

**Innov8tech**

***“Criação de IA generativa a fim de diagnosticar o histórico da figura pública de interesse antes da contratação para divulgação do produto/serviço.”***

**Ronald de Oliveira Farias - RM552364**

**Gustavo Carvalho Noia - RM552466**

**Lucas Serbato de Barros - RM551821**

**Phablo Isaías Santos - RM550687**

**Vitor Teixeira Silva - RM552228**

## Sumário

DEFINIÇÃO DO PROBLEMA .....	3
PROPOSTA DE SOLUÇÃO .....	3
PÚBLICO-ALVO .....	4
PRODUTOS SEMELHANTES .....	4
POTENCIAL DE MERCADO .....	4
COMO A TÉCNICA DE VIRTUALIZAÇÃO CONTRIBUIRÁ COM A ENTREGADO PROJETO? .....	5
NOSSO SERVIÇO DE NUVEM .....	6
TIPO DE SERVIÇO .....	6
OS RECURSOS QUE UTILIZAREMOS DO SERVIÇO PRESTADO EMNUVEM DE IAAS .....	6
ARQUITETURA MACRO DA SOLUÇÃO.....	7
ARQUITETURA DA SOLUÇÃO (IA) .....	7
Vídeo PITCH .....	10

## DEFINIÇÃO DO PROBLEMA

A solução proposta pelo grupo, visa ajudar em um problema que foi apresentado durante o *Kickoff* do *Challenge* pela empresa *Plusoft*. Enquanto o representante da citada contava sobre alguns *cases*, um deles foi sobre uma empresa que, após sua publicidade em um programa de televisão, acabou ocasionando uma repercussão negativa que foi identificada somente após o lançamento da publicidade; devido ao passado recente da figura pública escolhida para apresentar o produto deles. Em um cenário desses, uma marca pode ter sua imagem vinculada a algo que não faz parte de seus valores, ocasionando na perda de receita em caso de boicote dos seus produtos; em caso de lançamento do produto pode ser mais grave, uma vez que a empresa pode investir muito no desenvolvimento de um novo produto, identidade visual, marketing, publicidade e afins na espera de um bom feedback por parte da mídia.

Para se ter um pouco melhor de noção sobre o quanto pode afetar uma empresa, de acordo com um relatório de consultoria da empresa *Strategy&* do grupo *PWD* de 2016, as vinte empresas que mais investiam em pesquisa tecnológica por exemplo, somavam 179,3 bilhões de dólares investidos. Dessas vinte empresas, as três primeiras foram *Volkswagen* com 13,2 bi; *Samsung* com 12,7 bi e *Amazon* com 12,5 bi.

(Consulte a fonte no link: '<https://blog.onedaytesting.com.br/investimentos-empresas-inovadoras/>').

Já o investimento em publicidade e marketing, de acordo com a revista *Ad Age* em 2022, a *Amazon* apareceu em primeira posição investindo 16,9 bi; *Alibaba Group* em segundo com 14,2 bi e *L'Óreal* em terceiro com 12,5 bi investidos.

(Consulte a fonte no link: '<https://www.meioemensagem.com.br/marketing/conheca-os-25-maiores-anunciantes-do-mundo-em-2022/>').

Só envolvendo os valores investidos com marketing e desenvolvimento, já chegamos na casa dos bilhões; e como todo investimento é aguardado um retorno, esse retorno pode ser afetado dependendo de como e por quem o produto e serviço de uma empresa foi apresentado, prejudicando os retornos estimados pela empresa.

## PROPOSTA DE SOLUÇÃO

A proposta de solução da questão abordada anteriormente se constitui em desenvolver uma aplicação *mobile*, onde o cliente consiga pesquisar uma figura pública informando, por exemplo, o seu nome usuário no X – antigo *Twitter* -, o ramo em que a empresa atua e o produto que pretende realizar a ação de marketing, e, com essas informações, uma IA generativa analisaria essa figura e tentaria mensurar o quanto essa figura se relaciona com a empresa e qual a situação de sua imagem nos últimos tempos.

Entrando mais na parte técnica, o aplicativo irá conter telas como home; perfil do cliente; login; pesquisa; resultado e uma tela onde será possível verificar o histórico de pesquisa

desse cliente. A parte de login, recuperação e persistência dos dados e consulta na IA serão realizadas por meio de APIs desenvolvidas, provavelmente, em Java, C# e Python (que cuidará mais da parte envolvida com a IA). Enquanto no Javascript serão feitas as requisições assíncronas. Uma vez feita a consulta na IA, a intenção é que ela retorne para o cliente se a figura pública que ele pretende contratar é adequada à empresa e se ela tem uma boa imagem atualmente, ou se vêm com um histórico recente de escândalos anteriores. Evitando, assim, prejudicar a marca de alguma forma, arriscando todo o investimento no produto.

Atualizações: Devido ao fato de que a rede social X foi suspensa ao na tentativa de estar acima da constituição brasileira, o twitter foi suspenso como ordem do Supremo Tribunal Federal. Com isto, nosso projeto teve de mudar sua rota e por coincidência tivemos conhecimento de uma rede social chamada Bluesky, que possuem a mesma funcionalidade que o X (*twitter*).

## PÚBLICO-ALVO

O público-alvo da nossa solução seria qualquer pessoa jurídica que tenha interesse em contratar uma figura pública que, por meio de seus seguidores, aumente a projeção de venda dos seus produtos ou serviços ou até mesmo aumente sua visibilidade no mercado, fazendo com que sua marca chegue a pessoas que os meios tradicionais de publicidade não atingem.

## PRODUTOS SEMELHANTES

- **YouScan:** Oferece ferramentas de análise de mídia social para monitorar a reputação de figuras públicas e celebridades.
- **Crisp:** Especializada em monitoramento de mídia social para identificar e gerenciar crises de reputação online, o que pode incluir análise de figuras públicas.
- **Klear:** Fornece análise de influenciadores, incluindo métricas de engajamento, autenticidade e histórico de comportamento nas redes sociais.
- **Reputation.com:** Oferece soluções para gestão de reputação online, incluindo monitoramento de menções na mídia e análise de sentimentos.

## POTENCIAL DE MERCADO

Atualmente o Brasil é o país que mais consome conteúdo de influencers digitais. Um estudo recente feito pela Nielsen revela que o país possui mais de 500.000 influencers com ao menos 10.000 seguidores espalhados por diversas plataformas.

Outro levantamento de dados feito pelas agências *Hootsuite* e *We Are Social* demonstra que o Brasil é o segundo país no mundo que mais consome conteúdo de influenciadores (sendo **44,3%** dos usuários da internet), ficando apenas atrás das Filipinas (**51,4%**). Ao analisarmos esses dados podemos concluir que o potencial de mercado para a exploração desse nicho é enorme. Dando um adendo a isto, com a efetivação da nossa solução, é possível que a empresa consiga atingir uma expressiva rentabilidade através da divulgação de seus produtos e serviços pelos influenciadores digitais. Principalmente no Brasil, visto anteriormente, que é um dos países mais prósperos para esse tipo de nicho de mercado. Aliado a isto, é imprescindível a prevenção de possíveis riscos

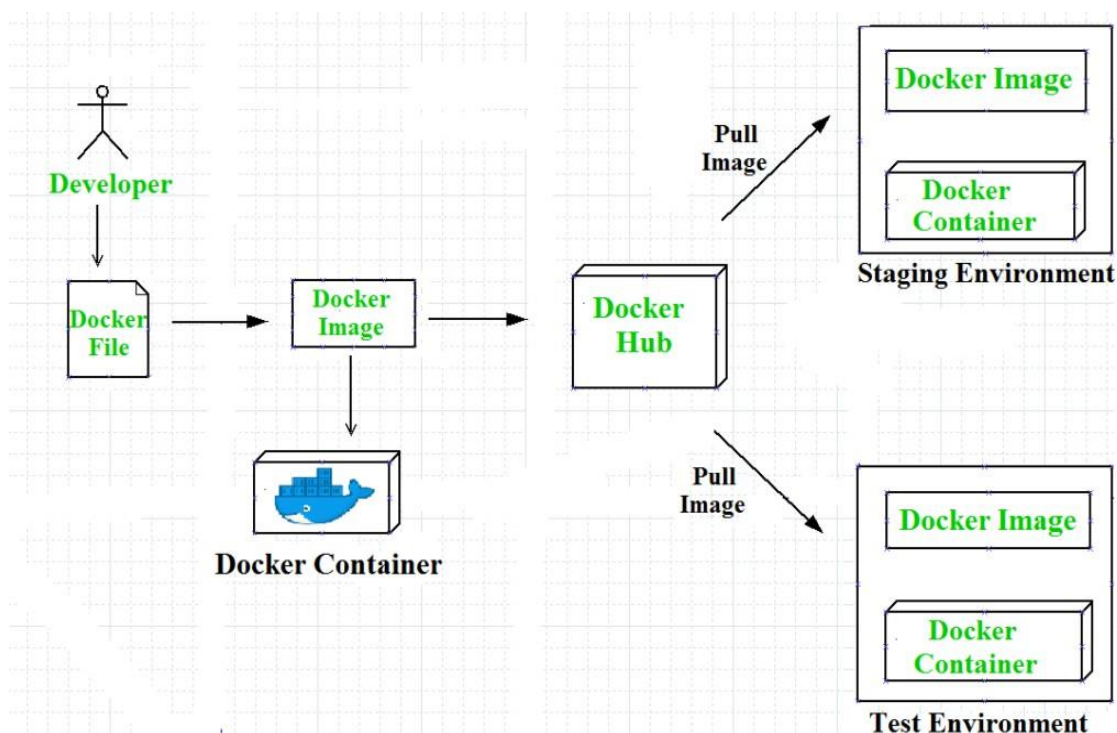
onde a nossa solução entra como complementação e evita possíveis estresses e perdas de lucro devido um deslize de não considerar o passado do influencer que foi escolhido.

(Consulte a fonte no link: '<https://veja.abril.com.br/comportamento/pesquisa-revela-que-o-brasil-e-o-pais-dos-influenciadores-digitais>').

## COMO A TÉCNICA DE VIRTUALIZAÇÃO CONTRIBUIRÁ COM A ENTREGA DO PROJETO?

O nosso projeto utilizará algumas APIs, e, nesse ponto, a virtualização irá auxiliar no desenvolvimento e qualidade da aplicação, uma vez que poderemos desenvolver nosso projeto sem preocupação de consumo de memória e espaço da máquina local, e a desnecessidade de vigilância frequente relacionado a desempenho por parte de cada um do grupo devido a algum fator externo. Acompanhado a isso, ao invés de termos as APIs localmente em nossas máquinas ou o projeto inteiro no mesmo lugar, iremos criá-los *containers* pelo *Docker*, que terão somente o necessário para o funcionamento de cada API, e assim, levantaremos a máquina em um ambiente cloud, como por exemplo, a *Azure*.

Exemplo de containerização do ambiente de desenvolvimento pelo *Docker*:



Dessa forma, cada API terá sua própria máquina em um ambiente de alta disponibilidade e, mesmo que ocorra algum problema com uma máquina, as outras poderão continuar trabalhando sem precisar parar a aplicação inteira. Vale ressaltar que com a virtualização, será mais fácil seguir boas práticas como utilizar o *Git Flow*, uma

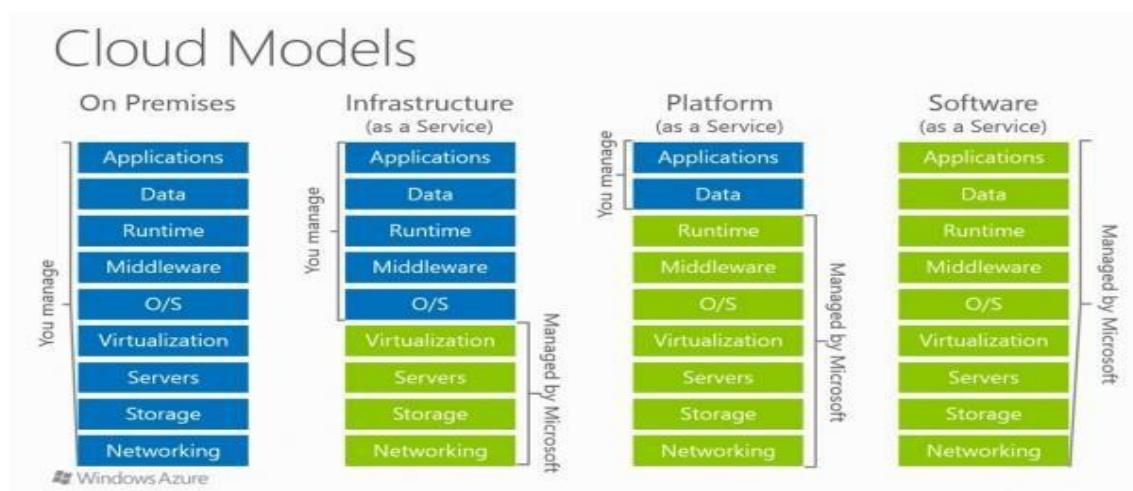
vez que é mais fácil levantar as máquinas e separar o ambiente de desenvolvimento do ambiente de testes, e depois que desenvolvidos e testados, encaminharmos para o ambiente de produção. Outras vantagens da virtualização utilizando cloud é que a preocupação e investimento com a obsolescência dos equipamentos, manutenção, cuidados e afins, deixam de ser nossos e passa para a ser preocupação da empresa que disponibiliza o serviço, o redimensionamento da máquina também fica simplificado, podendo aumentar ou diminuir facilmente conforme a demanda e distribuindo a aplicação pelo mundo, evitando quedas por desastre natural e afins.

## NOSSO SERVIÇO DE NUVEM

Devido ao fato de estarmos oferecendo um produto à empresa *Plusoft*, acreditamos que a nossa solução se encaixaria melhor como sendo pública - Partindo do ponto de vista se fôssemos a própria *Plusoft* -, visto que seria necessária uma massa de dados de tweets do X para mensurar o grau de relacionamento que a IA generativa irá estimar referente à popularidade da figura pública. Não teríamos renda o suficiente para ter capacidade de armazenar esses dados e garantir a segurança deles com data centers e uma infraestrutura de ponta a ponta.

## TIPO DE SERVIÇO

O tipo de serviço de nuvem que nosso projeto seria melhor aproveitado, é o serviço de IaaS, visto que não ficaríamos preocupados com questão de custos altos de manutenção dos dados, cabeamento, data centers e infraestrutura como um todo, pois o serviço de IaaS iria nos proporcionar tudo isso; não teríamos tanta preocupação referente a questão de segurança dos dados do nosso projeto por haver a facilidade de existirem plataformas hoje em dia que podem nos ajudar de acordo com a demanda ofertada, como por exemplo a Microsoft Azure, que oferecem diversos serviços de infraestrutura (e não somente) que atende a determinadas necessidades sob demanda.

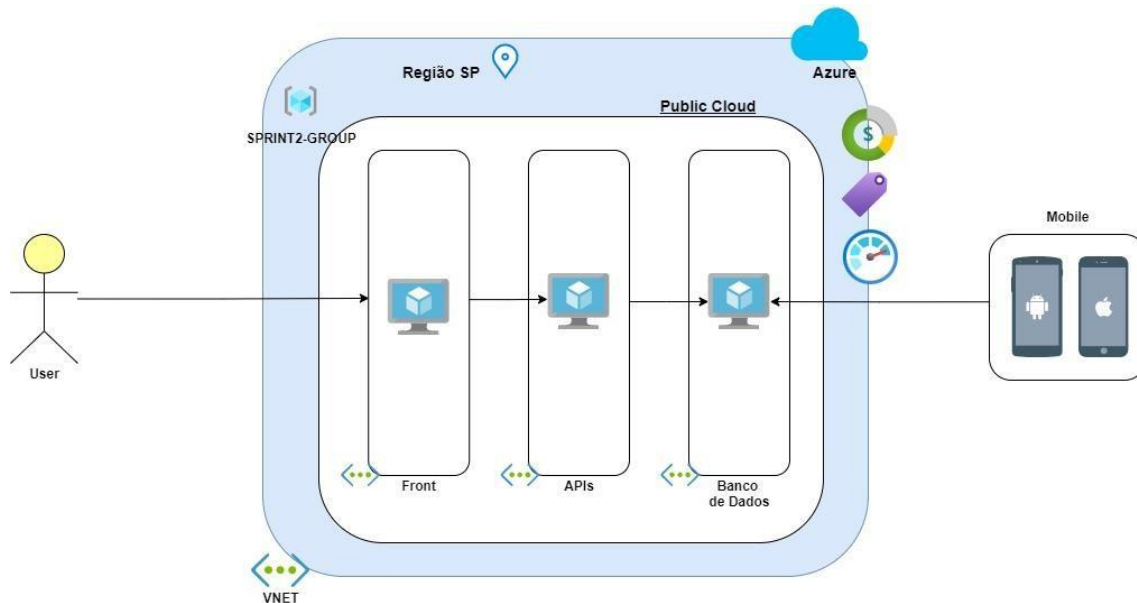


## OS RECURSOS QUE UTILIZAREMOS DO SERVIÇO PRESTADO EM NUVEM DE IAAS

- **Resource Groups:** para facilitar o gerenciamento de todos os nossos recursos do cloud, garantindo políticas de acesso como monitoramento e governança de forma mais eficiente.

- **Virtual Network:** recurso para gerenciamento de nossas próprias redes para o gerenciamento das conexões, a fim de garantir mais segurança como controle do acesso, e a conectividade entre os outros recursos da nossa solução. Garantindo uma comunicação confiável.
- **Virtual Machines:** a fim de centralizar o ambiente de desenvolvimento dos participantes, para que não tenha problemas com a utilização da máquina local dos desenvolvedores e nem conflito no momento de versionamento/manutenção do projeto.

## ARQUITETURA MACRO DA SOLUÇÃO



## ARQUITETURA DA SOLUÇÃO (IA)

### Tecnologias Utilizadas

- **BERT (Bidirectional Encoder Representations from Transformers):** Utilizado para gerar embeddings de postagens e capturar o significado contextual profundo de cada postagem de influenciadores.
- **VADER (Valence Aware Dictionary and Sentiment Reasoner):** Ferramenta de análise de sentimento usada para determinar a aceitação pública das postagens dos influenciadores.
- **Deep Learning:** O modelo de deep learning classifica e ordena influenciadores de acordo com sua relevância e engajamento.

### Processo de Implementação

- **Coleta de Dados:** A API do Bluesky coleta os influenciadores e suas postagens, buscando perfis relacionados ao nicho do produto inserido.
- **Análise de Sentimento:** O VADER é aplicado para calcular a percepção pública das postagens dos influenciadores, medindo a aceitação.
- **Geração de Embeddings:** Utilizando o BERT, as postagens são processadas

para gerar embeddings que captam o contexto semântico.

- **Classificação dos Influenciadores:** O modelo analisa as postagens e o sentimento para gerar uma classificação dos 10 influenciadores mais relevantes

## BASE DE DADOS UTILIZADA

A base de dados utilizada foi obtida por meio da API do Bluesky, uma rede social que funciona como o antigo twitter. O foco da coleta de dados está na identificação de influenciadores altamente relevantes para o nicho de mercado inserido pelo cliente. Esse processo visa garantir que as análises sejam o mais precisas possível em relação ao produto, ramo de atuação e imagem pública do influenciador.

## Detalhes da Coleta de Dados

Foram analisadas postagens recentes dos influenciadores para estabelecer a relevância do conteúdo em relação ao produto que o cliente deseja promover. O algoritmo foca na relação entre o conteúdo postado e o nicho de mercado do cliente, garantindo que a escolha dos influenciadores seja baseada não apenas em popularidade, mas na relevância específica do conteúdo divulgado.

Comentários, curtidas e outras formas de interação social foram avaliadas em cada postagem. Essa parte é importante para termos noção do tratamento do influenciador com o seu grupo indo além de simples números de seguidores. A análise busca quantificar o nível de engajamento e a qualidade das interações, ajudando a determinar quais influenciadores podem gerar maior impacto para a marca do cliente.

## Pré-processamento dos Dados

Os textos coletados passam por uma etapa de pré-processamento, onde são Os dados coletados passam por uma etapa de pré-processamento, essencial para que o modelo de IA extraia insights de forma eficiente e precisa:

- **Tokenização:** Todos os textos coletados são divididos em tokens, ou seja, fragmentos menores que permitem ao modelo compreender cada palavra, frase ou termo técnico.
- **Embeddings com BERT:** Após a tokenização, os dados são processados pelo modelo BERT para gerar embeddings. Esses embeddings são representações numéricas dos textos, que capturam o contexto semântico de cada palavra ou expressão, permitindo que o modelo compreenda o significado subjacente das postagens.
- **Análise de Sentimento:** Utilizando a ferramenta VADER, aplicamos uma análise de sentimento a cada postagem para avaliar a percepção do público em relação ao influencer, dividido em três categorias (neutro, positivo ou negativo).



## Reflexão (Autocrítica)

Nosso projeto foi elaborado com o objetivo de resolver a questão da escolha de influenciadores que seriam úteis para a marca e que apresentassem um bom histórico na mídia em sua atuação. Dessa forma, evitamos desperdícios e eventuais problemas decorrentes de influenciadores que estivessem passando por cancelamentos. Escolhemos, então, a rede social X, que, teoricamente, teria uma ótima aplicabilidade devido à sua relevância na comunidade internauta brasileira. Assim, ao analisar o público brasileiro que já utiliza essa rede social, conseguimos desenvolver um modelo de deep learning capaz de, por meio da análise de sentimento, estimar se um influenciador voltado ao nicho escolhido pelo cliente (empresa) poderia se adequar para realizar um serviço de publicidade para a marca.

No entanto, devido aos conflitos entre o CEO do X – Elon Musk – e o ministro do Supremo Tribunal Federal – Alexandre de Moraes –, a rede social foi bloqueada no Brasil por cerca de um mês, o que impactou a entrega do protótipo de código planejada para a Sprint 3. Dessa forma, optamos por usar uma rede social alternativa que conseguiu atrair uma parcela do público que antes estava no X e que agora havia migrado para a rede social Bluesky. Em teoria, tudo teria ocorrido conforme o planejado, pois essa nova rede social “supre” as funcionalidades da anterior (que estava, à época, banida por tempo indeterminado).

Todavia, enfrentamos outro problema: a rede social ainda estava em sua versão beta, com diversos bugs e problemas, e apresentava bem menos funcionalidades do que o antigo Twitter. Além disso, a sua API possuía um número limitado de requisições, o que nos impediu de aplicar nosso modelo de deep learning conforme planejado. Isso gerou um grande transtorno na entrega do modelo, que precisou ser apresentado com o código formulado, mas com um erro na coleta de dados para a análise de sentimento.

Na sprint atual (Sprint 4), tivemos a sorte do X voltar a operar no Brasil, o que nos trouxe mais conforto e tranquilidade para aplicar o modelo originalmente pensado em torno dessa rede social. No entanto, novamente enfrentamos um grande obstáculo. Embora a rede social permita o uso de uma versão gratuita de sua API – conforme o planejado –, ao aprofundarmos nossa pesquisa na documentação, percebemos que a versão gratuita limita o acesso a poucos tweets de qualquer perfil público do X. Além disso, grande parte dos recursos necessários para a análise de sentimento do modelo só está disponível na versão paga, o que novamente gera um grande transtorno em relação à aplicabilidade do projeto.

Se tivéssemos recursos, faríamos algo diferente, como elaborar uma nova ideia ou, futuramente, conseguir um fundo de investimento entre os integrantes do grupo para custear a versão paga da API, que infelizmente está fora do nosso orçamento atual. Uma nova ideia precisaria ser pensada, o que demandaria uma reformulação completa do projeto. Com orçamento adequado, o projeto teria funcionado extremamente bem, pois se alinha com a realidade das empresas

que buscam influenciadores para divulgar seus produtos. No entanto, o que não funcionou foi o impacto do banimento temporário do Twitter no Brasil e as limitações impostas pela API gratuita, que nos impedem de obter o acesso necessário para o desenvolvimento completo do modelo.

## **PITCH**

### **Vídeo PITCH**

<https://www.youtube.com/watch?v=dfO73PmUA74>

Atualizações PITCH (SPRINT 2)

<https://www.youtube.com/watch?v=IJn4Mlgskh4>

Atualizações PITCH (SPRINT 3)

<https://www.youtube.com/watch?v=lhwtv0N1dL0>

Atualizações PITCH (SPRINT 4)

<https://www.youtube.com/watch?v=rVD6d9yPTNY>