UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Análise do Engajamento de Postagens no Instagram relacionado ao tema Cirurgia Plástica

Carolina Toledo Ferraz

Monografia - MBA em Inteligência Artificial e Big Data

ICMC USP SÃO CARLOS

Carolina Toledo Ferraz

Análise do Engajamento de Postagens no Instagram relacionado ao tema Cirurgia Plástica

Monografia apresentada ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Concentration area: Inteligência Artificial

Orientador: Prof. Dr. Diego Furtado Silva

Coorientador: Prof. Dr. Claudio Luis Cruz de Oliveira

Versão original

São Carlos

2023

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

T649a

Toledo Ferraz, Carolina
Análise do Engajamento de Postagens no Instagram relacionado ao tema Cirurgia Plástica / Carolina Toledo Ferraz; orientador Diego Furtado Silva; coorientador Claudio Luis Cruz de Oliveira. -- São Carlos, 2023.
52 p.

Trabalho de conclusão de curso (MBA em Inteligência Artificial e Big Data) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2023.

1. Reconhecimento de Padrões. 2. Redes Neurais. 3. Aprendizado de Máquina. I. Furtado Silva, Diego , orient. II. Cruz de Oliveira, Claudio Luis, coorient. III. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

Carolina Toledo Ferraz

Engagement Analysis of Posts on Instagram Related to Plastic Surgery

Monograph presented to the Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, as part of the requirements for obtaining the title of Specialist in Artificial Intelligence and Big Data.

Concentration area: Artificial Intelligence

Advisor: Prof. Dr. Diego Furtado Silva

Original version

São Carlos

2023

Este trabalho é dedicado ao meu marido que sempre me incentivou na jornada do conhecimento.

AGRADECIMENTOS

Inicialmente, agradeço a Deus, por me proporcionar saúde e energia para alcançar o meu objetivo.

Agradeço ao meu orientador Prof. Dr. Diego Furtado, pelos momentos de discussão e aprendizado. Agradeço também meu co-orientador Prof. Dr. Claudio Oliveira pelos ensinamentos e discussões.

Agradeço aos meus queridos pais, Mercedes e João Baptista, que me educaram e me incentivaram na minha caminhada até aqui. Em especial ao meu pai (em memória) que sempre se orgulhava do meu trabalho e da minha dedicação.

Agradeço ao meu amado marido Marcelo que me apoiou em todos os momentos da minha vida e me incentivou a continuar os estudos.

Agradeço, por fim, a Empresa Cognitive que me deu oportunidade de trabalho e desenvolvimento deste trabalho.

"Cada escolha, uma oportunidade.
Cada queda, um aprendizado.
Cada atitute, uma consequência"
Autor Desconhecido

RESUMO

FERRAZ, C.T. **Análise do Engajamento de Postagens no Instagram relacionado ao tema Cirurgia Plástica**. 2023. 52p. Monografia (MBA em Inteligência Artificial e Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

Os brasileiros passam, em média, 46 horas por mês em mídias sociais, observando conteúdos, trocando informações e interagindo entre si. Um indicador relavante em mídias sociais é o engajamento, que é o conjunto de visualizações, *likes*, comentários e compartilhamentos. A predição do engajamento em mídia social usa os dados disponíveis de uma determinada postagem de mídia social como temas usados na mensagem, formato do conteúdo e uso de influenciadores na divulgação. Há um grande interesse nessa área, principalmente devido aos benefícios financeiros que são concedidos aos produtores de conteúdo que aumentam conforme sua audiência cresce e o impacto do engajamento no consumo. No entanto, o problema de predição do engajamento ainda é desafiador devido a fatores complexos, como a qualidade do conteúdo e a relevância para os expectadores, que são fatores difíceis de se medir. Este trabalho apresenta um estudo acerca da metodologias capazes de prever o engajamento de postagens no Intagram relacionadas ao tema Cirurgia Plástica com o objetivo de avaliar quais características gerariam maior engajamento e se a concatenação de características mostrariam relevância para o modelo. A metodologia adotada neste trabalho englobou os seguintes passos: obtenção das métricas via API, criação de novas métricas, concatenação das métricas que serão utilizadas para o treinamento do modelo, vetorização das métricas, modelagem utilizando 6 diferentes modelos de aprendizado de máquina e avaliação dos modelos. As melhores acurácias e F1 *score* obtidos mostraram valores acima de 70% e a AUC ROC acima de 80%. Dentre as diferentes métricas obtidas e criadas no desenvolvimento do método, percebemos que a característica *caption* das postagens teve uma grande importância para o modelo e que a concatenação desta métrica com características retiradas das imagens obtidas das postagens não agregou tanta importância para a metodologia.

Palavras-chave: Instagram, Engajamento, Aprendizado de Máquina, Classificadores

ABSTRACT

FERRAZ, C.T. **Engagement Analysis of Posts on Instagram Related to Plastic Surgery**. 2023. 52p. Monograph (MBA in Artificial Intelligence and Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

Brazilians spend, on average, 46 hours a month on social media, watching content, exchanging information and interacting with each other. A relevant indicator in social media is engagement, which is the set of views, likes, comments and shares. Predicting social media engagement uses available data from a given social media post such as themes used in the message, content format, and use of influencers in outreach. There is great interest in this area, mainly due to the financial benefits that are granted to content producers that increase as their audience grows and the impact of engagement on consumption. However, the problem of predicting engagement is still challenging due to complex factors, such as content quality and relevance to viewers, which are difficult to measure. This work presents a study on methodologies capable of predicting the engagement of posts on Instagram related to the topic of Plastic Surgery, with the objective of evaluating which features would generate greater engagement and whether the concatenation of features would show relevance for the model. The methodology adopted in this work encompassed the following steps: obtaining metrics via API, creating new metrics, concatenating the metrics that will be used for model training, vectorizing the metrics, modeling using 6 different machine learning models, and evaluating the models. The best accuracies and F1 score obtained showed values above 70% and AUC ROC above 80%. Among the different metrics obtained and created in the development of the method, we noticed that the caption feature of the posts was of great importance for the model and that the concatenation of this metric with features taken from the images obtained from the posts did not add so much importance to the methodology.

Keywords: Instagram, engagement, machine learning, classifiers.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

1.1 Contextualização

Hoje em dia as pessoas passam muito tempo em várias plataformas de mídia social, trocando informações e interagindo entre si. Como as mídias sociais ocupam grande parte do cotidiano das pessoas modernas, muitas pessoas estão interessadas em pesquisar como extrair dados das mídias sociais. Um exemplo de informação que pode ser obtida nas mídias sociais é a pontuação de popularidade. Especificamente, essa pontuação informa quantas pessoas visualizaram uma postagem e um número maior de visualizações significa uma alta influência. A predição de popularidade de mídia social tem como objetivo estimar a pontuação de popularidade usando os dados disponíveis de uma determinada postagem de mídia social. Há um grande interesse nessa área, devido aos benefícios financeiros que são concedidos aos produtores de conteúdo que aumentam conforme sua audiência cresce.

No entanto, o problema de predição da popularidade de postagens em mídias sociais ainda é desafiador devido a fatores complexos, por exemplo, a qualidade do conteúdo e a relevância para os espectadores são alguns dos fatores que são difíceis de medir (LIN *et al.*, 2022). Alguns métodos recentes tentam lidar com esses fatores complexos adicionando múltiplas modalidades (HESSEL; LEE; MIMNO, 2017; CHEN *et al.*, 2019b), como imagens (XU *et al.*, 2020), contexto temporal (WU *et al.*, 2017), *tags* e categorias.

Os primeiros estudos sobre predição de popularidade adotam principalmente métodos baseados em características (TSUR; RAPPOPORT, 2012), que extraem um grande número de características relacionadas a atributos do usuário, rede do usuário, conteúdo de texto e séries temporais e, em seguida, treinam modelos de aprendizado de máquina para prever a popularidade. Infelizmente, o desempenho desses métodos dependem muito da maneira com que são construídas as características, que é uma tarefa demorada e requer muito conhecimento especializado. Embora os primeiros estudos partem de modelos de regressão simples, alguns estudos recentes se concentram na modelagem da série temporal de popularidade usando processos estocásticos (MISHRA; RIZOIU; XIE, 2016). No entanto, eles usam apenas informações de séries temporais para a tarefa de previsão e ignoram outras informações valiosas que são importantes para a predição da popularidade.

Nos últimos anos, muitos pesquisadores se concentraram em explorar a combinação entre popularidade e metadados das postagens. Foi desenvolvido um método que combina as características dos metadados das postagens e dos usuários (HUANG *et al.*, 2018). Além disso foi explorado um método de classificação latente que considera as pistas visuais em imagens populares e impopulares (CAPPALLO; MENSINK; SNOEK, 2015). Embora as abordagens mencionadas acima tenham alcançado resultados promissores, elas trazem

algumas restrições, como por exemplo, a primeira abordagem considera principalmente os metadados das postagens e informações dos usuários, sem levar em consideração as imagens. O último método basicamente prevê a popularidade encontrando pistas visuais das imagens, mas negligencia o fato de pessoas famosas sempre ganharem uma alta popularidade. Com o desenvolvimento do aprendizado profundo, novos avanços foram feitos na extração de características e construção de modelos, usando mecanismos de auto-atenção (LIN *et al.*, 2022).

1.2 Justificativa e motivação

A postagem de conteúdos em mídias sociais e a iteração entre os usuários muitas vezes fazem com que determinados conteúdos tenham um impacto muito grande na Internet. A popularidade de conteúdos das mídias sociais pode ajudar órgãos privados e governamentais a entenderem o interesse público em determinados assuntos. Dessa forma, modelar e prever a popularidade de postagens tornou-se um tópico importante de pesquisa em análise de mídias sociais. Alguns exemplos que englobam esse tópico são: apoio a gestão de emergência, impacto de desastres naturais, terrorismos e crimes (POHL; BOUCHACHIA; HELLWAGNER, 2016); engajamento de determinadas marcas no Twitter (YAISAWAS *et al.*, 2018) e assistência governamental à petições online (SUBRAMANIAN; BALDWIN; COHN, 2018).

Em 2023, o Instagram tem mais de 1 bilhão de usuários ativos e, entre esses usuários, mais de 90% seguem pelo menos uma empresa na plataforma, o que significa que os usuários gostariam de ouvir as marcas. Pela mesma razão, o Instagram se tornou uma plataforma favorita para o marketing de influenciadores. Ao longo dos anos, devido ao aumento da conectividade com a internet em todo o mundo, imagens e vídeos têm contribuído para o *big data*, além da tradicional informação textual.

Os trabalhos atuais da literatura científica abrangem diferentes modalidades de características que buscam construir um modelo de predição de popularidade de postagens em mídias sociais que seja robusto. Isso nos motivou a explorar diferentes modalidades e diferentes metodologias com o objetivo de avaliar quais características geram maior engajamento nas postagens e consequentemente melhoram a predição da popularidade. Além disso, avaliar o quanto o aumento da complexidade das características afetará o modelo considerando as métricas definidas neste trabalho.

1.3 Questões de Pesquisa e Objetivos

O objetivo deste trabalho é propor, implementar e avaliar uma estrutura que extrai as características visuais-textuais das mídias sociais para a previsão da popularidade de postagens em mídias sociais. Mais especificamente, o *framework* contém três procedimentos, incluindo extração de características visual-textuais, fusão de características e a modelagem.

A entrada dos dados é um conjunto de valores, tais como, número de *likes*, comentários, *labels* de imagens e *captions* das postagens. Depois disso, fundimos algumas características e utilizamos um modelo de aprendizado de máquina para a previsão da popularidade. Diante dos desafios e problemas atualmente enfrentados em modelagens de predições de postagens em mídias sociais, foi elaborada a seguinte questão de pesquisa que norteará este projeto:

Q1: *A construção de um algoritmo para predição de postagens em mídias sociais com características multimodais, utilizando aprendizado de máquina, possui a capacidade de aprender e generalizar para outras postagens nunca antes vistas pelo modelo?*

Diante desta questão de pesquisa, são definidos os seguintes objetivos para o desenvolvimento deste trabalho:

- Mapear algoritmos de disponíveis na literatura, analisando suas lacunas e desempenhos de predições para popularidade de imagens, considerando as características e as métricas de engajamento mais utilizadas.
- Comparar os desempenhos obtidos a algoritmos de classificação do estado da arte utilizando diferentes métricas de avaliação do modelo.
- Avaliar quais características são mais importantes para a predição da popularidade.
- Discutir os resultados em nível de classificação das postagens.

A partir do modelo proposto, espera-se que o desempenho do algoritmo proposto tenha alta acurácia, baixo tempo computacional e precisão comparável ao estado da arte.

1.4 Organização do Trabalho

Este trabalho está distribuído em 5 capítulos, incluindo esta introdução, dispostos conforme a descrição que segue:

Capítulo 1: Apresenta uma introdução uma breve contextualização do problema, justificativa e motivação do trabalho, e por fim as questões de pesquisa.

Capítulo 2: Nesta parte, o sistema concebido é abordado de forma teórica apresentando os problemas encontrados na predição e classificação de postagens em mídias sociais e os modelos mais atuais da literatura.

Capítulo 3: Descreve os materiais utilizados, bem como todos os métodos utilizados para desenvolver a solução. Todas as etapas de desenvolvimento do projeto são detalhadas.

Capítulo 4: Os resultados do modelo desenvolvido são apresentados.

Capítulo 5: São apresentadas as conclusões do trabalho e sugestões de trabalhos futuros.

2 FUNDAMENTOS E TRABALHOS RELACIONADOS

2.1 Extração de Características visuais

As imagens das postagens em mídias sociais possuem uma variedade de estilos. Para utilizarmos estas imagens como parte das características para o modelo devemos extrair suas características. Para isso existem diversos métodos, dentro eles incluem estatísticas de ordem inferior, como os descritores de forma (MA *et al.*, 2010; KIM; KIM, 2000; CHEN *et al.*, 2019b), estatísticas de ordem superior, como características morfológicas (TADEJKO; RAKOWSKI, 2007) e também os descritores de textura (LOWE, 2004; OJALA; PIETIKäINEN; HARWOOD, 1996). Esses métodos de extração de características pertencem ao que é frequentemente chamado de descritores manuais. Eles são chamados dessa forma porque os algoritmos são projetados por pesquisadores para detectar características específicas consideradas importantes na análise de imagens. Estes descritores normalmente extraem características específicas para as quais foram desenvolvidas, e não são generalizáveis para todos os tipos de aplicações.

Além das características manuais, foram desenvolvidas técnicas de redes neurais que aprendem as características das imagens de forma automática. Essa classe de algoritmos também é amplamente usada para predição de popularidade (CHEN *et al.*, 2019a; QIAN *et al.*, 2022), mas tende a ser limitada em poder porque depende muito do conjunto de dados usado para treinamento. Esse problema pode ser superado treinando um conjunto de dados muito grande contendo um amplo conjunto de imagens para que o sistema aprenda uma ampla variedade de padrões diferentes. Dessa forma, as características aprendidas tornam-se independentes de qualquer conjunto de dados específico e podem ser considerados como extratores de características gerais. Assim como os descritores manuais mencionados acima, essas características aprendidas via redes neurais podem ser utilizadas sozinhas ou em combinação com outros conjuntos de características (por exemplo, descritores manuais). Alguns exemplos nesse sentido incluem a predição de mídias sociais utilizando *Hu-moments* como extrator de características visuais em um modelo multimodal (CHEN *et al.*, 2019b) e o modelo ResNet-50 como extrator de características das imagens também para a mesma finalidade (XU *et al.*, 2020).

2.2 Extração de características textuais

As características de texto correspondem às informações de texto no conteúdo da informação, incluindo por exemplo, o conteúdo da postagem, ou também, título da postagem, tópicos e categorias que o acompanham, etc.

As características do contexto e categóricas no texto podem ser extraídas usando

diferentes métodos.

Para extrair as características do contexto, os métodos mais comuns são:

- Modelos baseados em frequência de palavras, que incluem os métodos TF-IDF (*Term Frequency–Inverse Document Frequency*) (SALTON; BUCKLEY, 1988) e LDA (*Latent Dirichlet Allocation*) (BLEI; NG; JORDAN, 2003), são comumente utilizados. Cientistas desenvolveram um método no qual foi utilizado a matriz TF-IDF juntamente com o método *Princiapl Component Analysis* (PCA) com o objetivo de reduzir a dimensionalidade. Além disso, treinaram um modelo LDA com 20 tópicos para extrair as características textuais (WANG *et al.*, 2020a).

- Codificação *Word-Embedding* utilizam o método word2vec (MIKOLOV *et al.*, 2013) para treinar um modelo de rede neural com o objetivo de aprender associações de palavras a partir de um *corpus* de texto. Pesquisadores desenvolveram um método para prever a popularidade usando fusão visual-semântica com redes profundas. O modelo primeiro usa o *embedding* para calcular a média de um conjunto de palavras e, em seguida, alimenta a rede neural com os resultados agregados. As palavras são treinadas por três camadas completamente conectadas com diferentes pesos, melhorando assim as características de saída (SANJO; KATSURAI, 2017).

- Modelos de linguagem contextualizados como BERT (DEVLIN *et al.*, 2018) que extraem diretamente as informações do texto, obtendo características de alta-dimensão.

Além dos métodos citados acima, também podemos utilizar algoritmos que extraem características categóricas, por exemplo, o método *One-Hot-Enconding*. Este método usa 1 ou 0 para indicar se existe ou não a característica (HSU *et al.*, 2019). Também há os métodos diretos, no qual o algoritmo conta o tamanho do título da informação ou o número de rótulos para obter novas características do texto (XU *et al.*, 2020).

2.3 Extração de características emocionais

As características emocionais correspondem a informações de emoções referentes às postagens. Os algoritmos podem ser divididos em características de emoção visual e características de emoção de texto.

Alguns dos métodos mais comuns para a extração de características emocionais de imagens são:

- Detectores como o *Sentibank* (BORTH *et al.*, 2013) para identificar a emoção da imagem.

- VSO (*Visual Emotion Ontology*) para descobrir as emoções visuais obtidas durante a visualização da imagem. Para cada imagem, dois descritores (SentANPs e FateANPs) são extraídos para caracterizar características emocionais visuais (GELLI *et al.*, 2015).

As caracterização das emoções do texto varia de acordo com o tipo de informação. Os métodos mais comuns são:

- *SentiStrenght* (THELWALL *et al.*, 2010). É realizada uma análise de *tags* e títulos relacionados à imagem. O método gera resultados positivos (intervalo de 1 a 5) e negativos (intervalo de -1 a -5) para pontuações de emoção.
- *Tree-Net* (LI *et al.*, 2019). Este método foca nos emojis do texto para avaliar a intensidade emocional da informação do texto. O algoritmo utiliza uma rede neural profunda que aprende os mapas e prevê a intensidade do sentimento de qualquer entrada de texto com bastante precisão.

2.4 Extração de características temporais

As características temporais em geral se referem a qualquer característica associada ao tempo ou alterada ao longo do tempo. A codificação pode ser realizada em um formato de marca temporal (HSU *et al.*, 2019), onde a data de postagem é transferida em 5 subitens: horas, semana, dia, mês e ano. A codificação também pode ser feita utilizando características como: “HourInDay”, “HourInWeek”, “DayInWeek”, “DayIn-Month” e “WeekInYear” (LAI; ZHANG; ZHANG, 2020). Há também a possibilidade de criação do *timelapse*, que significa o intervalo de tempo desde a última postagem do usuário, e também o *num per time*, que seria o número de postagens que o usuário publica a cada vez (HE *et al.*, 2019).

Dentro deste tópico também podemos notar que usuários tendem a acompanhar e discutir um tópico específico por um período prolongado de tempo. Para isso, há métodos que realizaram uma média móvel deslizante (*Sliding Window Averaging*) sobre as características ordenadas no tempo, de modo que cada característica seja correlacionada com seus vizinhos anteriores. Como cerca de 75% dos usuários têm menos de 5 postagens no conjunto de dados, eles definiram o tamanho da janela abaixo de 5 (dependência de curto prazo) (WANG *et al.*, 2020b).

Também há métodos que usam o conceito de *Temporal Context Learning* que se divide em NTC (*Neighboring Temporal Context*) e PTC (*Periodic Temporal Context*), com o objetivo de aprender a coerência temporal do contexto temporal para predição. O NTC é usado para descrever flutuações de sentimento que mudam rapidamente no curto prazo. Já o PTC é um ponto descontínuo estabelecido pela série de dados anterior para indicar padrões populares cíclicos em um intervalo de tempo de longo prazo (WU *et al.*, 2017).

2.5 Extração de características dos usuários

As características do usuário correspondem às informações do usuário, que são usadas para caracterizar os dados pessoais dos usuários. Eles contêm principalmente a atividade e influência do usuário. As características digitais são frequentemente construídas por meio de classificação ou quantificação. A característica pode ser por meio da "contagem de ID do usuário" (HE *et al.*, 2019), que significa o número de vezes que cada usuário realizou uma postagem, com o objetivo de mostrar a atividade do usuário.

Também há métodos que podem quantificar a influência do usuário por seguidores e visualizações na página de um usuário ou pelo total de postagens e certificações do usuário (WANG *et al.*, 2020b; LAI; ZHANG; ZHANG, 2020; KANG *et al.*, 2019). Pode-se rastrear e selecionar informações do usuário, incluindo "*total_follower*" (o número de pessoas que seguem o usuário), "*total_following*" (o número de pessoas que o usuário segue), "*total_photos*" (o número de fotos postadas pelo usuário), "*total_views*" (o número total de visualizações das postagens do usuário) e "*total_tags*" (número total de *tags*), "*total_faves*" (o número de fotos que o usuário gosta), "*total_groups*" (o número de grupos que o usuário ingressou) e "*total_geotags*" (o número de geotags) (KANG *et al.*, 2019).

2.6 Modelos Preditivos

Após completar o passo de extração e construção das características utilizando diferentes modalidades, os métodos de previsão de popularidade de mídias sociais baseados em fusão de características requerem a integração das características resultantes e a modelagem da relação entre as características obtidas e a popularidade final para concluir a construção e o treinamento do modelo preditivo.

Os métodos de previsão de popularidade baseados em fusão de características usam modelagem com a combinação de características multimodais para obter uma previsão precisa de popularidade. Existem quatro tipos de métodos comuns:

- *Vector splice-regression training*. Neste caso os vetores de características obtidos por diferentes características multimodais são emendados e inseridos em modelos ou redes existentes para treinamento de regressão com o objetivo de obter pontuações de popularidade (HSU *et al.*, 2019; LAI; ZHANG; ZHANG, 2020; CHEN *et al.*, 2019b; HE *et al.*, 2019). Os modelos incluem os de aprendizado de máquina e redes de aprendizado profundo. Os modelos de aprendizado de máquina comumente usados para previsão de popularidade são Regressão logística (BISHOP, 2007), *CatBoost* (KANG *et al.*, 2019), *XGBoost* (CHEN; GUESTRIN, 2016) e *LightGBM* (KE *et al.*, 2017). As redes de aprendizado profundo usadas para previsão de popularidade são as DNN (*Deep Neural Networks*) (DING; WANG; WANG, 2019).

- *Joint embedding-deep learning*. Para representar características de diferentes dimensões, é necessário primeiramente combinar as características em um espaço de características *embedded*, para então treinar com uma rede neural profunda. A rede DTCN (*Deep Temporal Context Network*) foi utilizada para a predição de popularidade, usando MJE (*Multi-modal Joint Embedding*) com o objetivo de processar características visuais de alta-dimensão e características digitais dos usuários de baixa-dimensão e gerar uma representação unificada de características de alto-nível. A implementação do MJE engloba uma *Feed-foward neural network*, que contém dois canais correspondendo aos dados do usuário e imagens. Em cada canal, a rede *embedding* contém 4 camadas, das quais 2 são ocultas. O número de nós ocultos são 256 e 32, respectivamente. Para obter um mapeamento não linear de características do espaço original para o novo espaço é aplicada a função de ativação *tanh* em cada camada da rede. Também é utilizado o mecanismo de *dropout* para evitar o *overfitting*. Ao realizar a minimização da função *loss*, MJE incorpora as características do usuário e as características visuais geradas pela última camada da FNN. A representação da saída da rede *embedding* é um vetor de 64 dimensões que serve como entrada para as outras redes da arquitetura DCTN produzindo a pontuação de popularidade.

- Redes neurais profundas baseada em mecanismos de atenção. Enquanto os métodos de *joint embedding-deep learning* dá o mesmo peso para todas as características de diferentes modalidades. Já este tipo de método introduz um mecanismo de atenção com o objetivo de atribuir pesos a cada modalidade durante as fases de treinamento e inferência (XU *et al.*, 2020).

- Alinhamento de múltiplas sequências. Além das representações gerais de características vetoriais e de características tensoriais de diferentes dimensões, a extração de características também pode obter representações de características de sequência. Para esse tipo de características, podemos usar o método de alinhamento de várias sequências para modelar. O algoritmo wDTW-CD (LI *et al.*, 2019) realiza a correspondência de tempo entre o PTS (*Popularity Time Series*) e o STS (*Sentiment Time Series*), que são duas características de sequência. Com o intuito de prever a popularidade, os autores do algoritmo propuseram o modelo *Dual Autoregressive Integrated Moving Average* (*Dual ARIMA*) para considerar a relação entre PTS e STS.

3 DESENVOLVIMENTO

Neste capítulo, é apresentado o desenvolvimento deste trabalho, descrevendo a metodologia que será trabalhada para a execução dos experimentos.

O esquema geral da metodologia é apresentado na Figura 1.

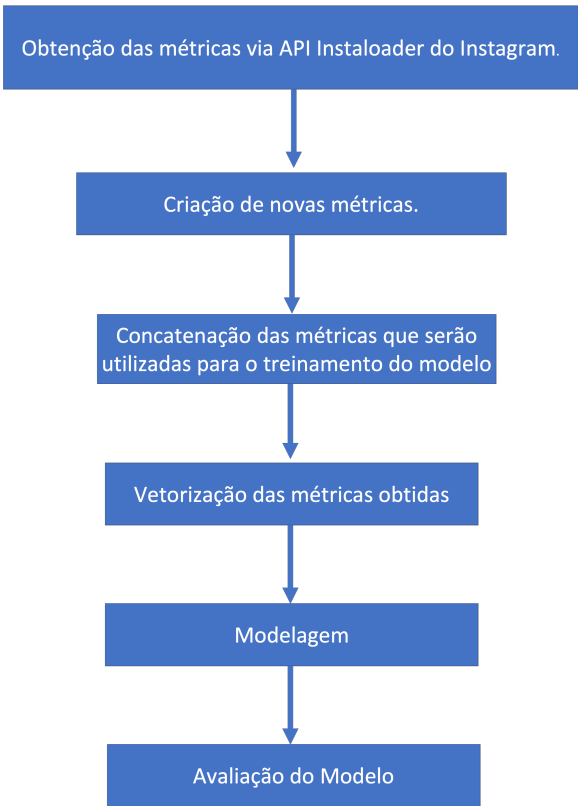


Figura 1 – Metodologia utilizada no projeto de TCC

Nas próximas sessões serão apresentados os ítens da metodologia mostrada na Figura 1.

3.1 Obtenção das métricas via API Instaloader do Instagram

A base de dados foi construída usando a API do Instagram chamada Instaloader [1]. Foi desenvolvido um código em Python com o objetivo de usar a API e baixar as métricas necessárias para o desenvolvimento do projeto.

Na Tabela 1 há características (que são os atributos) da base de dados construída. Nesta base de dados há as seguintes colunas:

- "post_url" representando o link da postagem;

[1] https://instaloader.github.io/

- “num_comments” informando o número de comentários;
- “num_likes” informando o número de likes;
- *“caption”* descrevendo o conteúdo daquela postagem;
- “image_url” contendo o link para a imagem do instagram;
- “is_video” contendo a informação se é vídeo ou imagem;
- “date” informando a data e horário da postagem;
- “caption_hashtag” descrevendo o hashtag contido no *caption* da postagem;
- “caption_mentions” informando as menções feitas naquela postagem.

Tabela 1 – Características da base de dados de postagens do Instagram

Nome	**Tipo**
post_url	url
num_comments	numérico
num_likes	numérico
caption	string
image_url	url
is_video	boleano
date	data/hora
caption_hashtag	string
caption_mentions	string

Foram escolhidos 27 perfis do Instagram com conteúdos relacionados à cirurgia plástica. Foram retirados no máximo 3000 postagens dos perfis. Ao todo foram capturadas 16421 postagens. As páginas escolhidas para a captura dos dados tinham perfis de quantidade de seguidores parecidas.

As informações capturadas de cada perfil foram salvas em um arquivo .csv. Os 27 arquivos foram concatenados para a geração de apenas um arquivo .csv no qual será utilizado para as implementações do modelo em Python.

3.2 Criação de novas métricas

A primeira métrica criada foi relacionada às imagens das postagens que obtivemos da API do Instagram. As imagens possuem características extremamente importantes para as aplicações na área de visão computacional e aprendizado de máquina. Neste trabalho, para extrairmos características das imagens, utilizamos a API do Google Vision[2]. Esta API oferece modelos pré-treinados de aprendizado de máquina com milhões de características

[2] https://cloud.google.com/vision?hl=pt-br

predefinidas. Com essa API extraimos *labels* das imagens e estes labels foram utilizados como características no modelo de aprendizado de máquina.

Também criamos uma nova métrica chamada de "cap_words" que tem como objetivo capturar palavras contidas dentro da métrica "caption" e remover toda a sua pontuação.

Outra métrica criada foi a métrica de *target*, que chamamos de *engament_score*. Esta métrica é uma medida ponderada entre o número de *likes* e comentários. Colocamos o peso em 60% para o número de *likes* e 40% para o número de comentários. A partir dessa métrica, criamos uma outra métrica binária, chamada de *engagement_high* que será identificada como 1 se a pontuação estiver acima do valor mediano da métrica de *engament_score*, caso contrário será 0. Ou seja, essa métrica rotulará os dados em 0 (não popular) ou 1 (popular). Essa métrica foi escolhida dessa forma inspirado em um *Survey* da área (SHAHID *et al.*, 2022). Além disso, observamos que a métrica *likes* é muito mais engajadora do que a métrica comentários. Por exemplo, um seguidor pode fazer um comentário desmerecendo a empresa ou a postagem, porém quando um *like* é dado em uma postagem a chance é bem maior do seguidor ter realmente gostado da postagem.

3.3 Concatenação das métricas que serão utilizadas para o treinamento do modelo

Concatenamos as métricas que serão utilizadas como o corpus da aplicação.

3.4 Vetorização das métricas obtidas

Para a vetorização das métricas obtidas, utilizamos o método TF-IDF é uma abreviação do termo *Term Frequency Inverse Document Frequency*. Este é um algoritmo muito comum para transformar texto em uma representação significativa de números que será usado como entrada no algoritmo de aprendizagem de máquina.

Foi implementado o método TF-IDF usando a biblioteca do Python *sklearn* com os parâmetros de *stop-words* em português.

3.5 Modelagem

Para a modelagem estatística foi utilizado o método de regressão logística. É um dos algoritmos de aprendizado de máquina mais simples e um dos mais usados para a classificação de duas classes.

Essa técnica tem como principal objetivo fornecer um modelo capaz de prever valores de uma variável dependente, que geralmente é binária, com base em outras variáveis independentes. Com base nesse modelo, são obtidas as probabilidades de ocorrência de um evento, dadas as variáveis aleatórias (observações).

Dado um conjunto de dados com instâncias $X = x_1, x_2, ..., x_n$, tal que n é a quantidade de instâncias, e os rótulos $Y = y_1, y_2, ..., y_n$, podendo pertencer às classes C=0,1, a função logística *sigma* e o modelo probabilístico são dados pelas equações 3.1 e 3.2.

$$\sigma(\theta) = \frac{1}{1 + e^{-\theta}} \tag{3.1}$$

$$Pr(y_n = 1|x_n) = \sigma(W.x_n) \tag{3.2}$$

Neste caso, θ é uma função linear que multiplica o vetor de pesos W ao vetor dos valores dos k atributos da instância x_n (Equação 3.3).

$$\theta = w_0 + w_1x_1 + w_2x_2 + ... + w_kx_k \tag{3.3}$$

Ao se utilizar a Regressão Logística para treinar um classificador, o procedimento consiste em ajustar a função ao conjunto de dados para aumentar a generalidade (BISHOP, 2007), iterativamente ajustando os pesos $W = w_0, w_1, ..., w_k$ utilizando a função de máxima verossimilhança descrita na equação 3.4.

$$Pr(y|x_n) = \prod_{n=1}^{N} \sigma(W.x_n)^{y_n}(1 - \sigma(W.x_n))^{(1-y_n)} \tag{3.4}$$

A função de erro (Equação 3.5) é então definida ao se pegar o logaritmo negativo da função de máxima verossimilhança, que dá a função entropia cruzada.

$$E(W) = -lnPr(y|x_n) = -\sum_{n=1}^{N}(y_nln\sigma(W.x_n) + (1 - y_n)ln(1 - \sigma(W.x_n))) \tag{3.5}$$

Em seguida, ao calcularmos o gradiente da função de erro $E(W)$ obtemos a Equação 3.6

$$\nabla E(W) = \sum_{n=1}^{N}(\sigma(W.x_n) - y_n)x_n \tag{3.6}$$

A partir de então, os pesos podem ser atualizados utilizando a função descrita a seguir, conhecida como mínimos quadrados médios, que considera que t é o tempo atual e $t + 1$ é o tempo da próxima iteração, sendo η um parâmetro a ser passado (Equação 3.7).

$$W^{t+1} = W^t - \eta\nabla E_n \tag{3.7}$$

Também implementamos outros métodos para realizarmos a comparação entre eles e também com a regressão logística. Os métodos são: Árvore de Decisão (QUINLAN, 1990), *Gradient Boosting* (AYYADEVARA, 2018), K-ésimo vizinhos mais próximos (MALEKI; ZEINALI; NIAKI, 2021), Rede Perceptron Multi-Camadas (TAUD; MAS, 2018) e *Random Forest* (CUTLER; CUTLER; STEVENS, 2012).

3.6 Avaliação do Modelo

Os critérios de avaliação utilizados neste trabalho são: validação cruzada (ZHANG; LIU, 2023), acurácia, área sob a curva ROC, F1 *score* e matriz de confusão.

Para a validação cruzada utilizamos o método *Repeated K-Fold cross validator* da biblioteca *sklearn* do Python. Este método repete o k-fold n vezes com randomização diferente em cada iteração. Utilizamos 10 subdivisões com 3 repetições em cada subdivisão.

A fórmula para o cálculo da acurácia é feita pela Equação 3.8.

$$\frac{VP + VN}{VP + VN + FP + FN} \tag{3.8}$$

onde VP indica os verdadeiros positivos, VN os verdadeiros negativos, FP os falsos positivos e FN os falsos negativos.

A área sob a curva ROC é derivada da curva ROC. A curva ROC é utilizada para classificação e possui dois parâmetros: taxa de verdadeiros positivos e taxa de falsos positivos. Uma curva ROC traça VP $\times$ FP em diferentes limiares de classificação.

Assim, para simplificar a análise da curva ROC, a área sob a curva é que uma maneira de resumir a curva ROC em um único valor com o objetivo de agregar todos os limiares da ROC. O valor da área sob a curva ROC varia entre 0 e 1, e quanto maior o valor, melhor se apresenta o modelo desenvolvido.

Já o F1 *score* é uma métrica que une a precisão e a revocação com o objetivo de trazer um único número que determine a qualidade do modelo. A equação 3.9 mostra a equação do F1 *score*.

$$F1 = \frac{2 * precisao * revocacoo}{precisao + revocacao} \tag{3.9}$$

A precisão (Equação 3.10) verifica de todos os dados classificados como positivos, quantos são realmente positivos.

$$Precisao = \frac{VP}{VP + FP} \tag{3.10}$$

A equação de revocação (Equação 3.11) nos mostra qual a porcentagem de dados classificados como positivos comparado com a quantidade real de positivos que existem em nossa amostra.

$$revocacacao = \frac{VP}{VP + FN} \tag{3.11}$$

A matriz de confusão é uma tabela que permite a visualização do desempenho de um algoritmo de classificação. Uma amostra da matriz de confusão pode ser vista na Tabela 2.

		Classificação atual	
		Não Popular	Popular
Classificação Prevista	Não Popular	VN	FP
	Popular	FN	VP

Tabela 2 – Matriz de confusão

4 RESULTADOS

Os resultados da metodologia proposta são apresentados neste Capítulo. Os experimentos foram conduzidos em um computador processador i7, 32 GB de RAM e uma placa de GPU GeForce GTX Titan XP. O código foi escrito em linguagem Python, usando a biblioteca *sklearn*.

Os resultados são divididos em seções, apresentando os diferentes métodos de classificação utilizados neste trabalho. Em cada seção são mostrados os resultados utilizando três diferentes formas de treinamento: 1) Utilizando dados apenas dos *captions* das imagens (chamaremos de Treinamento 1); 2) Utilizando dados apenas dos *labels* das imagens (chamaremos neste texto de Treinamento 2); 3) Utilizando dados de *captions* das postagens juntamente com os *labels* das imagens (chamaremos de Treinamento 3). Os três diferentes tipos de treinamentos foram realizados com o objetivo de medirmos a relevância da fusão de características.

Os métodos de aprendizado de máquina implementados foram os seguintes: Árvore de Decisão, *Gradient Boosting*, K-ésimo vizinhos mais próximos, Rede Perceptron Multi-Camadas, *Random Forest* e Regressão Logística. A classificação é feita por duas classes: 0 significa que a postagem não é popular e 1 alta popularidade.

4.1 Validação Cruzada K-Fold

Primeiramente analisamos o método *Repeated K-Fold* com 10 subdivisões e número de repetições $n = 3$ para avaliarmos a generalização do modelo. A Tabela 3 mostra os resultados do k-fold para os diferentes tipos de treinamento utilizando as seis metodologias de aprendizado de máquina descritas anteriormente. Podemos notar pela Tabela 3 que os melhores resultados foram para os treinamentos 1 e 3 com as metodologias *Random Forest* e Regressão Logística.

Tabela 3 – Acurácia média para o *Repeated K-Fold* e seus respectivos desvio-padrão

Metodologias	**Treinamento 1**	**Treinamento 2**	**Treinamento 3**
Árvore de Decisão	0.647 (0.009)	0.520 (0.014)	0.648 (0.012)
Gradient Boosting	0.698 (0.009)	0.524 (0.014)	0.699 (0.009)
K-ésimo vizinho mais próximo	0.567 (0.011)	0.512 (0.013)	0.562 (0.013)
Rede Perceptron Multi-Camadas	0.679 (0.011)	0.522 (0.016)	0.680 (0.013)
Random Forest	**0.729 (0.009)**	0.525 (0.014)	**0.728 (0.010)**
Regressão Logística	0.725 (0.011)	0.525 (0.014)	**0.728 (0.011)**

4.2 Cálculo das métricas

Para a realização do treinamento, dividimos o conjunto de dados escolhido em treinamento (80% dos dados) e teste (20% dos dados), de acordo com a métrica descrita na Seção 3.6.

As Tabelas 4, 5 e 6 mostram as métricas calculadas para os diferentes modelos de treinamento usando as diferentes metodologias de aprendizado de máquina.

Tabela 4 – Acurácia utilizando diferentes modelos de treinamento e a diferentes metodologias de aprendizado de máquina

Metodologias	Treinamento 1	Treinamento 2	Treinamento 3
Árvore de Decisão	0.6526	0.5129	0.6590
Gradient Boosting	0.7022	0.5156	0.7135
K-ésimo vizinho mais próximo	0.5683	0.5028	0.5659
Rede Perceptron Multi-Camadas	0.6794	0.5165	0.6837
Random Forest	**0.7485**	0.5175	**0.7458**
Regressão Logística	**0.7488**	0.5178	**0.7464**

Tabela 5 – AUC ROC utilizando diferentes modelos de treinamento e a diferentes metodologias de aprendizado de máquina

Metodologias	Treinamento 1	Treinamento 2	Treinamento 3
Árvore de Decisão	0.6526	0.52	0.6591
Gradient Boosting	0.7864	0.5263	0.7905
K-ésimo vizinho mais próximo	0.6170	0.5140	0.5909
Rede Perceptron Multi-Camadas	0.7392	0.5264	0.7459
Random Forest	**0.8192**	0.5299	**0.8156**
Regressão Logística	**0.8196**	0.5287	**0.8198**

Tabela 6 – F1 score utilizando diferentes modelos de treinamento e a diferentes metodologias de aprendizado de máquina

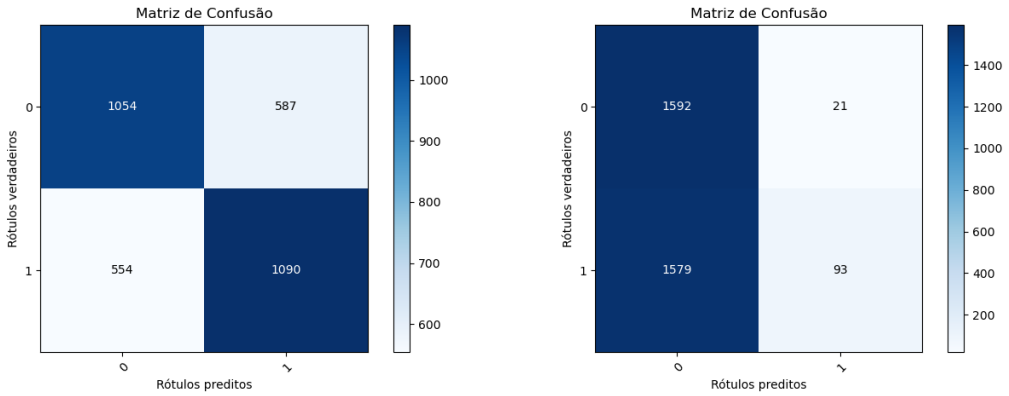
Metodologias	Treinamento 1	Treinamento 2	Treinamento 3
Árvore de Decisão	0.6564	0.1041	0.6591
Gradient Boosting	0.6711	0.1253	0.6881
K-ésimo vizinho mais próximo	0.6905	0.07053	0.5702
Rede Perceptron Multi-Camadas	0.6887	0.1207	0.6807
Random Forest	**0.7531**	0.1324	**0.7439**
Regressão Logística	**0.7487**	0.1334	**0.7432**

Analisando os resultados, notamos que as metodologias de Regressão Logística e *Random Forest* apresentaram os melhores resultados dos experimentos para todas as métricas calculadas. Podemos notar que não houve uma melhora significativa no treinamento 3, que é a fusão das características de *captions* e *labels* das imagens comparado ao treinamento 1, que são apenas características dos *captions* das imagens.

4.3 Matrizes de Confusão

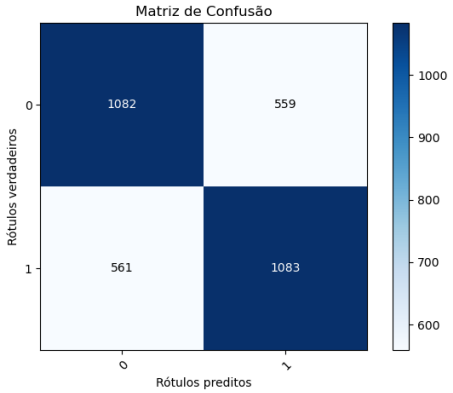
4.3.1 Árvore de Decisão

A Figura 2 mostra as matrizes de confusão para os diferentes tipos de treinamento (Treinamentos 1, 2 e 3) usando o método de classificação Árvore de Decisão.



(a) Treinamento 1: Usando apenas os *captions* das imagens

(b) Treinamento 2: Usando apenas os *labels* das imagens



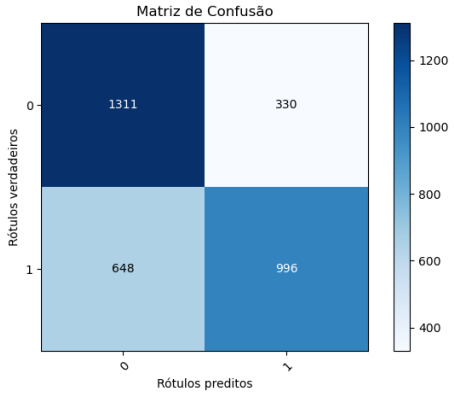
(c) Treinamento 3: Usando *caption* e *labels* das imagens

Figura 2 – Matrizes de Confusão usando o método Árvore de Decisão para diferentes tipos de treinamento

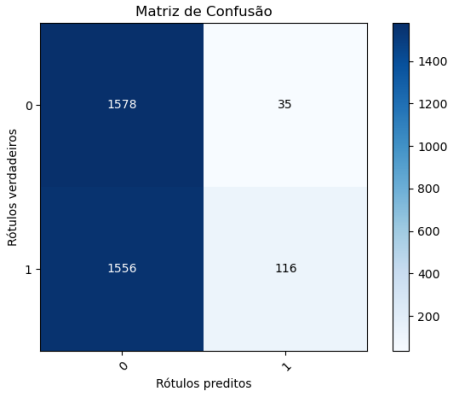
Analisando a Figura 2, podemos notar que os acertos para os 1 e 3 foram bem próximos. Enquanto o treinamento 1 acertou 1054 para a classe 0, o treinamento 3 acertou 1082. Já para a classe 1, o treinamento 1 acertou 1090 amostras, enquanto o treinamento 3, 1083 amostras. Por fim, avaliando o treinamento 2, percebemos uma boa quantidade de acerto para a classe 0, porém errou quase todas as amostras para a classe 1.

4.3.2 *Gradient Boosting*

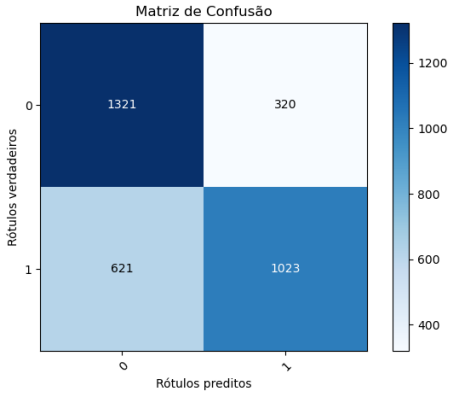
A Figura 3 mostra as matrizes de confusão para os diferentes modelos de treinamento usando o método de *Gradient Boosting*.



(a) Treinamento 1: Utilizando apenas os *captions* das imagens



(b) Treinamento 2: Utilizando apenas os *labels* das imagens



(c) Treinamento 3: Utilizando *caption* e *labels* das imagens

Figura 3 – Matrizes de Confusão usando o método *Gradient Boosting* para diferentes tipos de treinamento

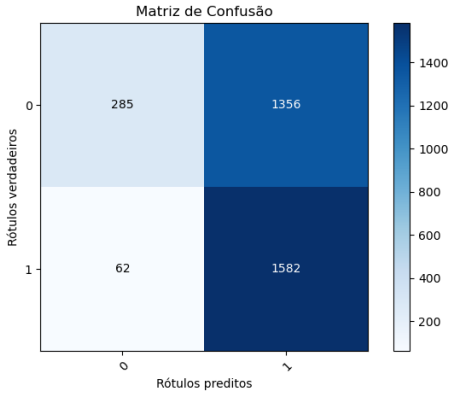
Analisando a Figura 3, podemos notar que o treinamento 3 mostrou-se melhor do que os treinamentos 1 e 2. Das 1641 amostras para a classe 0, o modelo acertou 1321 (80% do total de amostras da classe 0). Já para a classe 1, das 1644 amostras, o modelo acertou 1023 (62.22% do total), gerando uma acurácia, ROC AUC e F1 score um pouco melhor do que os outros treinamentos (vide Tabelas 4, 5 e 6, linha 2).

Comparando o treinamento 1 com o treinamento 3 em termos de falsos negativos, o treinamento 3 possui menos falsos negativos do que o treinamento 1. Analisando também os falsos positivos, podemos perceber que o treinamento 3 também apresenta falsos positivos do que o treinamento 1.

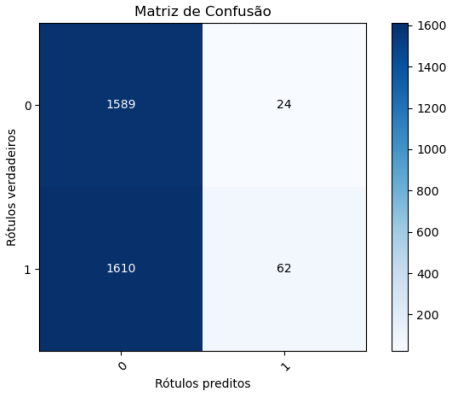
4.3.3 K-ésimo vizinho mais próximos

Foi implementado o número de vizinhos $k = 5$ para este experimento.

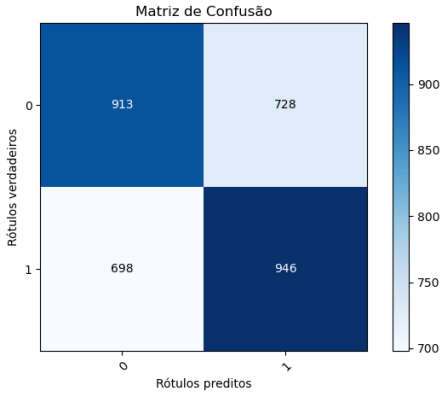
A Figura 4 mostra as matrizes de confusão para os diferentes modelos de treinamento usando o método de k-ésimo vizinhos mais próximos.



(a) Treinamento 1: Usando apenas os *captions* das imagens



(b) Treinamento 2: Utilizando apenas os *labels* das imagens



(c) Treinamento 3: Usando *caption* e *labels* das imagens

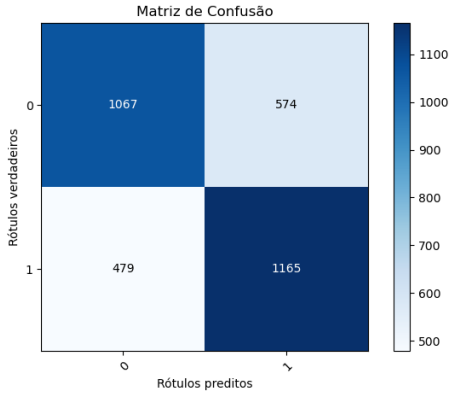
Figura 4 – Matrizes de Confusão usando o método K-vizinhos mais próximos para diferentes tipos de treinamento

Analisando a Figura 4, podemos notar que para o treinamento 3 houveram mais acertos do que erros comparado aos treinamentos 1 e 2. Apesar das métricas de acurácia, AUC ROC e F1 score (4, 5 e 6, linha 3) estarem bem próximas para os treinamentos 1 e 3, podemos observar que para o treinamento 1 - classe 0 houveram muito mais erros do que acertos.

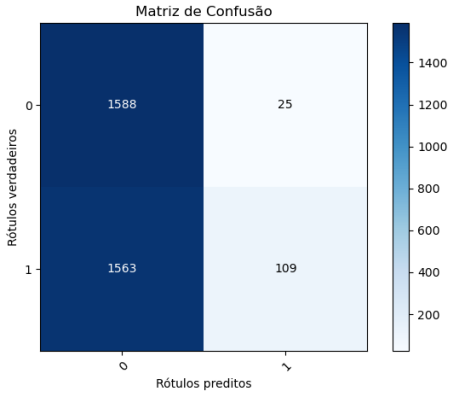
4.3.4 Rede Perceptron Multi-Camadas

A Figura 5 mostra as matrizes de confusão para os diferentes modelos de treinamento usando o método de Rede Perceptron Multi-Camadas.

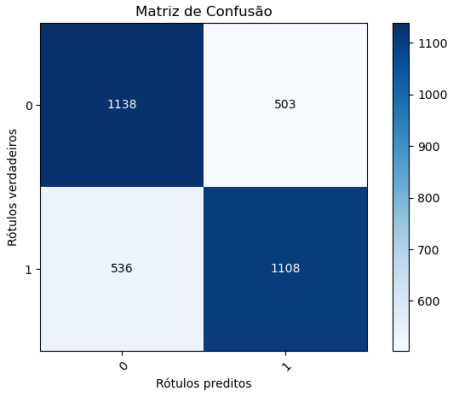
Observando a Figura 5, para a classe 0, o treinamento 3 apresentou mais acertos do que os treinamentos 1 e 2. Já para a classe 1, o treinamento 1 teve melhores resultados do que os outros tipos de treinamento.



(a) Treinamento 1: Utilizando apenas os *captions* das imagens



(b) Treinamento 2: Utilizando apenas os *labels* das imagens



(c) Treinamento 3: Utilizando *caption* e *labels* das imagens

Figura 5 – Matrizes de Confusão usando o método Rede Perceptron Multi-Camadas para diferentes tipos de treinamento

4.3.5 *Random Forest*

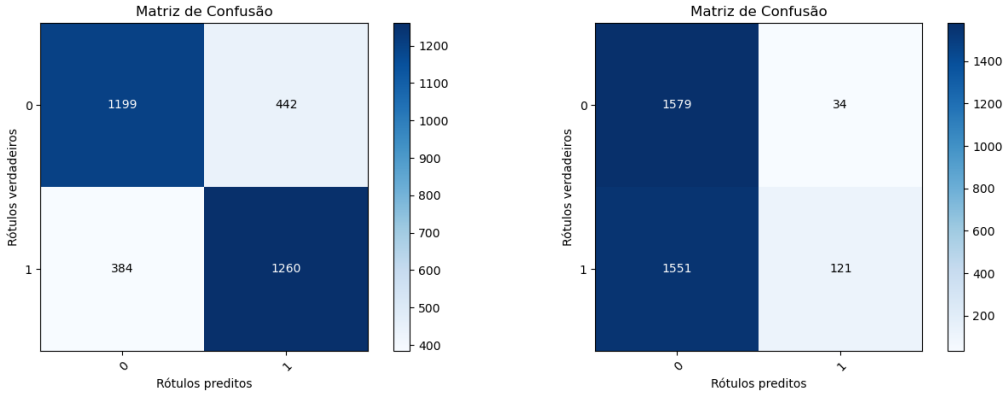
A Figura 5 mostra as matrizes de confusão para os diferentes modelos de treinamento usando o método *Random Forest*.

Observando a Figura 6 podemos notar que para a classe 0 o treinamento 2 apresentou melhores resultados, porém não acertou quase nenhuma amostra para a classe 1. Comparando os treinamentos 1 e 3, a classe 0 apresentou uma quantidade maior de acertos para o treinamento 3, enquanto que a classe 1 obteve melhores resultados para o treinamento 1.

4.3.6 Regressão Logística

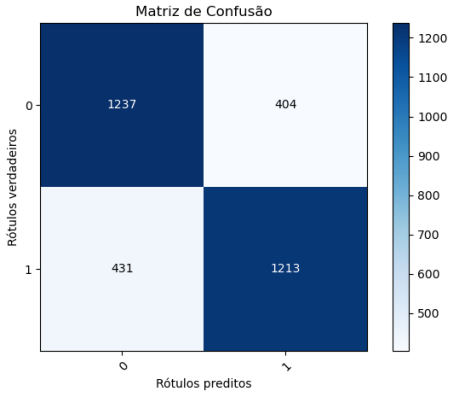
A Figura 7 mostra as matrizes de confusão para os diferentes modelos de treinamento usando o método Regressão Logística.

A Figura 7 mostra resultados parecidos com as matrizes de confusão da Figura 6, ou seja, para a classe 0 o treinamento 2 apresentou melhores resultados, porém não



(a) Treinamento 1: Utilizando apenas os *captions* das imagens

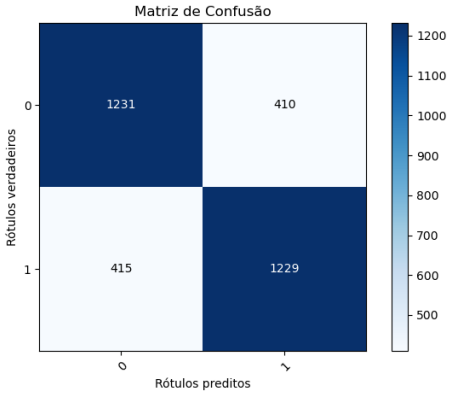
(b) Treinamento 2: Utilizando apenas os *labels* das imagens



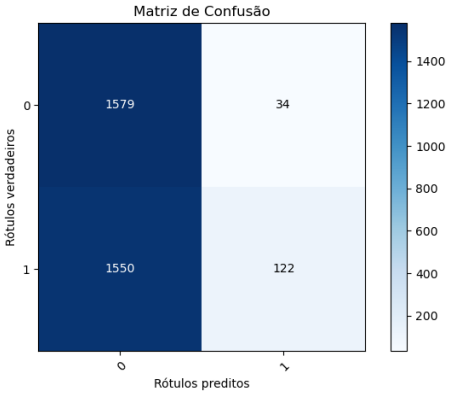
(c) Treinamento 3: Utilizando *caption* e *labels* das imagens

Figura 6 – Matrizes de Confusão usando o método *Random Forest* para diferentes tipos de treinamento

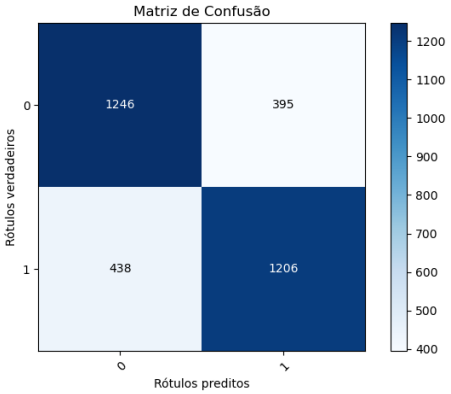
acertou quase nenhuma amostra para a classe 1. Comparando os treinamentos 1 e 3, a classe 0 apresentou uma quantidade maior de acertos para o treinamento 3, enquanto que a classe 1 obteve melhores resultados para o treinamento 1. Observando também as Tabelas (4, 5 e 6, linhas 5 e 6 podemos observar que os resultados para as metodologias *Random Forest* e Regressão Logística estão muito próximos.



(a) Treinamento 1: Utilizando apenas os *captions* das imagens



(b) Treinamento 2: Usando apenas os *labels* das imagens



(c) Treinamento 3: Utilizando *caption* e *labels* das imagens

Figura 7 – Matrizes de Confusão usando o método Regressão Logística para diferentes tipos de treinamento

5 CONCLUSÕES

Este trabalho foi um estudo para predizer a popularidade de postagens no Instagram na área de cirurgia plástica. Foram testadas seis diferentes métodos de aprendizado de máquina utilizando 3 tipos de treinamentos diferentes.

Os treinamentos utilizados neste trabalho foram:

- Treinamento 1: Usando apenas os *captions* das imagens.
- Treinamento 2: Usando apenas os *labels* das imagens.
- Treinamento 3: Usando os *labels* e os *captions* das imagens

As seis metodologias utilizadas neste trabalho foram:

- Árvore de Decisão
- *Gradient Boosting*
- K-ésimo vizinhos mais próximos
- Rede Perceptron Multi-Camadas
- *Random Forest*
- Regressão Logística.

Dentre as metodologias utilizadas, a *Random Forest* e a Regressão Logística mostraram melhores resultados. A acurácia mais alta foi 74% (treinamentos 1 e 3 para ambos os métodos de aprendizado de máquina). A métrica ROC AUC mais alta foi de 0.81 (treinamentos 1 e 3 para ambos os métodos de aprendizado de máquina). E por fim, o F1 Score mais alto foi de 0.75 para o método *Random Forest* (treinamento 1).

Para o método *Gradient Boosting* (terceiro melhor método em termos de acurácia e AUC ROC para os treinamentos 1 e 3) notamos que foi o único que apresentou valores mais baixos para falsos negativos e falsos positivos para o treinamento 3 quando comparado ao treinamento 1 do *Gradient Boosting*. Já comparando este método com todos os outros métodos estudados, vimos que ele apresentou menor número de falsos positivos (treinamento 3) do que todos os outros métodos analisados. Falsos positivos significa que a postagem seria ruim em termos de engajamento a longo prazo, pois o engajamento não seria sustentado, uma vez que os seguidores perceberão que a postagem não é relevante para eles.

Também notamos que o método *Random Forest* foi o método que apresentou menor número de falsos negativos (treinamento 1) do que todos os outros métodos estudados. Falsos negativos significa que uma postagem com alta relevância não seria postada pois o modelo errou neste caso.

No caso de postagens no Instagram, é interessante diminuir tanto os casos de falsos negativos quanto de falsos positivos.

Percebemos que a fusão de características (treinamento 3) teve resultados semelhantes ao treinamento 1 (utilizando apenas os *captions* das imagens). Ou seja, os *labels* das imagens não acrescentaram grande importância para o modelo de aprendizado de máquina e as *captions* foram características importantes para o aprendizado de acordo com a métrica de engajamento estipulada.

Como trabalho futuro pretendemos expandir nossos dados, utilizando outras APi's para captura de dados do Instagram, pois a APi Instaloader é bem limitada e com aquisição de poucas características. Dessa forma a fusão de uma grande quantidade de características poderá ser aplicada para análise da melhora de desempenho do método.

A principal limitação do trabalho é que os dados precisam ser atualizados ao longo do tempo. Todos os dias há milhares de postagens no Instagram, pois quanto mais postagens, mais engajamento há entre os influenciadores e seus seguidores. Para que o modelo continue robusto, há a necessidade da atualização do treinamento ao longo do tempo.

REFERÊNCIAS

AYYADEVARA, V. K. Gradient boosting machine. *In*: _____. **Pro Machine Learning Algorithms : A Hands-On Approach to Implementing Algorithms in Python and R**. Berkeley, CA: Apress, 2018. p. 117–134. ISBN 978-1-4842-3564-5. Available at: https://doi.org/10.1007/978-1-4842-3564-5_6.

BISHOP, C. M. **Pattern Recognition and Machine Learning (Information Science and Statistics)**. 1. ed. [*S.l.: s.n.*]: Springer, 2007. ISBN 0387310738.

BLEI, D. M.; NG, A. Y.; JORDAN, M. I. Latent dirichlet allocation. **Journal of machine Learning research**, v. 3, n. Jan, p. 993–1022, 2003.

BORTH, D. *et al.* Large-scale visual sentiment ontology and detectors using adjective noun pairs. *In*: **Proceedings of the 21st ACM International Conference on Multimedia**. New York, NY, USA: Association for Computing Machinery, 2013. (MM '13), p. 223–232. ISBN 9781450324045. Available at: https://doi.org/10.1145/2502081.2502282.

CAPPALLO, S.; MENSINK, T.; SNOEK, C. G. Latent factors of visual popularity prediction. *In*: **Proceedings of the 5th ACM on International Conference on Multimedia Retrieval**. New York, NY, USA: Association for Computing Machinery, 2015. (ICMR '15), p. 195–202. ISBN 9781450332743. Available at: https://doi.org/10.1145/2671188.2749405.

CHEN, G. *et al.* Npp: A neural popularity prediction model for social media content. **Neurocomputing**, v. 333, p. 221–230, 2019. ISSN 0925-2312. Available at: https://www.sciencedirect.com/science/article/pii/S0925231218314942.

CHEN, J. *et al.* Social media popularity prediction based on visual-textual features with xgboost. *In*: **Proceedings of the 27th ACM International Conference on Multimedia**. New York, NY, USA: Association for Computing Machinery, 2019. (MM '19), p. 2692–2696. ISBN 9781450368896. Available at: https://doi.org/10.1145/3343031.3356072.

CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. **CoRR**, abs/1603.02754, 2016. Available at: http://arxiv.org/abs/1603.02754.

CUTLER, A.; CUTLER, D. R.; STEVENS, J. R. Random forests. *In*: _____. **Ensemble Machine Learning: Methods and Applications**. New York, NY: Springer New York, 2012. p. 157–175. ISBN 978-1-4419-9326-7. Available at: https://doi.org/10.1007/978-1-4419-9326-7_5.

DEVLIN, J. *et al.* BERT: pre-training of deep bidirectional transformers for language understanding. **CoRR**, abs/1810.04805, 2018. Available at: http://arxiv.org/abs/1810.04805.

DING, K.; WANG, R.; WANG, S. Social media popularity prediction: A multiple feature fusion approach with deep neural networks. *In*: **Proceedings of the 27th ACM International Conference on Multimedia**. New York, NY, USA: Association for Computing Machinery, 2019. (MM '19), p. 2682–2686. ISBN 9781450368896. Available at: https://doi.org/10.1145/3343031.3356062.

GELLI, F. *et al.* Image popularity prediction in social media using sentiment and context features. *In*: **Proceedings of the 23rd ACM International Conference on Multimedia**. New York, NY, USA: Association for Computing Machinery, 2015. (MM '15), p. 907–910. ISBN 9781450334594. Available at: https://doi.org/10.1145/2733373.2806361.

HE, Z. *et al.* Feature construction for posts and users combined with lightgbm for social media popularity prediction. *In*: **Proceedings of the 27th ACM International Conference on Multimedia**. New York, NY, USA: Association for Computing Machinery, 2019. (MM '19), p. 2672–2676. ISBN 9781450368896. Available at: https://doi.org/10.1145/3343031.3356054.

HESSEL, J.; LEE, L.; MIMNO, D. M. Cats and captions vs. creators and the clock: Comparing multimodal content to context in predicting relative popularity. **CoRR**, abs/1703.01725, 2017. Available at: http://arxiv.org/abs/1703.01725.

HSU, C.-C. *et al.* Popularity prediction of social media based on multi-modal feature mining. *In*: **Proceedings of the 27th ACM International Conference on Multimedia**. New York, NY, USA: Association for Computing Machinery, 2019. (MM '19), p. 2687–2691. ISBN 9781450368896. Available at: https://doi.org/10.1145/3343031.3356064.

HUANG, F. *et al.* Random forest exploiting post-related and user-related features for social media popularity prediction. *In*: **Proceedings of the 26th ACM International Conference on Multimedia**. New York, NY, USA: Association for Computing Machinery, 2018. (MM '18), p. 2013–2017. ISBN 9781450356657. Available at: https://doi.org/10.1145/3240508.3266439.

KANG, P. *et al.* Catboost-based framework with additional user information for social media popularity prediction. *In*: **Proceedings of the 27th ACM International Conference on Multimedia**. New York, NY, USA: Association for Computing Machinery, 2019. (MM '19), p. 2677–2681. ISBN 9781450368896. Available at: https://doi.org/10.1145/3343031.3356060.

KE, G. *et al.* Lightgbm: A highly efficient gradient boosting decision tree. *In*: GUYON, I. *et al.* (ed.). **Advances in Neural Information Processing Systems**. Curran Associates, Inc., 2017. v. 30. Available at: https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

KIM, W.-Y.; KIM, Y.-S. A region-based shape descriptor using zernike moments. **Signal Processing: Image Communication**, v. 16, n. 1, p. 95–102, 2000. ISSN 0923-5965. Available at: https://www.sciencedirect.com/science/article/pii/S0923596500000199.

LAI, X.; ZHANG, Y.; ZHANG, W. Hyfea: Winning solution to social media popularity prediction for multimedia grand challenge 2020. *In*: **Proceedings of the 28th ACM International Conference on Multimedia**. New York, NY, USA: Association for Computing Machinery, 2020. (MM '20), p. 4565–4569. ISBN 9781450379885. Available at: https://doi.org/10.1145/3394171.3416273.

LI, J. *et al.* Senti2pop: Sentiment-aware topic popularity prediction on social media. *In*: **2019 IEEE International Conference on Data Mining (ICDM)**. [*S.l.: s.n.*], 2019. p. 1174–1179.

LIN, H.-H. *et al.* Social media popularity prediction based on multi-modal self-attention mechanisms. **IEEE Access**, v. 10, p. 4448–4455, 2022.

LOWE, D. G. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision**, v. 60, 2004. Available at: https://doi.org/10.1023/B:VISI.0000029664.99615.94.

MA, Y. *et al.* A novel shape feature to classify microcalcifications. *In*: **2010 IEEE International Conference on Image Processing**. [*S.l.: s.n.*], 2010. p. 2265–2268.

MALEKI, N.; ZEINALI, Y.; NIAKI, S. T. A. A k-nn method for lung cancer prognosis with the use of a genetic algorithm for feature selection. **Expert Systems with Applications**, v. 164, p. 113981, 2021. ISSN 0957-4174. Available at: https://www.sciencedirect.com/science/article/pii/S0957417420307594.

MIKOLOV, T. *et al.* Distributed representations of words and phrases and their compositionality. **Advances in neural information processing systems**, v. 26, 2013.

MISHRA, S.; RIZOIU, M.; XIE, L. Feature driven and point process approaches for popularity prediction. **CoRR**, abs/1608.04862, 2016. Available at: http://arxiv.org/abs/1608.04862.

OJALA, T.; PIETIKäINEN, M.; HARWOOD, D. A comparative study of texture measures with classification based on featured distributions. **Pattern Recognition**, v. 29, n. 1, p. 51–59, 1996. ISSN 0031-3203. Available at: https://www.sciencedirect.com/science/article/pii/0031320395000674.

POHL, D.; BOUCHACHIA, A.; HELLWAGNER, H. Online indexing and clustering of social media data for emergency management. **Neurocomputing**, v. 172, p. 168–179, 2016. ISSN 0925-2312. Available at: https://www.sciencedirect.com/science/article/pii/S092523121500613X.

QIAN, Y. *et al.* Popularity prediction for marketer-generated content: A text-guided attention neural network for multi-modal feature fusion. **Information Processing & Management**, v. 59, n. 4, p. 102984, 2022. ISSN 0306-4573. Available at: https://www.sciencedirect.com/science/article/pii/S0306457322001005.

QUINLAN, J. Decision trees and decision-making. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 20, n. 2, p. 339–346, 1990.

SALTON, G.; BUCKLEY, C. Term-weighting approaches in automatic text retrieval. **Information Processing & Management**, v. 24, n. 5, p. 513–523, 1988. ISSN 0306-4573. Available at: https://www.sciencedirect.com/science/article/pii/0306457388900210.

SANJO, S.; KATSURAI, M. Recipe popularity prediction with deep visual-semantic fusion. *In*: **Proceedings of the 2017 ACM on Conference on Information and Knowledge Management**. New York, NY, USA: Association for Computing Machinery, 2017. (CIKM '17), p. 2279–2282. ISBN 9781450349185. Available at: https://doi.org/10.1145/3132847.3133137.

SHAHID, A. *et al.* A survey on prevalent approaches to predict the popularity of social content. *In*: . [*S.l.: s.n.*], 2022. p. 1–7.

SUBRAMANIAN, S.; BALDWIN, T.; COHN, T. Content-based popularity prediction of online petitions using a deep regression model. **CoRR**, abs/1805.06566, 2018. Available at: http://arxiv.org/abs/1805.06566.

TADEJKO, P.; RAKOWSKI, W. Mathematical morphology based ecg feature extraction for the purpose of heartbeat classification. *In*: **6th International Conference on Computer Information Systems and Industrial Management Applications (CISIM'07)**. [*S.l.: s.n.*], 2007. p. 322–327.

TAUD, H.; MAS, J. Multilayer perceptron (mlp). *In*: _____. **Geomatic Approaches for Modeling Land Change Scenarios**. Cham: Springer International Publishing, 2018. p. 451–455. ISBN 978-3-319-60801-3. Available at: https://doi.org/10.1007/978-3-319-60801-3_27.

THELWALL, M. *et al.* Sentiment strength detection in short informal text. **Journal of the American Society for Information Science and Technology**, v. 61, n. 12, p. 2544–2558, 2010. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.21416.

TSUR, O.; RAPPOPORT, A. What's in a hashtag? content based prediction of the spread of ideas in microblogging communities. *In*: **Proceedings of the Fifth ACM International Conference on Web Search and Data Mining**. New York, NY, USA: Association for Computing Machinery, 2012. (WSDM '12), p. 643–652. ISBN 9781450307475. Available at: https://doi.org/10.1145/2124295.2124320.

WANG, K. *et al.* A feature generalization framework for social media popularity prediction. *In*: **Proceedings of the 28th ACM International Conference on Multimedia**. [*S.l.: s.n.*], 2020. p. 4570–4574.

WANG, K. *et al.* A feature generalization framework for social media popularity prediction. *In*: **Proceedings of the 28th ACM International Conference on Multimedia**. New York, NY, USA: Association for Computing Machinery, 2020. (MM '20), p. 4570–4574. ISBN 9781450379885. Available at: https://doi.org/10.1145/3394171.3416294.

WU, B. *et al.* Sequential prediction of social media popularity with deep temporal context networks. **CoRR**, abs/1712.04443, 2017. Available at: http://arxiv.org/abs/1712.04443.

XU, K. *et al.* Multimodal deep learning for social media popularity prediction with attention mechanism. *In*: **Proceedings of the 28th ACM International Conference on Multimedia**. New York, NY, USA: Association for Computing Machinery, 2020. (MM '20), p. 4580–4584. ISBN 9781450379885. Available at: https://doi.org/10.1145/3394171.3416274.

YAISAWAS, P. *et al.* Business popularity analysis from twitter. *In*: MEESAD, P.; SODSEE, S.; UNGER, H. (ed.). **Recent Advances in Information and Communication Technology 2017**. Cham: Springer International Publishing, 2018. p. 337–348. ISBN 978-3-319-60663-7.

ZHANG, X.; LIU, C.-A. Model averaging prediction by k-fold cross-validation. **Journal of Econometrics**, v. 235, n. 1, p. 280–301, 2023. ISSN 0304-4076. Available at: https://www.sciencedirect.com/science/article/pii/S0304407622000975.