

Allef Santos RM: 550341

Cássio Yuji..... RM: 551491

Débora Damaso..... RM: 97836

Paulo Barbosa RM: 550323

Yasmin Lopes..... RM: 552314

CHALLENGE



“Desvendando o Potencial dos Dados: Explorando a AI Generativa e o Deep Analytics
para Impulsionar o Crescimento Empresarial”

Documentação Geral

São Paulo - SP

2024

Sumário

1 - INTRODUÇÃO	3
1-2 Justificativa	3
2 – DEFINIÇÃO DO PROBLEMA.....	4
3 – PROJETO E OBJETIVOS.....	4
4 – PROTÓTIPO MOBILE DA APLICAÇÃO	5
5 – DEMONSTRAÇÃO DE TELA WEB EM C#	8
6 – CONSTRUÇÃO DO MODELO E APLICAÇÃO DE MACHINE LEARNING	9
6-1 CONCEITO DE PROCESSAMENTO DE LINGUAGEM NATURAL (NLP).....	9
6-2 APLICAÇÃO DE PROCESSAMENTO DE LINGUAGEM NATURAL (NLP)	10
6-3 BIBLIOTECAS / FRAMEWORKS UTILIZADOS.....	11
6-4 FERRAMENTAS ADICIONAIS.....	12
7 – FERRAMENTAS DOCKER	12
7-1 RECURSOS DA AZURE	13
8 – DIAGRAMA DA ARQUITETURA (ARCHIMATE)	14
9 – CONCLUSÃO	15

1 - INTRODUÇÃO

Nos últimos anos, a explosão de dados tem sido uma realidade inegável para as empresas em todos os setores. No entanto, o verdadeiro desafio não reside apenas na coleta massiva desses dados, mas sim em extrair insights valiosos que possam impulsionar o crescimento empresarial. É aqui que entram em cena duas tecnologias revolucionárias: Inteligência Artificial (IA) Generativa e Deep Analytics.

A IA Generativa, um ramo da inteligência artificial que se concentra na criação de novos dados, imagens, textos e até mesmo música, oferece um novo paradigma para a inovação. Por meio de algoritmos avançados, como redes neurais generativas adversariais (GANs), as empresas podem não apenas analisar dados existentes, mas também gerar novos dados sintéticos que podem ser utilizados para melhorar produtos, personalizar experiências do cliente e até mesmo prever tendências de mercado.

Por outro lado, o Deep Analytics vai além da análise superficial de dados e mergulha nas camadas mais profundas para desvendar insights significativos. Utilizando técnicas de machine learning avançadas, como redes neurais profundas e algoritmos de aprendizado não supervisionado, o Deep Analytics é capaz de identificar padrões complexos, correlações sutis e anomalias ocultas nos dados empresariais. Isso permite que as organizações tomem decisões mais informadas, otimizem processos internos e identifiquem novas oportunidades de negócios.

Ao combinar IA Generativa e Deep Analytics, as empresas podem criar um ciclo virtuoso de inovação e crescimento. Por exemplo, uma empresa de varejo pode usar IA Generativa para criar simulações de comportamento de compra de clientes e, em seguida, aplicar Deep Analytics para identificar os drivers subjacentes desse comportamento. Com esses insights, a empresa pode desenvolver estratégias de marketing mais eficazes, melhorar a segmentação de clientes e otimizar o mix de produtos, tudo levando a um aumento tangível nas receitas e lucros.

No entanto, é importante notar que, embora o potencial dessas tecnologias seja imenso, elas também trazem consigo desafios únicos, como questões éticas relacionadas à privacidade dos dados e viés algorítmico. Portanto, as empresas devem adotar uma abordagem responsável ao implementar IA Generativa e Deep Analytics, garantindo que os benefícios sejam maximizados enquanto os riscos são mitigados.

Em suma, o futuro do crescimento empresarial está intrinsecamente ligado à capacidade das organizações de desvendar o potencial dos dados por meio da IA Generativa e do Deep Analytics. Ao abraçar essas tecnologias de maneira estratégica e ética, as empresas podem posicionar-se na vanguarda da inovação e prosperar em um mundo cada vez mais orientado por dados.

1-2 Justificativa

Nosso objetivo é dar uma solução de alcance maximizado, ou seja, que possa ser utilizada por todos os tipos de clientes que a Plusoft presta serviços e não somente em um nicho específico.

Com isso, há diversos benefícios e não somente a Plusoft, mas também com os clientes e seus consumidores, já que as empresas terão um mapeamento sentimental dos compradores aos adquirirem um produto ou serviço.

2 – DEFINIÇÃO DO PROBLEMA

A coleta e análise de feedback são cruciais para o sucesso de qualquer empresa. No entanto, diversos desafios podem dificultar esse processo, impedindo a organização de alcançar todo o potencial do feedback do cliente:

- **Feedbacks Ausentes:** Uma grande quantidade de produtos com feedbacks ausentes indica baixo engajamento do consumidor. Isso torna difícil identificar problemas e oportunidades de melhoria, além de limitar a capacidade da empresa de capitalizar os aspectos positivos dos produtos.
- **Análise Complexa:** A vasta quantidade de dados coletados, tanto estruturados quanto não estruturados, torna desafiador extrair insights relevantes e acionáveis. A falta de expertise em análise de dados dificulta a identificação dos motivos e tópicos subjacentes aos feedbacks positivos e negativos.
- **Demanda por agilidade:** A demanda por soluções rápidas exige a capacidade de identificar feedbacks recentes, sinais de alerta e oportunidades em tempo real. A dificuldade nesse aspecto pode levar à perda de chances de fidelizar clientes e aprimorar a experiência do usuário.

Esses desafios podem ter consequências negativas para a empresa, como dificuldade na identificação e resolução de problemas, perda de oportunidades de melhoria, insatisfação do cliente, dificuldade na fidelização de clientes, desvantagens competitivas.

3 – PROJETO E OBJETIVOS

O projeto visa implementar um sistema avançado de processamento de dados para análise sentimental dos feedbacks dos clientes. O foco é duplo: avaliar tópicos relevantes nas opiniões dos usuários que já forneceram seu parecer sobre produtos ou serviços e também engajar aqueles que ainda não o fizeram. O objetivo é extrair insights valiosos que permitam às empresas compreender as razões por trás das avaliações — sejam elas positivas, negativas, neutras ou ausentes.

Será desenvolvido um mecanismo de incentivos (cashbacks, descontos) para encorajar os consumidores a adotarem o hábito de avaliar suas experiências de compra, reduzir o número de avaliações ausentes e incentivar novas compras. Isso se baseia na premissa de que as avaliações de consumidores anteriores são um elemento chave na decisão de novos clientes de adquirir um produto.

Para atingir esses objetivos, usaremos recursos de programação das áreas de Front-End, Back-End, Machine Learning, além de a infraestrutura para que a aplicação possa funcionar de maneira mais segura.

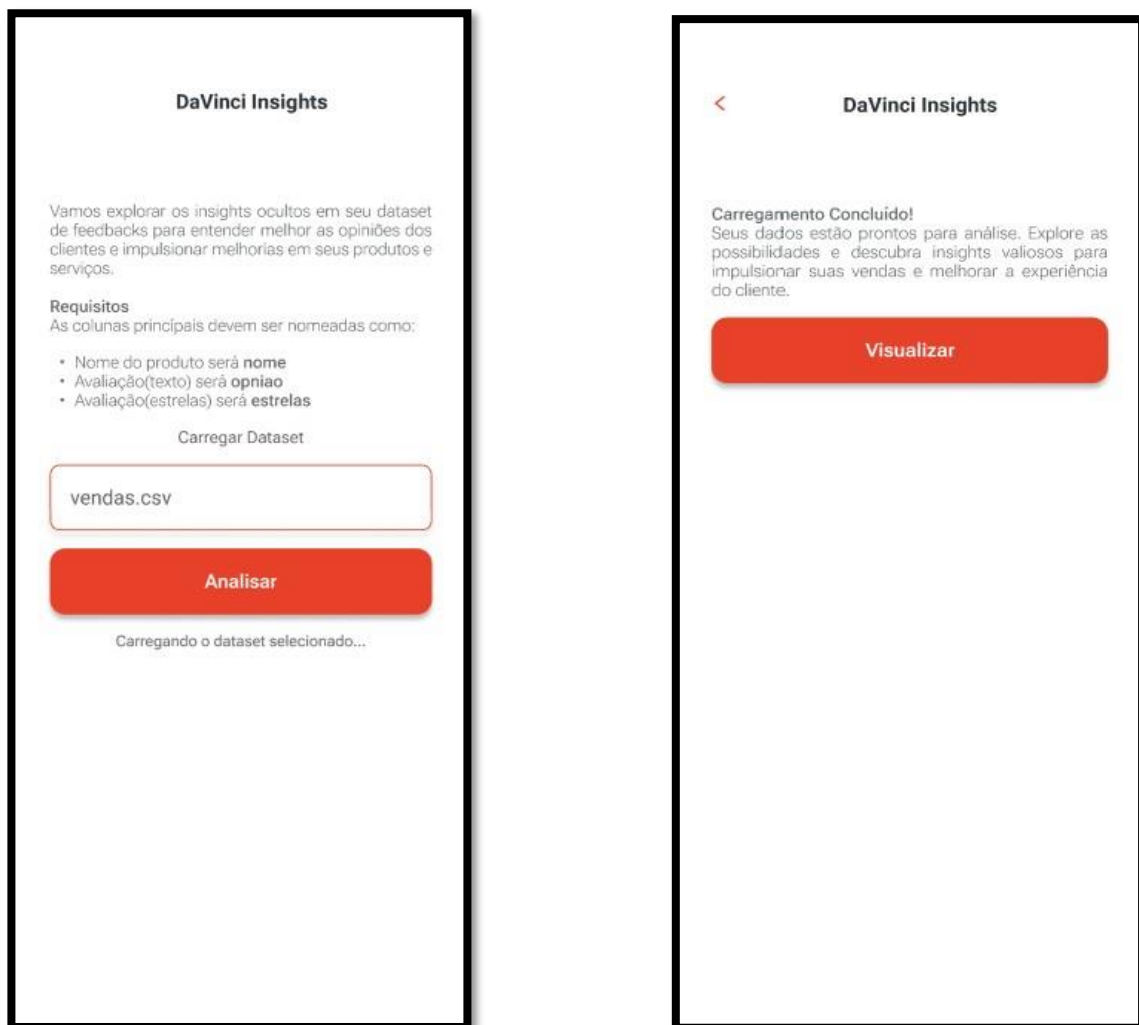
Nossa solução irá iniciar com a entrada do DataSet da empresa em nossa aplicação e posteriormente, passar pelo processo de Data Analytics com todas as etapas que desenvolvemos e por fim, um gráfico para melhorar a visualização dos resultados em relação aos feedbacks e quais foram as palavras mais utilizadas em cada polo de avaliação (positivo e negativo). Com essas informações centralizadas, a empresa poderá investigar por exemplo (fictício), porque um certo produto possui uma melhor avaliação sendo vendido no estado de São Paulo do que no estado de Pernambuco (mesmo o produto sendo exatamente o mesmo), neste caso, podem apresentar problemas com a experiência de entrega, seja ela o preço elevado, devido a dificuldade de acessar o local, como não ter entregas para esses endereços.

4 – PROTÓTIPO MOBILE DA APLICAÇÃO

O design de nosso aplicativo é focado nos conceitos UX (User Experience) e UI (User Interface), desenvolvemos uma interface extremamente sofisticada e intuitiva, facilitando não só o acesso as informações, como o preenchimento delas que no caso será a entrada do DataSet.

Em relação à proposta sugerida, algumas telas de demonstração desenvolvidas em nosso projeto são representadas abaixo:

- **Entrada de dados:**



Com base nestas telas, elas fornecem ao usuário um feedback claro sobre o status do carregamento do dataset e orienta sobre os próximos passos a serem tomados para iniciar a análise e explorar os dados com mais detalhes.

- **Análise do dataset:**

A tela de análises do dataset apresenta os seguintes elementos:



- **Menu:**

No canto superior esquerdo da tela. Ao ser clicado, este botão apresenta apenas a opção "Selecionar outro Dataset", permitindo ao usuário carregar um novo conjunto de dados para análise.

- **Resultados:**

Logo abaixo do menu, uma seção intitulada "Resultados" informa ao usuário sobre o número total de avaliações presentes no dataset carregado, destacando a oportunidade de alavancar vendas com base nos insights obtidos a partir desses dados.

- **Campo de Pesquisa:**

Um campo de pesquisa com a mensagem "Pesquisar por um produto" permite ao usuário buscar por produtos específicos dentro do dataset carregado.

- **Potenciais Produtos para Alavancagem:**

Uma seção apresenta três produtos identificados como potenciais oportunidades para alavancar vendas. Esses produtos são destacados por terem poucas ou nenhuma avaliação, apesar de possuírem um alto número de compras.

- **Botão "Expandir":**

Embora ainda não esteja disponível, o botão com um ícone de olho "Expandir" é uma implementação futura para expandir as análises e visualizar mais detalhes sobre os dados.

- **Média de Avaliações por Produto:**

Mostra a média do número de opiniões e estrelas por produto, fornecendo uma visão geral da qualidade média das avaliações.

- **Análise de Sentimento por Produto:**

Um gráfico em formato de pizza exibe a distribuição das avaliações por sentimento (neutra/nula, positiva e negativa) para cada produto, ajudando o usuário a compreender a percepção geral dos clientes em relação aos produtos.

- **Botão "Impulsionar":**

Por fim, um botão "Impulsionar" direciona o usuário para a tela "Impulsionar", onde são fornecidas orientações sobre como aumentar as avaliações e vendas dos produtos.

- **Impulsionar:**

A tela impulsionar apresenta os seguintes elementos:

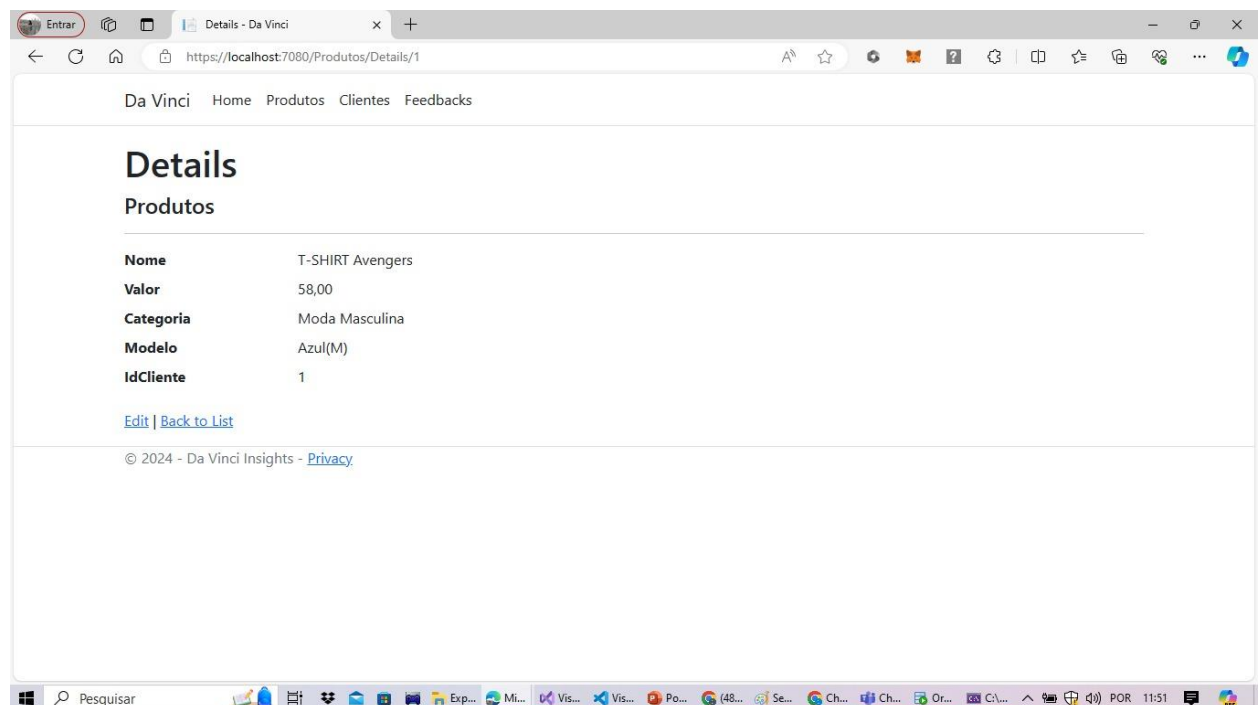
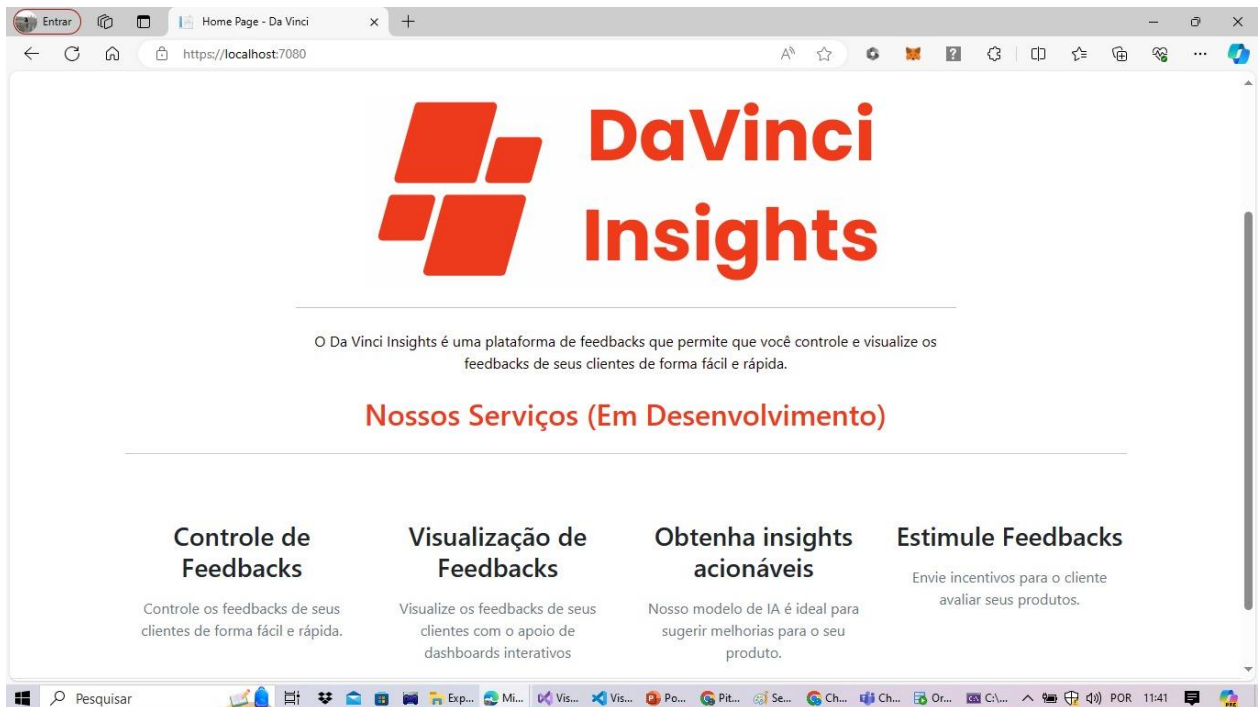


- **Impulsionar Vendas:**
Uma seção destacada na tela apresenta uma mensagem de incentivo para impulsionar as vendas.
- **Potencial Produto para Alavancagem:**
Esta seção exibe um produto específico identificado como uma oportunidade de alavancagem de vendas. Este produto é destacado por ter menos avaliações ou nenhuma, apesar de ter registrado um alto número de vendas. O produto é acompanhado da proposta ao cliente, sendo sugerido o uso de cupons de desconto exclusivos para clientes que deixarem avaliações de produtos. Essa estratégia incentiva os clientes a compartilharem suas opiniões e os recompensa com economia em futuras compras.
- **Botão "Voltar":**
Um botão "Voltar" é disponibilizado na tela para permitir que o usuário retorne à página de análises e continue explorando os dados do dataset carregado.

Diante disso, essa tela oferece ao usuário insights sobre como impulsionar as vendas e fornece orientações sobre como aproveitar ao máximo as oportunidades identificadas no dataset.

5 – DEMONSTRAÇÃO DE TELA WEB EM C#

Demonstração de página funcional do projeto utilizando a linguagem C#, que permite algumas operações de CRUD de maneira inicial (Cliente, Produto, Feedback), e será aprimorada futuramente nas próximas fases.



6 – CONSTRUÇÃO DO MODELO E APLICAÇÃO DE MACHINE LEARNING

6-1 CONCEITO DE PROCESSAMENTO DE LINGUAGEM NATURAL (NLP)

O processamento de linguagem natural (PLN), ou em inglês Natural Language Processing (NLP), combina linguística computacional, modelagem baseada em regras da linguagem humana e modelos estatísticos, de machine learning e de deep learning, para permitir que os computadores processem a linguagem humana na forma de dados de texto ou voz e “entendam” seu significado integral, pode ser usado para identificar a **intenção e o sentimento** do locutor ou escritor.



O PLN envolve várias tarefas, que incluem tradução automática (traduzir texto de um idioma para outro), reconhecimento de fala (converter fala falada em texto escrito), geração de linguagem natural (converter informações de bancos de dados em linguagem compreensível), e muitos outros.

Dentre os desafios da PLN, é necessário observar que a linguagem humana é repleta de ambiguidades, como homônimos, homófonos, sarcasmo, expressões idiomáticas e metáforas, diante disso, programadores devem estar atentos à ensinar aplicativos orientados por linguagem natural a reconhecer e entender essas nuances.

Existem vários processamentos de linguagem natural (PLN) que podem ser aplicados a um texto para extrair informações, entender o significado e realizar tarefas específicas. Aqui estão alguns dos principais exemplos:

1. Tokenização: Divide o texto em unidades menores, como palavras ou frases.
2. Análise Morfológica: Identifica as partes constituintes das palavras, como raízes e sufixos.
3. Análise Sintática: Analisa a estrutura gramatical das sentenças, identificando a relação entre as palavras.
4. Análise Semântica: Compreende o significado das palavras e frases em um contexto específico.
5. Desambiguação Semântica: Resolve ambiguidades de significado em palavras ou frases.

6. Reconhecimento de Entidades Nomeadas (NER): Identifica e classifica entidades, como nomes de pessoas, locais e organizações.
7. Extração de Relacionamentos: Descobre relações semânticas entre entidades no texto.
8. Classificação de Texto: Atribui categorias ou rótulos a textos com base no conteúdo.
9. Agrupamento de Texto: Agrupa documentos ou sentenças semelhantes.
10. Sumarização de Texto: Cria resumos concisos do conteúdo de um texto.
11. Tradução Automática: Converte texto de uma língua para outra.

6-2 APLICAÇÃO DE PROCESSAMENTO DE LINGUAGEM NATURAL (NLP)

No contexto do nosso projeto, o Processamento de Linguagem Natural será utilizado para a compreensão e interpretação de feedback não estruturado, assim como permitir uma análise mais aprofundada das opiniões (comentários em texto) e identificar sentimentos (negativos, positivos e neutros) dos clientes em relação ao produto.

Diante disto, é sugerido criar o modelo em python com base em alguns passos:

- **Análise exploratória de dados:**

Para treinar o modelo, num primeiro momento, vamos explorar o dataset de vendas da Olist. Este dataset contém diversos dados sobre clientes, geolocalização, pedidos, pagamentos, avaliações, produtos e vendedores. Será necessário entender as demandas dos produtos vendidos e então analisar os comentários do consumidor à respeito dos produtos, incluindo avaliações de 1 a 5 estrelas, verificar correlações existente entre as colunas. Será analisado os padrões e tópicos relevantes presentes no feedback dos clientes.

Exemplo fictício: um produto específico tem a média de 5 estrelas relacionadas ao tópico “Entrega do produto” na cidade de São Paulo, enquanto que em Pernambuco tem 3 estrelas para o mesmo tópico. Sugestão: a otimização da entrega do produto na área de maior deficiência pode aumentar a avaliação geral.

- **Pré-processamento de Dados:**

Limpeza e normalização dos dados é um passo crucial para o treinamento de um bom modelo. Para lidar com processamento de texto serão necessários alguns recursos de NLP: Remoção de pontuações, Tokenização, remoção de stopwords, normalização de maiúsculas e minúsculas, stemming.

- **Bag-Of-Words:**

O modelo bag-of-words (BOW) é uma representação que transforma texto arbitrário em vetores de comprimento fixo, contando quantas vezes cada palavra aparece. Este processo é muitas vezes referido como vetorização.

- **Construção do Modelo de Análise de Sentimentos:**

Utilização de algoritmos de aprendizado supervisionado, como Naive Bayes, SVM ou Redes Neurais e selecionar o de melhor performance. Treinamento do modelo utilizando os dados históricos de feedback dos clientes. Identificar a polaridade das opiniões dos clientes

(positivo, negativo ou neutra)

- **Avaliação do Modelo:**

Utilização de métricas de avaliação, como acurácia, precisão, recall e F1-score.

Métrica	Descrição	Fórmula	Legenda: <ul style="list-style-type: none">• TP: Verdadeiros positivos• TN: Verdadeiros negativos• FP: Falsos positivos• FN: Falsos negativos
Acurácia	Porcentagem de predições corretas sobre o total de amostras	$(TP + TN) / (TP + TN + FP + FN)$	
Precisão	Porcentagem de verdadeiros positivos entre os positivos preditos	$TP / (TP + FP)$	
Recall	Porcentagem de verdadeiros positivos identificados corretamente	$TP / (TP + FN)$	
F1 Score	Média harmônica da precisão e do recall	$2 * (Precisão * Recall) / (Precisão + Recall)$	

- **Deploy do Modelo:**

Integração do modelo em uma API para ser utilizado em tempo real e monitoramento contínuo do desempenho do modelo e re-treinamento conforme necessário. Será possível integrar a API na aplicação mobile, ou em sites web, de acordo com a necessidade.

6-3 BIBLIOTECAS / FRAMEWORKS UTILIZADOS

Dentre as bibliotecas e frameworks que podem ser utilizados no desenvolvimento do nosso modelo, é possível citar:

Para Machine Learning:

- Scikit-learn

Para a implementação de algoritmos de aprendizado de máquina na categorização e análise do feedback dos clientes.

- TensorFlow e Keras

Deep learning e para a construção de modelos de IA personalizados que podem identificar padrões e tendências no feedback do usuário.

Para manipulação e análise de dados:

- Pandas e NumPy

É necessário manipular dados para proporcionar uma base sólida para a análise de feedback qualitativo.

Visualização de Dados:

- Matplotlib ou Seaborn

Para a visualização de gráficos e obter insights.

- Plotly

Exibir gráficos interativos, nuvens de palavras, mapas de calor e gráficos de tendências. É possível explorar a distribuição dos sentimentos, principais tópicos e aspectos mencionados pelos clientes, evoluções, entre outros.

NLP (Processamento de Linguagem Natural):

- NLTK (Natural Language Toolkit):

Processamento de linguagem natural em Python, que fornece acesso a recursos e algoritmos de PLN amplamente utilizados.

- TextBlob

Simplifica operações de PLN, como classificação de sentimento, extração de frases-chave e tradução.

- spaCy

Suporta tarefas como tokenização, parsing, nomeação de entidades, e vetorização de palavras. Facilita a integração com outras bibliotecas de machine learning.

- Hugging Face Transformers:

Tem modelos de linguagem pré-treinados que podem ser utilizados para tarefas específicas de PLN.

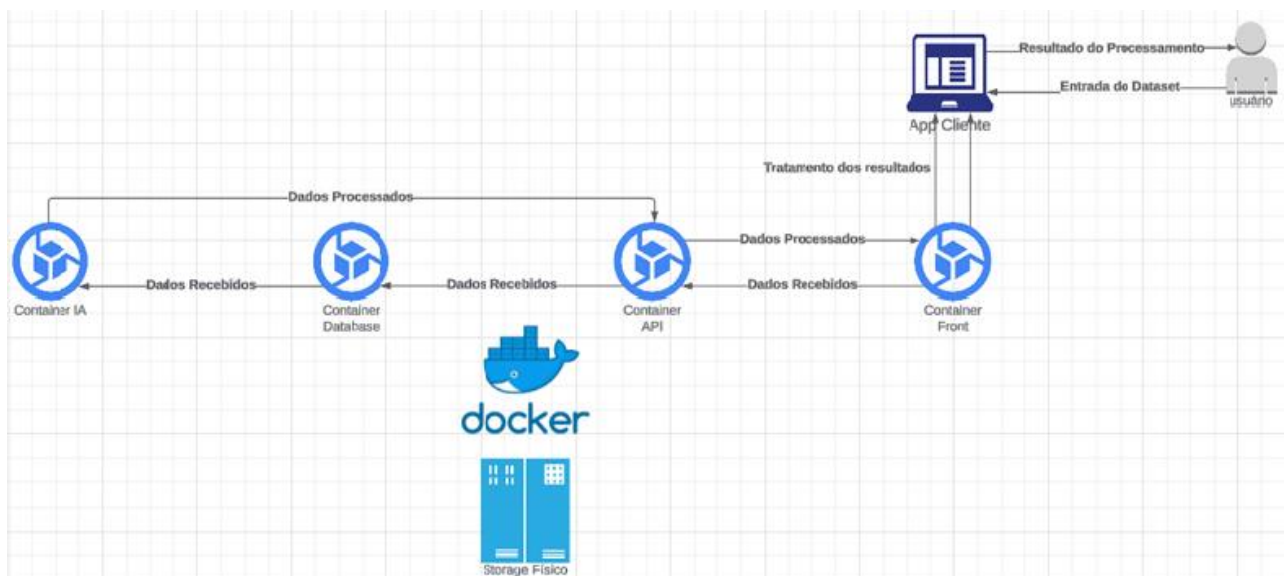
6-4 FERRAMENTAS ADICIONAIS

- Power BI ou Tableau:

Ferramentas de visualização de dados que podem ser usadas para criar dashboards e gráficos interativos a partir dos insights extraídos.

7 – FERRAMENTAS DOCKER

Como o projeto analisa o grande fluxo de dados de feedbacks e comentários, faz-se necessário a implementação de uma infraestrutura para a virtualização em contêineres e micro serviços. Os contêineres Docker proporcionam isolamento entre os diferentes componentes da aplicação, aumentando a segurança e a estabilidade do sistema. Este sistema consiste em um frontend desenvolvido em uma aplicação web moderna e uma API em Java no backend, que atua como uma camada intermediária entre o frontend e a IA.



Descrição da infraestrutura docker:

Entrada:

- Storage Físico: máquina com os hardwares necessários para a virtualização, como o processador, memória RAM, disco e placa de rede.
- Docker: Ferramenta que utilizaremos para o uso e gerenciamento dos containers de modo mais eficiente e intuitivo.
- App Cliente: aplicativo final e interface que o usuário irá interagir.
- Container Front: container responsável por ter todas as dependências e entrada do front, e envia esses dados para API.
- Container API: container responsável pelo funcionamento da API, que irá intermediar a comunicação do front, Database e IA.
- Container Database: container que irá receber os dados trazidos pela API e depois enviar esses dados para a IA.
- Container IA: container que irá receber todos os dados e é responsável por todo o processamento deles.

Saída:

- Com o resultado do processamento, eles serão levados pela API até o front, onde passará por um tratamento para que os insights sejam mais visuais e de fácil entendimento do usuário, como gráficos e pequenas descrições dos resultados com alguns indicativos de melhora.

7-1 RECURSOS DA AZURE

- Virtual Machine (IaaS)

As Máquinas Virtuais são um exemplo clássico de IaaS. Elas nos permitirão que possamos configurar e gerenciar nossos servidores virtuais, proporcionando flexibilidade e escalabilidade. Com as VMs, temos controle sobre o sistema operacional, armazenamento, e a configuração da rede, permitindo ajustar os recursos de computação (como CPU, memória e disco) de acordo com as necessidades específicas. Isto elimina a necessidade de adquirir e manter hardware físico, oferecendo uma infraestrutura sob demanda.

- Load Balancer (IaaS)

Garante a disponibilidade e distribuição eficiente de tráfego entre as máquinas virtuais, iremos utilizar o Load Balancer, ou seja, um balanceador. Este serviço é responsável por distribuir automaticamente as solicitações entre as VMs (Front e Back), de forma que nenhuma máquina seja sobrecarregada e que nossa solução permaneça responsiva mesmo durante picos de demanda.

- VNet (IaaS)

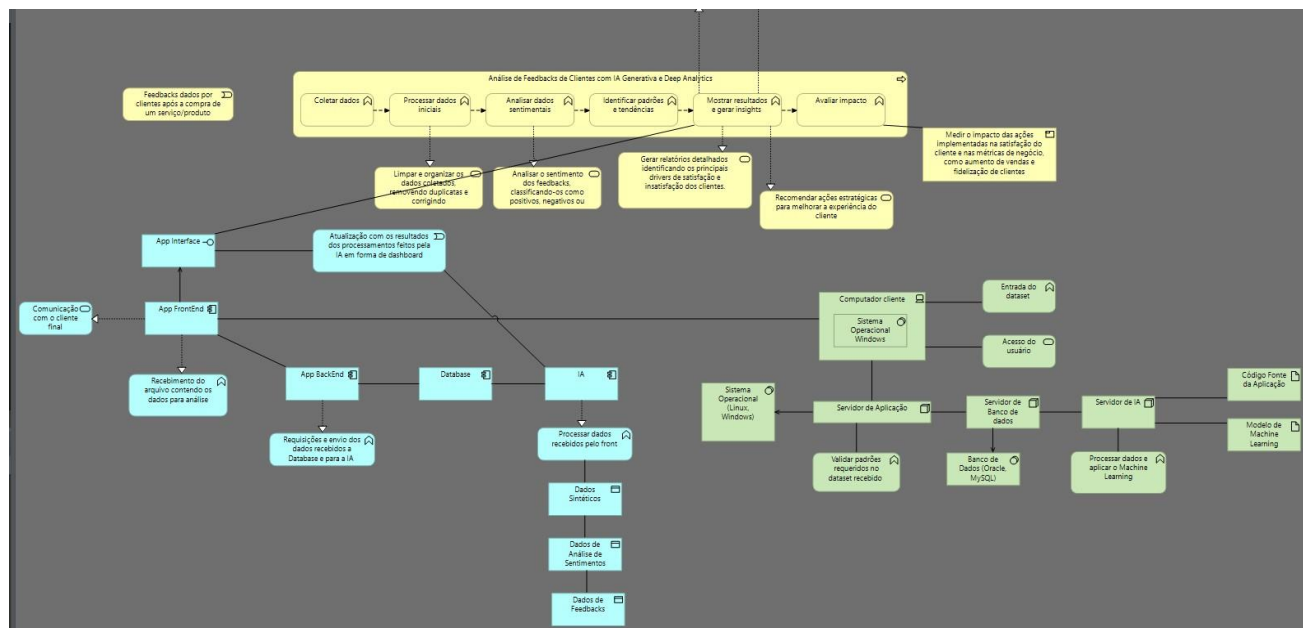
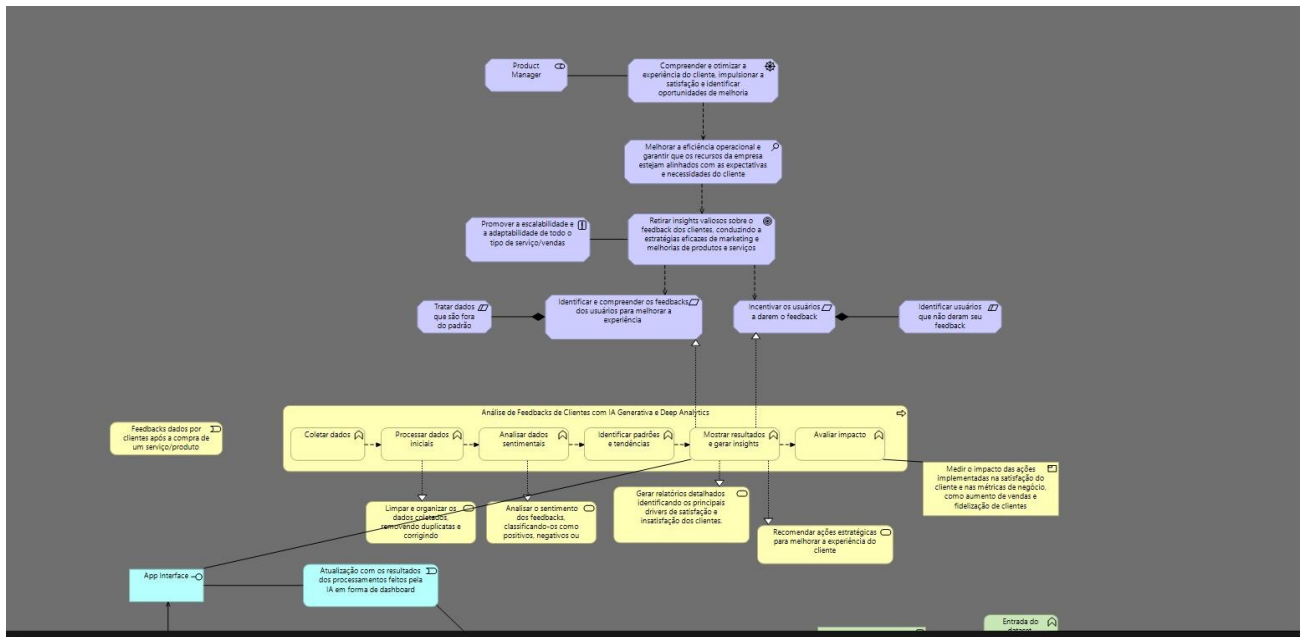
É responsável por garantir um isolamento e proteger a infraestrutura da aplicação, permite a comunicação entre as VM's, balanceador de carga e outros recursos de rede, assim como facilitar a implementação de políticas de segurança como firewalls e grupos de segurança de rede.

- Storage Account (PaaS)

Será utilizado para armazenar datasets, resultados de análises, logs e backups. Além de oferecer alta disponibilidade, redundância geográfica, acesso seguro e a possibilidade de compartilhamento de grandes massas de dados.

8 – DIAGRAMA DA ARQUITETURA (ARCHIMATE)

Para garantir uma visualização mais ampliada deste diagrama, as imagens e archimate do mesmo será também disponibilizada dentro da pasta da entrega, de modo separado deste documento.



9 – CONCLUSÃO

Em suma, este projeto destaca a importância do processamento de dados para compreender e agir sobre os feedbacks dos usuários. Ao analisar esses feedbacks, as empresas podem identificar as razões por trás das avaliações, desenvolver estratégias para melhorar a satisfação do cliente e, conseqüentemente, atrair novos compradores. Ao reconhecer que as avaliações dos clientes desempenham um papel crucial na decisão de compra, é necessário também estimular os clientes à adotarem o hábito de avaliar as suas compras, como também é fundamental investir em análise de feedbacks para o sucesso e a competitividade das empresas no mercado. Com esta infraestrutura, estaremos bem equipados para transformar feedbacks em insights acionáveis, ajudando as empresas a melhorar a satisfação do cliente e aumentar suas vendas.