo para prever os churn e auxiliar nos melhores descontos para os clientes, garantindo a retenção dos mesmos e sempre visando um lucro máximo para a empresa.

Expectativas: Anderson Silva espera poder utilizar a IA de forma eficiente como ferramenta para ajudar no seu desenvolvimento dentro da empresa em relação à saída de clientes.

Oportunidades: No processo de desenvolvimento de uma IA para prever Churn de clientes, enxergamos oportunidades na integração da ferramenta com outras tecnologias e objetivos diferentes, de forma integrada e eficiente.

Responsabilidades: As responsabilidades da equipe seriam: ajustar o treinamento da IA, trabalhar os dados em questão e monitorar o aprendizado da ferramenta de forma que os resultados sejam relevantes e úteis.

Fase 1 - Identificação do problema: -Identificar a base do problema e seus critérios que poderão ser resolvidos por meio da IA. -Entender os critérios de negócios da empresa, buscando o entendimento entre o negócio e o modelo preditivo.

Fase 2 - Analise: -Depois da identificação do problema, trabalhar na análise mais precisa dos dados, selecionando os que podem ser analisados por uma IA. -Com os dados já identificados, trabalhar na filtragem deles.

Fase 3 - Estrategia: -Discutir com o time de desenvolvimento uma maneira de integrar o modelo de forma eficiente. -Criar uma estratégia integrada ao modelo preditivo que solucione o problema em questão.

Fase 4 - Finalização: -A solução deve ser apresentada e estar solucionando o problema em questão dentro dos padrões fornecidos. -A equipe de desenvolvimento, junto à de marketing, combinar estratégias que evitem chuns futuros.

Fase 5 - Feedback e reavaliação: -Com base nos feedbacks finais, trazer melhorias e inovações ao produto. Tornando o uma ferramenta sempre moderna e utilizável. -Buscar novas integrações tecnológicas sempre que possível, fazendo com que a IA fique cada vez mais dinâmica, imersiva e precisa.

> Figura 9 - Jornada de Usuário 2



NOME DA PERSONA: Amanda Silva

Cenário: Utilização do modelo preditivo de IA para prever o churn de clientes e desenvolver estratégias de retenção que incluem a oferta de

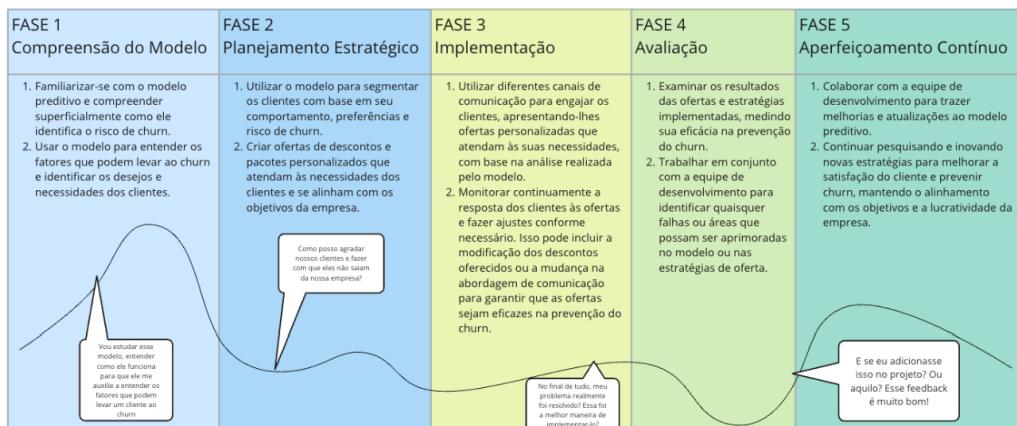
Expectativas

Amanda espera utilizar o modelo preditivo de IA para identificar e prever os riscos de churn, melhorando a classificação do cliente e contribuindo para a



descontos personalizados, mantendo um equilíbrio entre a satisfação do cliente e os objetivos de lucro da empresa.

Fase de churn, mantendo o equilíbrio do cliente e contribuindo para retenção, sem comprometer a lucratividade da empresa.



Oportunidades

Amanda tem a oportunidade de inovar na forma como a empresa interage com seus clientes. Utilizando a IA, ela pode identificar padrões de comportamento, criar ofertas personalizadas e desenvolver abordagens proativas para a retenção de clientes, integrando a tecnologia de forma criativa e eficaz.

Responsabilidades

As responsabilidades incluem a compreensão profunda do modelo preditivo de IA, a criação e implementação de estratégias de retenção de clientes com base nas previsões, a comunicação eficaz com os clientes, e o monitoramento e ajuste contínuo das estratégias para assegurar que elas estejam alinhadas com os objetivos da empresa.

Fonte: Autoria própria

Nome: Amanda Silva

Cenário: Utilização do modelo preditivo de IA para prever o churn de clientes e desenvolver estratégias de retenção que incluem a oferta de descontos personalizados, mantendo um equilíbrio entre a satisfação do cliente e os objetivos de lucro da empresa.

Expectativas: Amanda espera utilizar o modelo preditivo de IA para identificar e prever os riscos de churn, melhorando a satisfação do cliente e contribuindo para a retenção, sem comprometer a lucratividade da empresa.

Oportunidades: Amanda tem a oportunidade de inovar na forma como a empresa interage com seus clientes. Utilizando a IA, ela pode identificar padrões de comportamento, criar ofertas personalizadas e desenvolver abordagens proativas para a retenção de clientes, integrando a tecnologia de forma criativa e eficaz.

Responsabilidades: As responsabilidades incluem a compreensão profunda do modelo preditivo de IA, a criação e implementação de estratégias de retenção de clientes com base nas previsões, a comunicação eficaz com os clientes, e o monitoramento e ajuste contínuo das estratégias para assegurar que elas estejam alinhadas com os objetivos da empresa.

Fase 1 - Compreensão do modelo: -Familiarizar-se com o modelo preditivo e compreender superficialmente como ele identifica o risco de churn. -Usar o modelo para entender os fatores que podem levar ao churn e identificar os desejos e necessidades dos clientes.

Fase 2 - Planejamento estratégico: -Utilizar o modelo para segmentar os clientes com base em seu comportamento, preferências e risco de churn. -Criar ofertas de descontos e pacotes personalizados que atendam às necessidades dos clientes e se alinhem com os objetivos da empresa.

Fase 3 - Implementação: -Utilizar diferentes canais de comunicação para engajar os clientes, apresentando-lhes ofertas personalizadas que atendam às suas necessidades, com base na análise realizada pelo modelo. -Monitorar continuamente a resposta dos clientes às ofertas e fazer ajustes conforme necessário. Isso pode incluir a modificação dos descontos oferecidos ou a mudança na abordagem de comunicação para garantir que as ofertas sejam eficazes na prevenção do churn

Fase 4 - Avaliação: -Examinar os resultados das ofertas e estratégias implementadas, medindo sua eficácia na prevenção do churn. -Trabalhar em conjunto com a equipe de desenvolvimento para identificar quaisquer falhas ou áreas que possam ser aprimoradas no modelo ou nas estratégias de oferta.

Fase 5 - Aperfeiçoamento contínuo: -Colaborar com a equipe de desenvolvimento para trazer melhorias e atualizações ao modelo preditivo. -Continuar pesquisando e inovando novas estratégias para melhorar a satisfação do cliente e prevenir churn, mantendo o alinhamento com os objetivos e a lucratividade da empresa.

4.1.8 Política de Privacidade

A Lei Geral de Proteção de Dados (LGPD) define os direitos do titular dos dados com relação a eles. É regulamentada pela Autoridade Nacional de Proteção de Dados (ANPD) e visa assegurar o uso adequado das informações coletadas. A LGPD abrange a política de privacidade, que é específico sobre como os dados são coletados e usados.(ALBERTACCI, 2021). As informações necessárias para fornecer o serviço dependem do produto ou serviço e é baseado em um contrato, quando exigido por lei, mediante consentimento do cliente. Segundo a política de proteção de dados da Fortum, os seguintes dados são coletados:

Dados pessoais: nome, endereço, número de telefone, endereço de e-mail, sexo, idade, língua materna, nacionalidade, ocupação, interesses, segmento de cliente e número de segurança social necessários para verificar a identidade.

Dados de contratos e transações – informações sobre contratos, pedidos, compras, status de pagamento e faturas; ligações gravadas e transcritas; informações sobre pedidos e objeções aos dados; solicitações de serviço e trocas com serviço de atendimento ao cliente.

Informações de pagamento e crédito – cartão de pagamento e dados bancários necessários para confirmação de compra e reembolsos, e informações de solvabilidade.

Dados e identificadores de tráfego da Web – informações coletadas por meio de cookies ou tecnologias semelhantes sobre como os serviços foram usados, como eventos e segmentos de navegação no site, endereço IP, ID de cookie, ID de telefone celular, navegador

Serviços foram usados, como eventos e segmentos de navegação no site, endereço IP, ID de cookie, ID de telefone celular, navegação, dispositivo e informações de localização.

Dados de Segurança – informações usadas para proteger o uso de serviços e instalações, como senhas e informações de login e registros de segurança.

Dados técnicos e dados de consumo – por exemplo, dados relacionados com a utilização de um dispositivo ou aplicação, incluindo a medição do consumo e produção de eletricidade e outras mercadorias, bem como dados gerados por estações de carregamento e dispositivos inteligentes, como dados de sensores (por exemplo, temperatura).

Esses dados são coletados a partir de informações cedidas pelo usuário, quando encomenda ou usa os serviços, preenche formulários de contato, participa de uma pesquisa ou concurso, cria uma conta, visita nosso site ou interage com a empresa. Além disso, os dados também são coletados de registros de endereços públicos, agências de referência de crédito, agências de cobrança de dívidas, parceiros de instalação, parceiros de marketing, companhias de eletricidade e seguros.

Os dados são coletados com a finalidade de prestação de serviços, atendimento ao cliente, vendas, marketing, anúncios online, comunicação com stakeholders e clientes, desenvolvimento de produtos e serviços, obrigações legais como leis contábeis e fiscais, além de defesa dos direitos legais e manutenção da segurança dos serviços, prevenindo ataques, fraudes e abusos.

A Fortum armazena os dados no próprio banco de dados e, eventualmente, compartilha com empresas do Grupo Fortum, com parceiros comerciais, como empresas de redes de eletricidade e agências de cobrança de dívidas, subcontratados e autoridades, como a polícia, na medida exigida por lei. Esses ficam armazenados de 1 a 20 anos, a depender do tipo de dado.

Com a ajuda de cookies e tecnologias semelhantes, a Fortum é capaz de contar visitantes individuais em seus sites, ler certas informações do dispositivo do usuário e, em alguns casos, determinar quando o usuário visualizou determinado conteúdo ou abriu um determinado e-mail, associar informações de rastreamento ao perfil do usuário e identificar o endereço IP. A Fortum utiliza 4 tipos de cookies, são eles:

Cookies necessários: essenciais ao funcionamento do site, usados apenas para responder a ações do usuário. Não armazenam informações de identificação pessoal e, embora o usuário possa negar, algumas partes do site não funcionarão sem eles.

Cookies de desempenho: são usados para contar visitas ao site e fontes de tráfego, a fim de medir e melhorar o desempenho. São definidos pela Fortum ou por prestadores de serviços externos. Se esses cookies não forem permitidos, a Fortum não saberá quando o site foi visitado e não poderá monitorar seu desempenho. Cookies Funcionais: servem para melhorar a funcionalidade do site e torná-lo mais pessoal.

Cookies de segmentação: podem ser usados por essas empresas para criar um perfil de seus interesses e exibir anúncios relacionados a você em outros sites. Eles não armazenam nenhuma informação pessoal diretamente, mas são baseados exclusivamente na identificação do navegador do usuário e do dispositivo de rede. Se esses cookies não forem permitidos, haverá menos publicidade direcionada.

Cookies de rede social: servem para identificar o usuário para que ele possa, por exemplo, compartilhar artigos no serviço pressionando um botão. São capazes de rastrear o navegador do usuário para outros sites e criar um perfil de seus interesses. Isso pode afetar o conteúdo e as mensagens que aparecem em outros sites que o usuário visita. Se esses cookies não forem permitidos, talvez não seja possível usar essas ferramentas de compartilhamento.

Os cookies podem ser de sessão, que são excluídos quando o usuário fecha o navegador ou de persistentes, que permanecem no dispositivo do usuário por um determinado período de tempo ou até que o usuário os exclua e podem ser verificados a partir do administrador de cookies. Além disso, ferramentas como Google Analytics e Google Firebase são usadas para monitorar o uso dos serviços Fortum.

A fim de manter a segurança dos dados, a empresa utiliza um modelo de gestão de segurança cibernética que descreve funções e responsabilidades, realizando campanhas de conscientização para melhorar a capacidade dos funcionários da Fortum de proteger a privacidade e a segurança, além de exigir o mesmo nível de medidas de segurança em vigor por parte de empresas que tenham acesso aos dados.

Como titular dos dados, o usuário tem o direito de acesso aos dados, à retificação, à correção, atualização, apagamento e à portabilidade de dados pessoais. Também possui o direito de retirar o seu consentimento, direito de se opor ou restringir o processamento.

Todos os direitos estão explicitados na página de política de privacidade da empresa. Se o usuário quiser mais informações ou quiser exercer seus direitos como titular do dado, é possível entrar em contato através do endereço físico, de email, e através do serviço de atendimento ao cliente, por telefone. Além disso, pedido de informação, ou seja, acesso à cópia dos dados que a Fortum possui sobre o usuário é facilitado e de fácil acesso através do site quando comparado com outras empresas.

De modo geral, fica claro a transparência com o tratamento de dados, pois possui um site intuitivo e que permite fácil acesso aos direitos, políticas e contatos para maiores esclarecimentos. Não há um contato direto para o encarregado de dados, mas há o direcionamento quanto a quem contatar em caso de dúvidas.

4.2. Compreensão dos Dados

4.2.1. Exploração de dados

Com base nos dados fornecidos pela PowerCO, que foram coletados de clientes que realizaram ou não o churn, temos acesso a três tabelas cruciais:

base_clientes.csv: essa tabela representa o núcleo das informações, abrangendo um total de 32 colunas. Essas colunas oferecem insights detalhados sobre os clientes, incluindo dados relacionados ao consumo de energia, projeções de consumo, preços previstos e informações sobre as datas de início e término dos contratos.

base_hist_churn.csv: esta tabela atua como um registro completo de todos os clientes, indicando se eles efetuaram ou não o churn. Composta por apenas 2 colunas, essa tabela fornece informações críticas sobre o comportamento de churn dos clientes.

base_precos.csv: esta tabela é dedicada aos preços praticados em diferentes intervalos de tempo. Com um total de 8 colunas, ela documenta as variações de preços ao longo do tempo, permitindo uma análise abrangente das tendências de preços.

Identificação das colunas numéricas e categóricas

Baseado nessas tabelas foi feita a análise de todos os dados distribuindo-os entre numéricos e categóricos, sendo eles:

Numéricos: cons_12m, cons_gas_12m, cons_last_month, date_activ, date_end, date_first_activ, date_modif_prod, date_renewal, forecast_base_bill_ele, forecast_base_bill_year, forecast_bill_12m, forecast_cons, forecast_cons_12m, forecast_cons_year, forecast_discount_energy, forecast_meter_rent_12m, forecast_price_energy_p1, forecast_price_energy_p2, forecast_price_pow_p1, imp_cons, margin_gross_pow_ele, margin_net_pow_ele, nb_prod_act, net_margin, num_years_antig, pow_max.

Categóricos: id, activity_new, campaign_disc_ele, channel_sales, has_gas, origin_up, churn.

Estatística descritiva das colunas

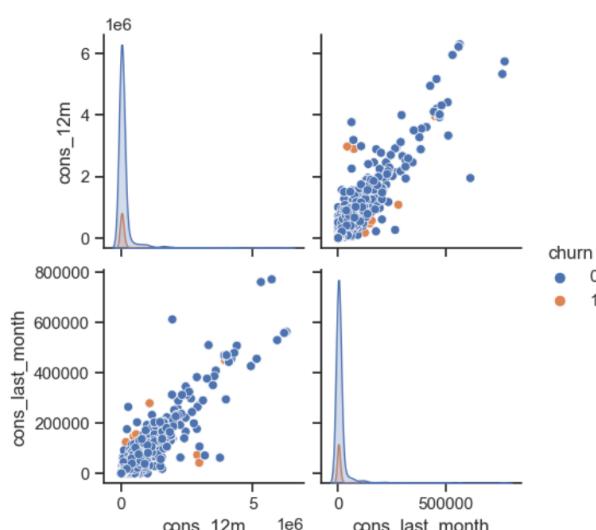
Após a identificação das colunas, procedeu-se à análise descritiva das tabelas. Neste estágio, devido à ausência de tratamento prévio dos dados, identificou-se uma notável disparidade entre os valores, evidenciando a presença de diversos outliers e campos com valores nulos. Adicionalmente, merece destaque a análise das colunas em que os desvios padrão superam consideravelmente as médias, tais como "cons_last_month" e "forecast_cons_12m", indicando uma variabilidade substancial nos dados. Por fim, algumas colunas apresentam uma considerável diferença entre os valores máximo e mínimo, acentuando ainda mais a presença de outliers. Essas análises ressaltam a necessidade de futuros passos na análise e tratamento dos mesmos.

Gráficos gerados para visualização da relação entre as variáveis escolhidas

Após a realização das análises e dos procedimentos de tratamento de dados, conseguimos identificar as variáveis que foram selecionadas para a compreensão dos aspectos iniciais relacionados ao churn dos clientes. Isso se alinha com a preparação para as hipóteses que serão detalhadas posteriormente. Como parte desse processo, geramos os seguintes gráficos:

- Diferença entre o consumo de eletricidade nos últimos 12 meses (cons_12m) e o consumo no último mês (cons_last_month)

Figura 10 - Consumo de energia anterior e atual



Fonte: Autoria própria

Este gráfico ilustra uma tendência notável: os clientes que tiveram um menor consumo nos últimos meses parecem ser mais propensos a realizar o churn, enquanto é rara a ocorrência de churn entre os clientes com alto consumo.

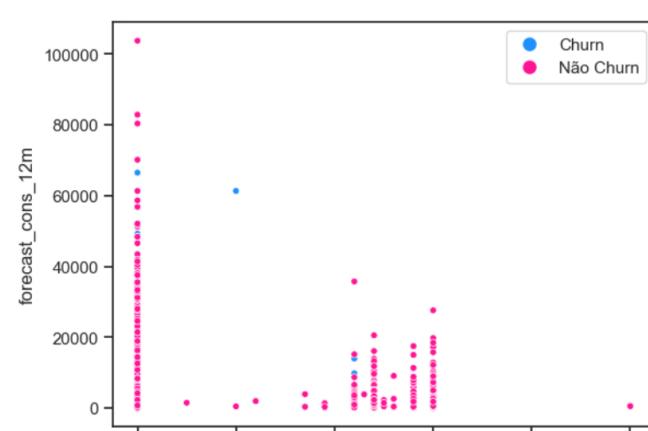
- Relação entre o valor do desconto previsto (forecast_discount_energy) e a previsão de consumo no próximo ano (forecast_cons_year)

Figura 11 - Relação entre o valor do desconto previsto (forecast_discount_energy) e a previsão de consumo no próximo ano (forecast_cons_year)

![Previsão de Desconto e Consumo Previsto 1](<https://github.com/2023M3T7-Inteli/grupo5/blob/main/assets/Previs%C3%A3o%20de%20Desconto%20e%20Consumo%20Previsto%201.png?raw=true>)

Fonte: Autoria própria

Figura 12 - Previsão de desconto e consumo previsto



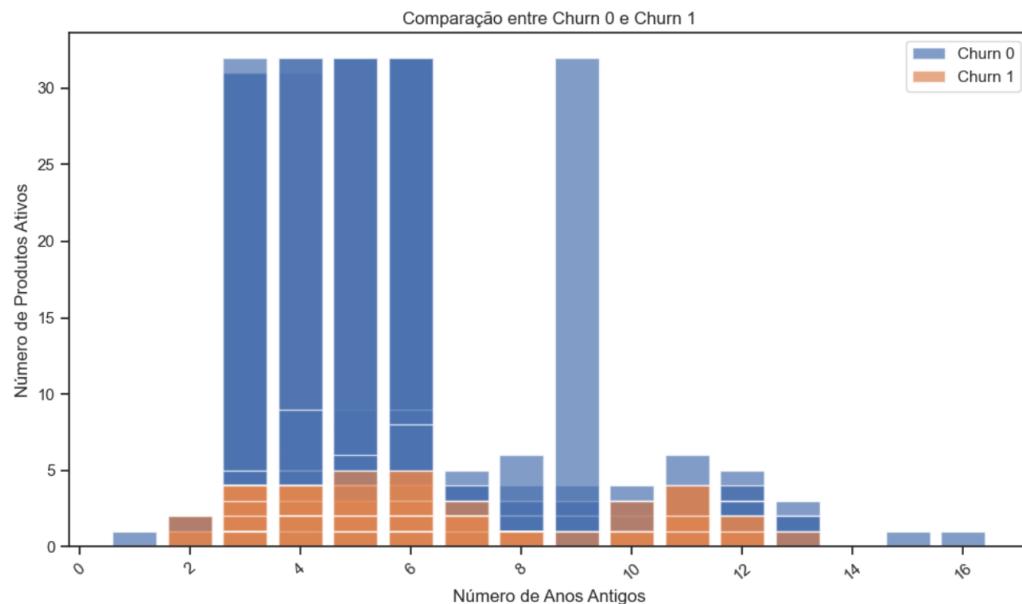


Fonte: Autoria própria

Esses gráficos revelam que, quando os clientes apresentam um consumo elevado, mas não desfrutam de descontos, têm uma tendência maior a realizar churn. Em contrapartida, os clientes que recebem descontos tendem a ter uma taxa menor de churn, como evidenciado pela concentração significativa de não churn, particularmente na faixa de desconto entre 20% e 30%.

- Relação entre clientes com maior antiguidade (num_years_antig) e número menor de produtos e serviços ativos (nb_prod_act)

Figura 13 - Número de Produtos Ativos e Antiguidade do Cliente



Fonte: Autoria própria

Esse gráfico evidencia uma tendência de churn notável entre clientes com uma antiguidade de 3 a 6 anos e um número limitado de produtos ativos. Além disso, é interessante notar que não há registros de churn entre clientes com uma antiguidade superior.

4.2.2. Pré-processamento dos dados

Processamento em relação às variáveis:

Em relação ao pré-processamento dos dados, iniciamos com a verificação dos dados repetidos de modo geral. Essa ação nos ajudou identificar quais colunas são categóricas (criptografadas ou não) ou numéricas. Após isso, separamos esses dados em colunas de forma que recebessem valores binários (1 ou 0) sempre que um respectivo ID fosse correspondente a um respectivo valor categórico. Em seguida, retiramos do nosso DataFrame todas as colunas que julgamos não serem importantes ou colunas que possuam muitos valores nulos, são elas: 'campaign_disc_ele', 'forecast_base_bill_ele', 'forecast_base_bill_year', 'forecast_bill_12m', 'forecast_cons', 'date_first_activ', 'has_gas', 'forecast_cons_year'. Entretanto, nosso time de desenvolvimento está trabalhando em análises para fazermos a repescagem de algumas colunas. Adicionamos também, colunas trimestrais e médias que nos fornecessem a média trimestral de preços de um determinado cliente. Dessa forma, conseguimos enxugar e reorganizar uma tabela de 200 mil linhas para 20 mil linhas. Ao final de todo o processamento de colunas, fizemos um "murch" da base de clientes com a base de preços e churn, agregando com a identificação da saída de cada cliente e nas médias trimestrais.

Processamento em relação aos valores:

Do lado mais numérico, nós transformamos todos os valores negativos em positivos. Havia muitos dados com essa característica, como datas e valores. Muitos dados nulos foram alterados pela média de suas colunas ou pelo valor 0, mas nunca deixados com o valor "NaN" (valor nulo). Em seguida, limitamos todos os valores com vírgula para terem somente duas casas decimais. Foi adicionado também no DataFrame dois campos chamados dias e diasrenovacao. Basicamente esses campos são a quantidade de dias do cliente dentro da empresa e a quantidade de dias para o fim do contrato. Por fim, preparamos os dados em ordem decrescente de churn, com o objetivo de facilitar o treinamento da IA e a nossa compreensão dos dados.

4.2.3. Hipóteses

Consumo de Energia Anterior e Atual: A diferença entre o consumo de eletricidade nos últimos 12 meses (cons_12m) e o consumo no último mês (cons_last_month) pode estar relacionada ao churn. Um aumento ou diminuição significativa no último mês pode indicar insatisfação ou mudança nas necessidades do cliente.

Previsão de Desconto e Consumo Previsto: A relação entre o valor do desconto previsto (forecast_discount_energy) e a previsão de consumo no próximo ano (forecast_cons_year) pode influenciar o churn. Clientes com altos descontos, mas com previsões de consumo elevadas, podem ser mais propensos a permanecer com a empresa.

Número de Produtos Ativos e Antiguidade do Cliente: Clientes com maior antiguidade (num_years_antig) e número menor de produtos e serviços ativos (nb_prod_act) podem ser mais propensos ao churn, indicando que a falta de engajamento com produtos adicionais pode resultar em menos fidelidade.

III - Preparação dos Dados e Modelagem

a) Modelagem para o problema (proposta de features com a explicação completa da linha de raciocínio).

Para a preparação do modelo, a princípio, removemos features com "type" data, pois já haviam a contagem de dias (subtração) para cada relação com datas. Logo, as features removidas foram: 'date_activ', 'date_end', 'date_modif_prod', 'date_renewal'. Removemos também as colunas de ID e Churn. A primeira é por motivos claros, os IDs dos clientes não trazem relevância para o treinamento, apenas atrapalhariam. Já a coluna de 'Churn' é por ser a nossa variável de análise, por isso removemos. Em relação às nossas features de treinamento, escolhemos as seguintes features: 'consumption_p1', 'consumption_p2', 'consumption_p3', 'activity_new', 'channel_sales', 'cons_12m', 'cons_gas_12m', 'cons_last_month', 'forecast_cons_12m', 'forecast_cons_year', 'forecast_discount_energy', 'forecast_meter_rent_12m', 'forecast_price_energy_p1', 'forecast_price_energy_p2', 'forecast_price_pow_p1', 'imp_cons', 'margin_gross_pow_ele', 'margin_net_pow_ele', 'nb_prod_act', 'net_margin', 'num_years_antig', 'pow_max', 'contract_tenure', 'renovation_tenure', 'contract_tenure_year', 'bill_dev', 'cons_pattern'.

```
df.fillna(0,inplace=True)
y = df["churn"]
X = df.drop(labels = ["id", "date_activ", "date_end", "date_modif_prod", "date_renewal", "churn"],axis = 1)
```

Em seguida, utilizamos o método/cláusula "OneHotEncoder" da biblioteca Sklearn. Com o "OneHotEncoder" transformamos todos os dados categóricos em colunas que recebem valores binários (1 ou 0) de acordo com seu dado categórico.

```
from sklearn.preprocessing import LabelEncoder

label_encoder2 = LabelEncoder()
df['channel_sales'] = label_encoder2.fit_transform(df['channel_sales'])
df['activity_new'] = label_encoder2.fit_transform(df['activity_new'])
df['origin_up'] = label_encoder2.fit_transform(df['origin_up'])
df['has_gas'] = df['has_gas'].replace( {'f':0, 't':1} )
```

Fizemos esse processo, também chamado de "Label Encoding", nas seguintes features: activity_new, channel_sales e origin_up, totalizando mais de 38 colunas (essas 38 colunas aumentam quando o código entra no processo de modelagem e treinamento da IA). Utilizamos também, da biblioteca imblearn, o método SMOTE. Com ele, basicamente, aumentamos o tamanho do nosso DataFrame para uma melhor compreensão do treinamento da IA. Em nosso caso, tínhamos em torno de 10% de dados avaliados com Churn (1). Dessa forma, o SMOTE cria dados de Churn = 1, baseados com os antigos dados já treinados e aplica-os no nosso DataFrame principal.

```
smote = SMOTE(random_state=42)
X_resampled, y_resampled = smote.fit_resample(X, y)
```

Em seguida, dividimos os dados de treinamento e teste para o processo. Mantivemos uma proporção de 80/20. 80% dos dados para treinamento, e os outros 20% para teste e validação. Após a separação do conjunto de dados, iniciamos o modelo escolhido, o Random Forest (explicação do modelo com mais detalhes nos próximos parágrafos). E por fim, realizamos o treinamento do modelo com todas features hiperparâmetros escolhidos.

```
X_train, X_test, y_train, y_test = train_test_split(X_resampled, y_resampled, test_size=0.2, random_state=42, stratify=y_resampled)

model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
```

b) Métricas relacionadas ao modelo (conjunto de testes, pelo menos 3).

Consideremos o churn como classe positiva nas seguintes métricas:

A matriz de confusão mostra a frequência das classificações das diferentes classes avaliadas, permitindo uma análise mais completa do desempenho do modelo. Ela mostra as classificações e compara se estavam corretas se comparado ao valor real. A matriz de confusão mostra a frequência das classificações de churn e não churn para cada modelo. No projeto, temos os seguintes casos:

True Positive (TP): o modelo previu corretamente que (%) clientes cancelariam o contrato.

False Positive(FP): (%) não cancelou mas o modelo previu que cancelaria.

True Negative(TN): o modelo previu corretamente que (%) não cancelaria o contrato.

False Negative(FN): (%) cancelou mas o modelo previu que não cancelaria. Dessa forma, aplicando a matriz de confusão no nosso modelo, conseguimos obter os seguintes resultados:

Figura 14 - Matriz de Confusão





Fonte: Autoria própria

É possível notar um possível sobreajuste, esse fator foi levado em consideração e o modelo será aprimorado com a utilização de hiperparâmetros. Assim, espera-se encontrar resultados ainda mais satisfatórios em uma próxima avaliação.

Precision

Como o próprio nome já diz, busca identificar o nível de precisão do modelo. Calcula o número de exemplos classificados como pertencentes a uma classe, que realmente são daquela classe. É obtido através da seguinte fórmula:

$$\text{Precision} = \frac{TP}{TP + FP}$$

No modelo preditivo tratado neste documento, a precisão será calculada através do número de chuns previstos corretamente, dividido pelo número de chuns previstos corretamente somado ao número de chuns previstos incorretamente.

Recall

Busca identificar o nível de acurácia do modelo para prever positivos. Calcula o número de exemplos classificados como pertencentes a uma classe, que realmente são daquela classe, dividido pela quantidade total de exemplos que pertencem a esta classe, mesmo que sejam classificados em outra. É obtido através da seguinte fórmula:

$$\text{Recall} = \frac{TP}{TP + FN}$$

No modelo preditivo tratado neste documento, a precisão será calculada através do número de chuns previstos corretamente, dividido pelo número de chuns previstos corretamente somado ao número de não chuns previstos incorretamente.

F1-score

Ao considerar o F1-score como métrica, avaliamos a relação entre a precisão(precision) e a revocação(recall). O F1-score é dada por

$$2 \times \frac{\text{precisão} \times \text{revocação}}{\text{precisão} + \text{revocação}}$$

e mostra um balanço (média harmônica) das 2 métricas. Por considerar outras 2 métricas, oferece um modo mais rico de medir o sucesso do modelo.

ROC

A curva ROC(Receiver Operating Characteristic) mostra o quão bom o modelo é em distinguir entre duas classes, no nosso caso, distinguir quem deu churn e quem não deu. É usada para modelos de classificação e leva em conta dois parâmetros:

- Taxa de verdadeiro positivo = verdadeiros positivos / (verdadeiros positivos + falsos negativos)
- Taxa de falso positivo = falso positivos / (falso positivos + verdadeiros negativos)

AUC

A AUC(Area Under the Curve) avalia a área debaixo da curva ROC. Varia de 0 a 1, com limiar no 0,5, isto é, acima desse limite, o algoritmo classifica em uma classe e abaixo na outra classe. Quanto maior é a AUC, melhor é o modelo.

Unindo essas métricas de avaliação e aplicando elas em nosso modelo, obtivemos os seguintes resultados:

c) Apresentar o primeiro modelo candidato, e uma discussão sobre os resultados deste modelo (discussão sobre as métricas para esse modelo candidato).

O primeiro modelo que consideramos foi o Random Forest, escolhido devido à sua eficácia na previsão de churn e aos resultados satisfatórios nas métricas de avaliação. No entanto, identificamos um problema de overfitting, o que sugere que o modelo pode estar fazendo previsões excessivamente otimistas, já que a matriz de confusão está equilibrada. Mesmo assim, em comparação com outros modelos, o Random Forest se destacou em nossas análises.

O modelo Random Forest combina as saídas de várias árvores de decisão para produzir uma única previsão e permite a utilização de diversos tipos de dados para classificação. Por ser o somatório de diversas árvores, ele é capaz de reduzir o overfitting e oferece maior flexibilidade. No projeto, utilizamos as seguintes características (colunas) como entrada:

```
'price_date', 'price_p1_var', 'price_p2_var', 'price_p3_var', 'price_p1_fix', 'price_p2_fix', 'price_p3_fix', 'consumption_p1', 'consumption_p2', 'consumption_p3', 'activity_new', 'channel_sales', 'cons_12m', 'cons_gas_12m', 'cons_last_month', 'date_activ', 'date_end', 'date_modif_prod', 'date_renewal', 'forecast_cons_12m', 'forecast_cons_year', 'forecast_discount_energy', 'forecast_meter_rent_12m', 'forecast_price_energy_p1', 'forecast_price_energy_p2', 'forecast_price_pow_p1', 'has_gas', 'imp_cons', 'margin_gross_pow_ele', 'margin_net_pow_ele', 'nb_prod_act', 'net_margin', 'num_years_antig', 'origin_up', 'pow_max', 'churn', 'contract_tenure', 'renovation_tenure', 'contract_tenure_year', 'bill_dev', 'cons_pattern'.
```

Apesar dos desafios encontrados com o overfitting, o modelo Random Forest se destacou como a melhor opção em nossa análise, dada a diversidade e a quantidade de recursos disponíveis para a previsão de churn. No entanto, é importante continuarmos a refinar o modelo para melhorar seu desempenho e confiabilidade.

4.4. Comparação de Modelos

O problema apresentado é de classificação, visando prever a saída de clientes de uma empresa, abordado com um modelo supervisionado.. Considerando as características apresentadas, foram escolhidas como métrica as curvas AUC-ROC e F1-Score.

Visto que acurácia, precisão e revocação por si só não fornecem insights tão abrangentes sobre a qualidade do modelo, escolhemos o F1-Score, capaz de fornecer uma média entre precisão e revocação e usamos AUC-ROC, visto que não é afetada pela distribuição desigual das classes, como é o caso do dataset (com uma proporção de 1:9). Ambos fornecem um meio mais objetivo de comparar modelos preditivos.

A fim de melhorar as métricas, usamos o GridSearch e o RandomSearch. O primeiro testa todas as possibilidades de combinação de hiperparâmetros e o segundo testa algumas combinações aleatoriamente. Devido ao custo computacional relacionado ao GridSearch, aplicamos apenas ao modelo com menor quantidade de hiperparâmetros, o KNeighborsClassifier. No RandomForest e no XGB, aplicamos o RandomSearch. Veja a seguir os resultados encontrados através dessas buscas e as métricas.

K-NN (K-Nearest Neighbors): O K-NN, ou K-Nearest Neighbors, é um algoritmo de aprendizado supervisionado que faz previsões com base na proximidade de pontos, onde cada ponto é um dado. Especificamente, ele identifica os "K" pontos mais próximos a um ponto de teste e atribui à classe predominante ou a média das respostas, considerando a quantidade de pontos escolhidos. Nesse modelo, depois de aplicar o GridSearch, escolhemos os seguintes hiperparâmetros:

'metric': 'manhattan': define a métrica de distância que o KNN usará para calcular a proximidade entre os pontos de dados. No caso, manhattan significa que o algoritmo calculará a distância usando a soma das diferenças absolutas nas coordenadas (eixos) entre os pontos de dados.

'n_neighbors': 11: determina o número de vizinhos mais próximos a serem considerados ao fazer uma previsão para um novo ponto de dados. O algoritmo KNN considerará os 5 vizinhos mais próximos para tomar uma decisão de classificação ou regressão para um novo ponto de dados.

'weights': 'uniform': O hiperparâmetro 'weights' determina como os vizinhos contribuem para a decisão. Quando definido como 'uniform', indica que todos os pontos tem a mesma influência.

Figura 15 - Métricas de treino do KNN

	precision	recall	f1-score	support
0	1.00	1.00	1.00	10577
1	1.00	1.00	1.00	10577
accuracy			1.00	21154
macro avg	1.00	1.00	1.00	21154
weighted avg	1.00	1.00	1.00	21154

Fonte: Autoria própria

Figura 16 - Métricas de teste do KNN

	precision	recall	f1-score	support
0	0.90	0.62	0.73	2628
1	0.11	0.40	0.17	305
accuracy			0.60	2933
macro avg	0.50	0.51	0.45	2933
weighted avg	0.82	0.60	0.67	2933

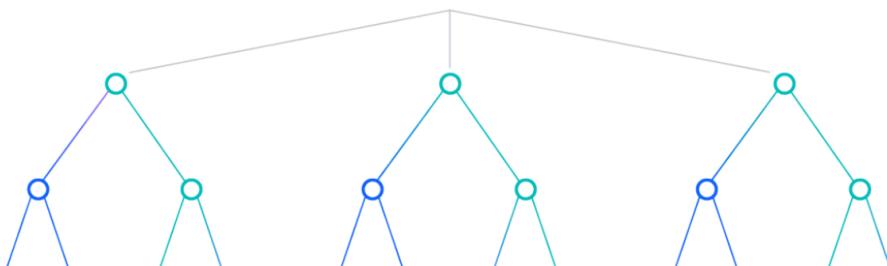
Fonte: Autoria própria

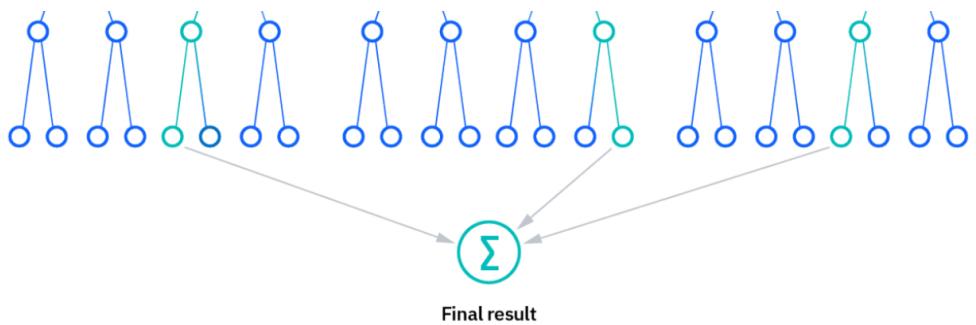
AUC-ROC: 0.50

Atingimos um resultado que não foi muito satisfatório e continuamos testando. Além disso, encontramos overfitting, problema que foi melhor solucionado com o uso do XGBoost.

Random Forest: O Random Forest é um algoritmo de aprendizado de máquina que opera por meio de um conjunto de árvores de decisão. Cada árvore é gerada utilizando um subconjunto aleatoriamente selecionado de atributos e registros do conjunto de dados, uma abordagem caracterizada como "bootstrap aggregating" ou simplesmente "bagging". Durante a fase de predição, o algoritmo submete uma entrada a todas as árvores e coleta suas decisões. A classe (em problemas de classificação) ou valor médio (em problemas de regressão) mais votada pelas árvores é então escolhida como a saída final. Este método de agregação reduz a variância, minimizando o risco de overfitting e otimizando a capacidade de generalização em conjuntos de dados não observados durante o treinamento.

Figura 17 - RandomForest modelo





Fonte: ibm.com

Ao ajustar o modelo para lidar com o desequilíbrio nas classes, foram escolhidos os seguintes hiperparâmetros:

`n_estimators = 50`: Define quantas árvores de decisão serão usadas no modelo.

`max_depth = 10`: Controla a profundidade máxima das árvores, ou seja, os "galhos".

`min_samples_split = 20`: Define o valor mínimo para um grupo ser considerado uma amostra.

`min_samples_leaf = 10`: Determina o menor número de amostras requerido em um nó terminal, chamado de folha.

`class_weight = 'balanced'`: Indica o peso de cada classe no treinamento do modelo. Nesse caso, significa que o modelo deve balancear as classes.

Figura 18 - Métricas do conjunto de treino do RandomForest

	precision	recall	f1-score	support
0	0.73	0.50	0.59	10577
1	0.62	0.81	0.70	10577
accuracy			0.66	21154
macro avg	0.67	0.66	0.65	21154
weighted avg	0.67	0.66	0.65	21154

Fonte: Autoria própria

AUC-ROC: 0.58

Figura 19 - Métricas do conjunto de teste do RandomForest

	precision	recall	f1-score	support
0	0.90	0.43	0.59	2628
1	0.11	0.58	0.18	305
accuracy			0.45	2933
macro avg	0.50	0.51	0.38	2933
weighted avg	0.82	0.45	0.54	2933

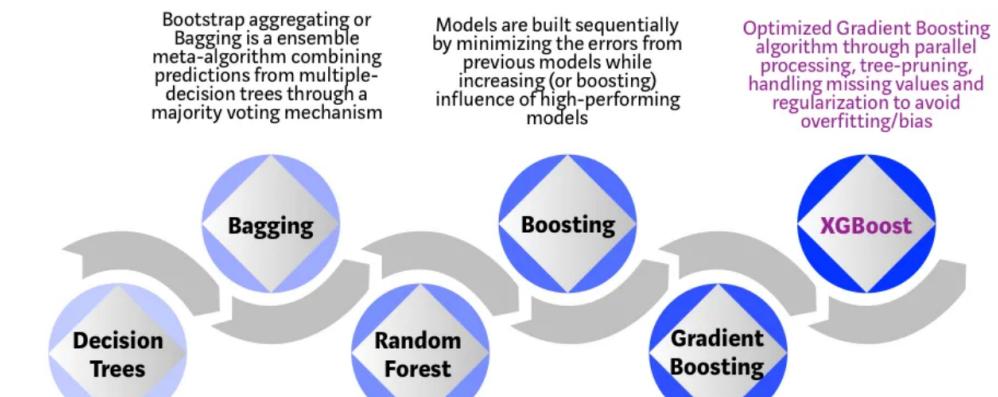
Fonte: Autoria própria

AUC-ROC: 0.50

XGB(Extreme Gradient Boost): O XGBoost, ou eXtreme Gradient Boosting, é uma técnica que busca construir, de forma iterativa, uma série de árvores de decisão, onde cada uma tenta corrigir os erros das anteriores, otimizando uma função de perda. Este processo é executado até que um critério de parada seja alcançado.

A imagem a seguir demonstra um pouco a diferença entre o XGB e o RandomForest:

Figura 20 - Evolução dos algoritmos baseados em árvores



A graphical representation of possible solutions to a decision based on certain conditions

Bagging-based algorithm where only a subset of features are selected at random to build a forest or collection of decision trees

Gradient Boosting employs gradient descent algorithm to minimize errors in sequential models

Fonte: towardsdatascience.com

A seguir os hiperparâmetros e os valores escolhidos para cada um:

objective='binary: logitraw': define a função de objetivo usada pelo algoritmo. Neste caso, 'binary:logitraw' indica que o modelo é de classificação binária e a função de objetivo é o logit crú (ou seja, a saída antes da aplicação da função logística).

learning_rate=0.1: define a taxa na qual o modelo aprende durante o treinamento. Um valor baixo, como 0.1, significa que o modelo fará atualizações menores nos pesos a cada iteração.

max_depth=5: define a profundidade máxima da árvore de decisão. Árvores mais profundas podem aprender relações mais complexas nos dados, mas aumentam o risco de overfitting (ajuste excessivo) nos dados de treinamento.

n_estimators=500: é número de árvores de decisão que serão construídas durante o treinamento do modelo.

n_jobs=-1: define quantos núcleos de CPU serão usados durante o treinamento. No caso, -1 significa usar todos os núcleos disponíveis, o que ajuda a acelerar significativamente o treinamento.

Figura 21 - Métricas do conjunto de treino do XGB

	precision	recall	f1-score	support
0	0.73	0.50	0.59	10577
1	0.62	0.81	0.70	10577
accuracy			0.66	21154
macro avg	0.67	0.66	0.65	21154
weighted avg	0.67	0.66	0.65	21154

Fonte: Autoria própria

AUC-ROC: 0.97

Figura 22 - Métricas do conjunto de teste do XGB

	precision	recall	f1-score	support
0	0.90	0.43	0.59	2628
1	0.11	0.58	0.18	305
accuracy			0.45	2933
macro avg	0.50	0.51	0.38	2933
weighted avg	0.82	0.45	0.54	2933

Fonte: Autoria própria

AUC-ROC: 0.62

Considerando os resultados observados, o modelo escolhido foi o XGBoost, por apresentar melhores resultados em comparação com os outros modelos avaliados. Isto é, apresentou um F1-score mais elevado e, de forma particularmente significativa, em sua AUC_ROC de 0.72 durante o teste. Além disso, o XGBoost compartilha algumas características valiosas com outros algoritmos, como processamento paralelo, bagging e boosting, proporcionando uma avaliação mais dinâmica e abrangente.

4.5. Avaliação ↗

Iniciamos com a proposta de desenvolver um modelo preditivo para identificar o churn de clientes da Fortum, uma empresa finlandesa de energia elétrica. Para abordar esse desafio, optamos por criar um modelo de classificação, dada a natureza binária do problema (churn ou não churn).

Com esse objetivo em mente, conduzimos uma análise detalhada do comportamento dos clientes, com a finalidade de antecipar suas ações futuras. A eficácia desse modelo tem o potencial de otimizar significativamente a receita da empresa e possibilitar o desenvolvimento de estratégias direcionadas à retenção de clientes, incluindo a criação de campanhas de marketing específicas e eficazes. Esse processo é fundamental para fortalecer o relacionamento com os clientes e maximizar o valor de longo prazo que eles representam para a organização.

Optamos pelo modelo XGBoost devido ao seu desempenho superior em termos de métricas e à sua capacidade de lidar eficazmente com grandes volumes de dados, alta dimensionalidade e valores ausentes. O XGBoost é amplamente reconhecido por sua precisão, resistência ao overfitting e suporte a diversas funções de perda, tornando-o uma escolha sólida em uma variedade de cenários de aprendizado de máquina. As métricas obtidas confirmaram a eficácia desse modelo para o nosso contexto específico.

Para utilizar este modelo, é fundamental compreender as métricas que são geradas como resultado. Portanto, quando o modelo fornece um resultado, é essencial avaliar as métricas de forma abrangente e buscar consistência para obter insights válidos. Para enfrentar possíveis desafios e falhas, é crucial contar com um plano de contingência. Nossas recomendações incluem:

Treinamento: É imprescindível que as equipes envolvidas na utilização do modelo passem por treinamento prévio. Isso permitirá que

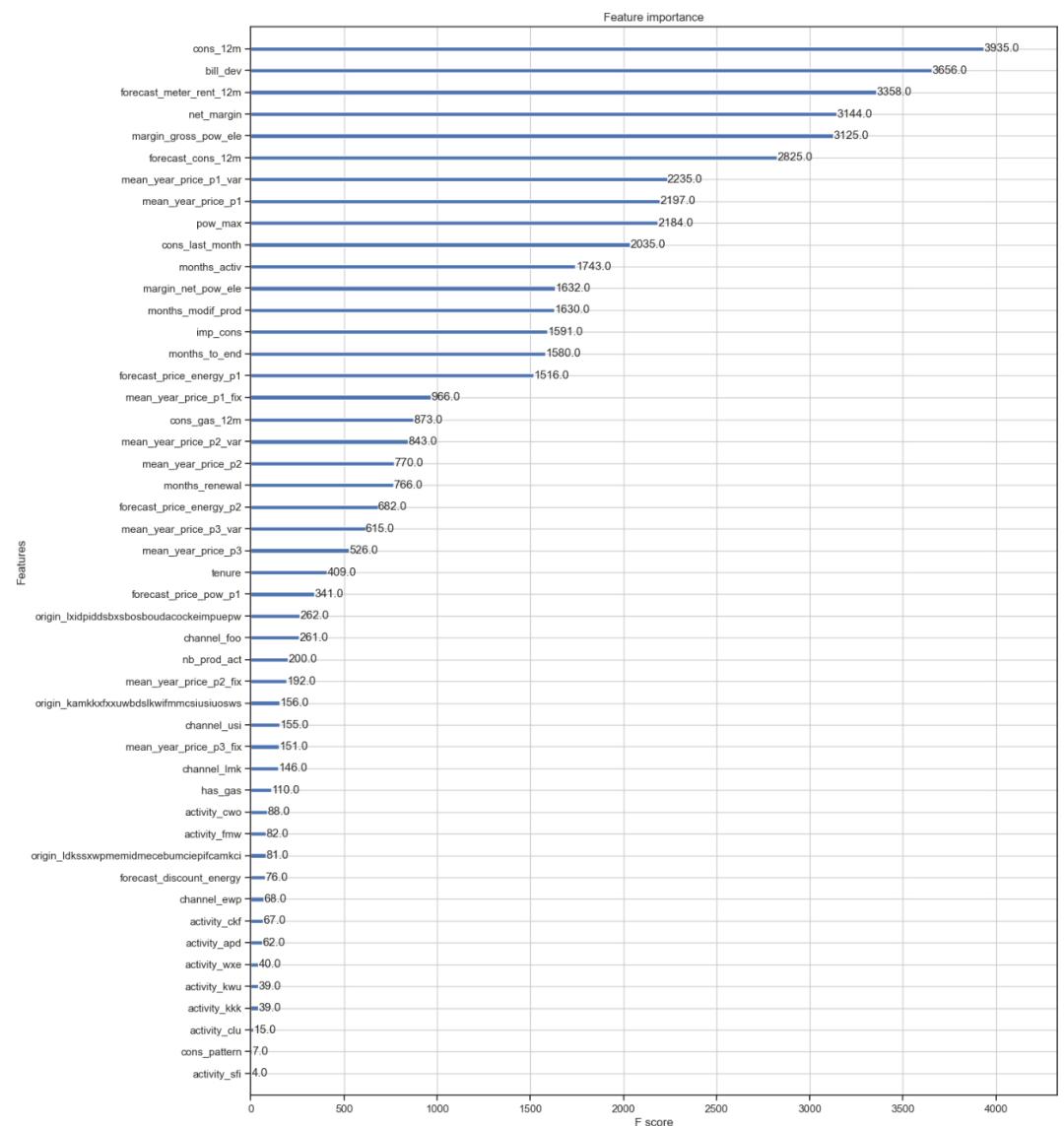
compreendam como emprega-lo adequadamente e desenvolvam um senso crítico para identificar erros sutis, como altas taxas de acurácia que não refletem um desempenho eficaz.

Governança: A governança de dados desempenha um papel fundamental, assegurando uma tomada de decisão eficaz e ágil, especialmente quando se trata de lidar com situações de erro. Além disso, garante a segurança dos dados envolvidos.

Comitê de Crise: A criação de um comitê de crise é fundamental para abordar situações desafiadoras de maneira proativa. Este comitê é responsável por agir assertivamente em tais casos, garantindo a eficácia das medidas adotadas.

Com base nos gráficos e análises realizadas, as hipóteses mencionadas anteriormente neste documento em relação ao consumo de energia anterior e atual foram validadas. A análise estatística revelou que a média da diferença é significativamente maior para os clientes que realizaram churn em comparação com aqueles que não o fizeram. Além disso, o teste de hipótese confirmou que essa diferença é numericamente significativa, indicando que a variação abrupta no consumo é um preditor relevante de churn. Esse resultado pode ser ainda mais evidenciado por meio do gráfico gerado de feature importance, no qual a variável 'cons_12m' aparece como a mais importante em influenciar o churn.

Figura 23 - Gráfico de Feature Importance



5. Conclusões e Recomendações

Os resultados que obtivemos com o desenvolvimento do nosso projeto foram: com nossa calibração, conseguimos 65% na métrica AUC-ROC - algo relativamente bom para um DataFrame desbalanceado como o de Churn na empresa PowerCO. De forma mais detalhada, alcançamos as métricas esperadas e conseguimos desenvolver um algoritmo capaz de contribuir para a análise e previsão da saída de clientes na PowerCO.

Em relação ao uso do modelo, é possível carregá-lo com ajuda da biblioteca do Joblib. Dessa forma, o modelo estará pronto para receber inputs novos e fazer suas previsões usando novos datasets. A seguir uma lista que preparamos com alternativas para incrementá-lo na empresa junto ao time de marketing :

- Analisar, junto os times envolvidos, as principais features que impactam a rotação de clientes na empresa.
- Analisar quais relações e manipulações podem afetar a rotação desses clientes.
- Analisar, quais descontos são relevantes, de forma individual ou generalizada, para a rotação de clientes na empresa.

Acreditamos que, com essas medidas, nosso modelo terá uma ótima capacidade de trabalhar e desenvolver seu papel na prática com

maestria.

Além disso, é muito importante destacarmos que nosso modelo tem como objetivo não prever a saída de um tipo específico de clientes, mas sim de clientes sem características específicas - clientes que o modelo julgar propícios a darem Churn ou não. E como indicação de recomendações a serem tomadas a fim de evitar a rotação desses clientes, a seguir, listamos algumas estratégias que podem ser tomadas com base em análises estatísticas e de relações humanas:

Análise estatística:

- Preço da energia por Kw/h
- Potência disponibilizada por contrato.
- Notificar os clientes em relação ao consumo mensal e anual.
- Fornecer medidas de apoio aos clientes que não geram a maior margem de lucro líquida.
- Alertar os clientes quando o contrato estiver perto de acabar. Observação: O gráfico utilizado para fazer as medidas de apoio com base na análise estatística é o mesmo da Figura 23 - Gráfico de Feature Importance.

Observação: As recomendações com base na análise estatística foi desenvolvida de acordo com a importância de cada Feature para o modelo. Dessa forma, conseguimos calcular qual Feature de fato pode ser inovada e reavaliada por parte da empresa. Entretanto, ressaltamos que as melhores medidas de apoio são fornecidas com base nas relações humanas.

Relações humanas:

- Melhorar a experiência do usuário na venda e na experimentação do serviço;
- Fornecer atendimento personalizado para cada cliente;
- Inovar a qualidade do produto com o auxílio de tecnologias;
- Treinamento e capacitação da equipe de atendimento ao cliente;
- Monitorar constantemente e antecipar sinais de Churn na empresa.

6. Referências

- ALBERTACCI, Laura. *Devo me preocupar com a LGPD?* Jus.com.br. Disponível em: <https://jus.com.br/artigos/91636/devo-me-preocupar-com-a-lgpd>. Acesso em: 12 set. 2023.
- AFFONSO, Aline. *Você sabe o que é a metodologia CRISP DM?* Blog Voitto. Disponível em: <https://www.voitto.com.br/blog/artigo/crisp-dm>. Acesso em: 11 set. 2023.
- CARDOSO, Bruno. 5 forças de Porter: modelo, análise, como aplicar e muito mais!. Agências de Resultados, 22 fev. de 2022. Disponível em: <https://resultadosdigitais.com.br/agencias/as-cinco-forcas-de-porter/>. Acesso em: 8 ago. 2023.
- Fortumin evästekäytäntö. Disponível em: <https://www.fortum.fi/tietosuoja/evasteet>. Acesso em: 07 ago. 2023.
- FORTUM. HENKILÖTETOJEN SÄILYTYSAJAT / FORTUM (FINLAND). [s.l.: s.n.]. Disponível em: <https://www.fortum.fi/media/21635/download?attachment=1>. Acesso em: 07 ago. 2023.
- HARLEY, Aurora. *Personas Make Users Memorable for Product Team Members*. Nielsen Norman Groupo. Disponível em: <https://www.nngroup.com/articles/persona/>.
- IBM. *Visão geral da ajuda do CRISP-DM*. Disponível em: <https://www.ibm.com/docs/pt-br/spss-modeler/18.4.0?topic=dm-crisp-help-overview>. acesso em: 11 set. 2023.
- MAGRETTA, Joan, *Entendendo Michael Porter*, Rio de Janeiro : Alta Books, 2019.
- NAPOLEÃO, Bianca. *Matriz de Riscos (Matriz de Probabilidade e Impacto)*, Ferramentas da Qualidade. Disponível em: <https://ferramentasdqualidade.org/matriz-de-riscos-matriz-de-probabilidade-e-impacto/>.
- OSTERWALDER, Alexander; PIGNEUR, Yves; BERNARDA, Gregory. *Value Proposition Design: Como Construir Propostas de Valor Inovadoras*. Rio de Janeiro: Editora Alta Books, 2014.
- OSTERWALDER, Alexander ; YVES PIGNEUR. *Business model generation : inovação em modelos de negócios : um manual para visionários, inovadores e revolucionários*. Rio De Janeiro: Alta Books, 2011.
- Powering a thriving world. Disponível em: <https://www.fortum.com/>. Acesso em: 8 ago. 2023.
- VISHAL MORDE. XGBoost Algorithm: Long May She Reign! Medium. Disponível em: <https://towardsdatascience.com/https-medium-com-vishalmorde-xgboost-algorithm-long-she-may-rein-eddf99be63d>.

Anexos

Lean Inception:

https://miro.com/welcomeonboard/b1NuelBGRkFleXp2R1RuOVBlcDl3dzBkQWNwYmFQY0jqUTR4YnhPcXNJJaDk5eUladGVPQ0Y0WEjOVUzbWg0QnwzNDU4NzY0NTUxOTkxNjl1ODk2fDI=?share_link_id=173781648717