



INSTITUTO FEDERAL
São Paulo Pirituba

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE SÃO PAULO

CURSO DE ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

Iago Vitor Araújo

Victor Moreira dos Santos

Vinicius Augusto da Silva

Identificação de manifestações de Clientes de instituições bancárias em redes sociais utilizando inteligência artificial

São Paulo

2021

Iago Vitor Araújo

Victor Moreira dos Santos

Vinicius Augusto da Silva

**Identificação de manifestações de Clientes de instituições bancárias em redes sociais
utilizando inteligência artificial**

Trabalho de Conclusão de curso apresentado ao Instituto Federal de São Paulo, Campus São Paulo Pirituba, como parte dos requisitos para a obtenção do título de Tecnólogo em Análise de Desenvolvimento de Sistemas.

Orientador: Prof. Dr. Fabio Teixeira

São Paulo

2021

Ficha catalográfica preparada pela Coordenadoria de Biblioteca
Instituto Federal de Educação, Ciência e Tecnologia de São Paulo
Câmpus São Paulo Pirituba

Araujo, Iago Vitor de

A663i Identificação de manifestações de clientes de instituições
bancárias em redes sociais utilizando inteligência artificial / Iago
Vitor de Araujo, Victor Moreira dos Santos, Vinicius Augusto da
Silva. – São Paulo, 2021.
23 f. : il.

Orientador: Prof. Dr. Fabio Teixeira

Trabalho de Conclusão de Curso (Tecnólogo em Análise e
Desenvolvimento de Sistemas) – Instituto Federal de Educação,
Ciência e Tecnologia de São Paulo, Câmpus São Paulo Pirituba.

1. Inteligência artificial. 2. Instituições financeiras. 3. Redes
sociais. I. Santos, Victor Moreira dos. II. Silva, Vinicius Augusto da.
III. Teixeira, Fabio. IV. Título.

ATA N.º 23/2021 - ADS-PTB/DAE-PTB/DRG/PTB/IFSP

Ata de Defesa de Trabalho de Conclusão de Curso - Graduação

Tecnologia em Análise e Desenvolvimento de Sistemas - ADS

Aos quatro dias do mês de dezembro de dois mil e vinte e um, às onze horas na sala virtual <https://conferenciaweb.rnp.br/webconf/igor-calebe-zadi>, realizou-se a sessão pública de defesa do Trabalho de Conclusão de Curso intitulado Identificação de manifestações de Clientes de instituições bancárias em redes sociais utilizando inteligência artificial apresentado pelos alunos Iago Vítor Araújo, Victor Moreira dos Santos e Vinicius Augusto da Silva do curso SUPERIOR DE TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS (Campus Pirituba). Os trabalhos foram iniciados às 11:00h pelo professor presidente da banca examinadora, constituída pelos seguintes membros: Fabio Oliveira Teixeira (Presidente/Orientador), Regivaldo Sousa Ferreira (Examinador 1) e Rogério Aparecido Campanari Xavier (Examinador 2).

Itens avaliados	Orientador(a)	Examinador 1	Examinador 2
Solução tecnológica (0 a 4)	4	4	4
Trabalho escrito (0 a 4)	4	3	2
Apresentação oral (0 a 2)	2	2	2
Nota final (NF) (0 a 10)	NF1: 10.0	NF2: 9.0	NF3: 8.0

No item SOLUÇÃO TECNOLÓGICA, foram avaliados itens como: tecnologias inovadoras, nível de dificuldade, impacto da solução, contribuição da solução para a sociedade, coerência da solução com a proposta, arquitetura da solução.

No item TRABALHO ESCRITO, foram avaliados itens como: organização sequencial, argumentação, profundidade do tema, relevância e contribuição acadêmica da pesquisa, correção gramatical, clareza, apresentação estética, adequação aos aspectos formais às normas da ABNT e template do ADS/IFSP-PTB.

No item APRESENTAÇÃO ORAL, foram avaliados itens como: domínio do conteúdo, organização da apresentação, habilidades de comunicação e expressão, capacidade de argumentação, uso dos recursos audiovisuais, correção gramatical e apresentação estética do trabalho.

MÉDIA FINAL: A média final foi calculada pela soma das três notas finais (NF1, NF2 e N3) dividida por três.

A banca examinadora, tendo terminado a apresentação do conteúdo da monografia, passou à arguição dos candidatos. Em seguida, os examinadores reuniram-se para avaliação e deram o parecer final sobre o trabalho apresentado pelos alunos, tendo sido atribuído o seguinte resultado:

Aprovado(a) Reprovado(a) Média final: 9.0

Proclamados os resultados pelo presidente da banca examinadora, foram encerrados os trabalhos e, para constar, eu lavrei a presente ata que assino juntamente com os demais membros da banca examinadora.

Câmpus Pirituba, 6 de dezembro de 2021

Avaliador externo: Sim Não

Documento assinado eletronicamente por:

- Fabio Oliveira Teixeira, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 07/12/2021 16:51:02.
- VINICIUS AUGUSTO DA SILVA, PT3005402 - Discente, em 07/12/2021 17:03:34.
- Victor Moreira dos Santos, PT3006646 - Discente, em 07/12/2021 17:26:50.
- IAGO VITOR DE ARAUJO, PT3006689 - Discente, em 07/12/2021 18:25:15.
- Rogerio Aparecido Campanari Xavier, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 07/12/2021 20:10:45.
- Regivaldo Sousa Ferreira, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 07/12/2021 22:06:40.

Este documento foi emitido pelo SUAP em 07/12/2021. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifsp.edu.br/autenticar-documento/> e forneça os dados abaixo:

Código Verificador: 268388
Código de Autenticação: be6d1cd692



RESUMO

Objetivo: Identificar automaticamente o relato da experiência de Clientes com produtos de instituições bancárias no Twitter por meio da aplicação de técnicas de Inteligência Artificial. **Métodos:** Foram classificados manualmente 4.374 tweets relacionados a produtos oferecidos por bancos e respectivas manifestações negativas de clientes. Estes tweets foram utilizados para o treinamento de um classificador baseado em cálculos probabilísticos. Um grafo com os resultados da classificação e conexões oriundas da rede social foi construído e apresentado por meio de um sítio eletrônico. **Resultados:** A classificação final da base de textos pelo classificador atingiu 70% de precisão, 67% para revocação e 69% no F-measure. **Conclusão:** As técnicas de Inteligência Artificial relacionadas ao reconhecimento de entidades em redes sociais apresentadas neste artigo são promissoras e alcançaram níveis de precisão e revocação satisfatórios, no ponto de vista que as postagens passam despercebidas e não agregam valor aos serviços prestados pelas instituições bancárias .

Palavras-chave: finanças; entidades; twitter; inteligência artificial; análise de dados; aprendizado de máquina;

ABSTRACT

Objective: Automatically identify the report of Customer experience with products from banking institutions on Twitter through the application of Artificial Intelligence techniques. **Methods:** There were manually classified 4,374 tweets related to products offered by banks and respective customer statements. These tweets were used to train a classifier based on probabilistic calculations. A graph with the classification results and connections from the social network was constructed and presented through an electronic website. **Results:** The final classification of the text base through the classifier reached 0.70 precision, 0.67 for recall and 0.69 for F-measure. **Conclusion:** The Artificial Intelligence techniques related to the recognition of entities in social networks presented in this article are promising and have reached satisfactory levels of accuracy and recall.

Keywords: finance; entities; twitter; artificial intelligence; data analysis; machine learning;

LISTA DE ABREVIATURAS E SIGLAS

IBM (International Business Machines)

PLN (Processamento de Linguagem Natural)

NER (Named Entity Recognition)

IA (Inteligência Artificial)

API (Application Programming Interface)

CRF (Conditional Random Fields)

DB (Database)

MB (Megabyte)

GB (Gigabyte)

RAM (Random Access Memory)

CSV (Comma-separated Values)

JSON (JavaScript Object Notation)

AVG (Average)

LISTA DE PUBLICAÇÕES

Artigo publicado em congresso

SANTOS, V. M.; ARAÚJO, I. V.; SILVA, V. A.; TEIXEIRA, F. **Identificação de manifestações de clientes de instituições bancárias em redes sociais utilizando inteligência artificial.** In: 12º CONGRESSO DE INOVAÇÃO, CIÊNCIA E TECNOLOGIA DO IFSP - 2021. 17 out. 2021.

Resumo: Identificar automaticamente o relato da experiência de Clientes com produtos de instituições bancárias no Twitter por meio da aplicação de técnicas de Inteligência Artificial. Foram identificados e classificados manualmente produtos oferecidos por bancos e respectivas manifestações de clientes em 4.374 tweets. Estes tweets foram utilizados para o treinamento de um classificador baseado em cálculos probabilísticos. E um grafo com os resultados da classificação e conexões oriundas da rede social, foi construído e apresentado por meio de um sítio eletrônico. A classificação final da base de textos por intermédio do classificador atingiu 70% de precisão, 67% para revocação e 69% para F-measure. As técnicas de Inteligência Artificial relacionadas ao reconhecimento de entidades em redes sociais apresentadas neste artigo são promissoras e alcançaram níveis de precisão e revocação satisfatórios.

SUMÁRIO

1. INTRODUÇÃO	11
1.1. APRESENTAÇÃO DO PROBLEMA	13
1.2. OBJETIVOS	13
2. REVISÃO DA LITERATURA	13
3. JUSTIFICATIVA	15
4. MATERIAIS E MÉTODOS	15
5. RESULTADOS E DISCUSSÃO	19
6. CONCLUSÃO	22
7. REFERÊNCIAS	23

1. INTRODUÇÃO

O uso das redes sociais, em especial o Twitter, vem aumentando exponencialmente com o passar do tempo. O Brasil, em 2019, ocupava a segunda posição no ranking de países cuja população passa mais tempo nas redes sociais (VIESBA et al., 2020).

Com isso, seguindo a visão do cliente, grandes quantidades de informações são publicadas diariamente. Dessa forma, as mídias sociais têm um papel de extrema importância na comunicação e troca de informações dos mais variados tipos, uma vez que em poucos segundos uma mensagem pode atingir uma quantidade considerável de pessoas (JUNQUEIRA, 2014). Dentre estas postagens estão inseridos tweets de avaliação e opinião das experiências que os usuários tiveram com determinados produtos ou serviços oferecidos por uma empresa. A repercussão de uma mensagem informal publicada na Internet pode ser mais notória que uma reclamação feita diretamente na central de atendimento ao consumidor das empresas, já que diversas outras pessoas podem ter compartilhado da mesma experiência e isso resulta no compartilhamento da mensagem por diversos outros indivíduos, alcançando proporções escaláveis.

Pela perspectiva empresarial, a análise do feedback de seus clientes no Twitter serve como parâmetro de avaliação de seus serviços ou produtos, mostrando tanto os pontos positivos quanto os negativos que devem ser melhorados ou corrigidos. É possível analisar os dados e identificar como seus concorrentes de mercado estão sendo avaliados e como tirar uma vantagem competitiva da situação (JUNQUEIRA, 2014).

A análise de dados não estruturados se torna cada vez mais importante devido ao aumento de acessos às redes sociais, uma vez que as opiniões são expressas através de textos, imagens ou até mesmo vídeos (OLIVEIRA, 2012). O desafio proposto neste projeto consiste na utilização do reconhecimento de entidades, por meio do aprendizado de máquina, para filtrar informações retiradas de *tweets* a respeito dos serviços oferecidos por organizações bancárias, desenvolvendo uma aplicação com o auxílio da inteligência artificial que seja capaz de identificar serviços, avaliações e o banco relacionado, além de relacionar as

informações postadas pelos usuários e criar mecanismos de visualização para potencializar as análises, por exemplo, utilizando técnicas de grafos.

Segundo a IBM, *Machine Learning* é uma tecnologia onde os computadores têm a capacidade de aprender de acordo com as respostas esperadas por meio associações de diferentes dados, os quais podem ser imagens, números e tudo que essa tecnologia possa identificar (MUELLER; MASSARON, 2021).

Ou seja, o Aprendizado de Máquina é uma área da Inteligência Artificial (IA) voltada para a criação de sistemas capazes de absorver conhecimento de forma automática (MONARD, 2003). O sistema de aprendizagem da IA ocorre por intermédio da experiência adquirida de parâmetros já utilizados por outros sistemas bem sucedidos na resolução de problemas. Os parâmetros podem ser definidos de forma particular de acordo com o objetivo, linguagem, metodologia, paradigma e premissa.

Outra ferramenta importante de IA utilizada no projeto é o Processamento de Linguagem Natural (PLN). O objetivo da PLN é proporcionar aos computadores a habilidade de compreender textos, sons e diversos aspectos da comunicação humana (ABRANTES, 2017).

Através da *Named Entity Recognition* (NER), sub-área do Processamento de Linguagem Natural, é possível analisar e classificar mensagens em formato de texto, a fim de catalogar essas entidades em categorias pré-definidas, por exemplo: nomes, organizações, problemas, adjetivos, locais, expressões, etc (GONZALEZ; STRUBE DE LIMA, 2003).

O grande desafio das organizações é filtrar de um grande volume de dados as informações que estão presentes nessas mídias (ABRANTES, 2017). Quando examinados, podem atingir de forma considerável na criação de valor e vantagem competitiva para as empresas, sendo capaz de gerar novas maneiras de comunicação com seus clientes ou no desenvolvimento de novos produtos, serviços e estratégias, aumentando a sua lucratividade.

1.1. APRESENTAÇÃO DO PROBLEMA

Atualmente um grande volume de dados é gerado diariamente por meio da utilização de redes sociais (JUNQUEIRA, 2014). Elas são utilizadas para realizar críticas positivas ou negativas em relação aos serviços ou produtos de instituições bancárias, que, por sua vez, não conseguem lidar com a quantidade de informações gerada por tais plataformas (ABRANTES, 2017). Por meio do aprendizado de máquina e técnicas de processamento de linguagem natural aplicadas às postagens realizadas pelos usuários, é possível criar uma solução capaz de organizar os dados das redes sociais e proporcionar às instituições bancárias uma vantagem competitiva com as informações coletadas.

1.2. OBJETIVOS

Objetivo Geral: Identificar automaticamente o relato da experiência de Clientes com produtos de instituições bancárias no Twitter por meio da aplicação de técnicas de Inteligência Artificial.

- Objetivo específico 1: Criar um mecanismo automático de coleta e armazenamento de dados não estruturados
- Objetivo específico 2: Construir um algoritmo para identificação de entidades nomeadas a partir de técnicas de Inteligência Artificial
- Objetivo específico 3: Criar um painel de visualização de dados com filtragens avançadas
- Objetivo específico 4: Realizar testes contínuos para aperfeiçoamento da IA.

2. REVISÃO DA LITERATURA

Na literatura existe uma variedade de artigos que estudam o reconhecimento de entidades nomeadas. O Twitter destaca-se como fonte de dados de alguns destes estudos, pois oferece mecanismos automatizados para coleta de informações presentes na plataforma. A utilização de técnicas relacionadas à Inteligência Artificial é um ponto em comum em muitos destes trabalhos.

No artigo (PIVETA, 2013), o autor faz uso de uma Application Programming Interface (API), pública de streaming do Twitter para a obtenção de tweets em tempo real. O objetivo deste trabalho é unir duas técnicas para agrupar tweets de acordo com uma determinada classificação. A primeira técnica é o uso de mecanismo de *Story Link Detection*, esse método baseia-se na utilização de semelhança de cosseno entre vetores de palavras do documento com peso TF-IDF. Já a segunda técnica é o reconhecimento de entidades, onde foi aplicado um código aberto chamado Stanford NER (FINKEL; GRENAGER; MANNING, 2005).

A análise de sentimento foi o ponto abordado pela pesquisa de Abrantes (ABRANTES, 2017), onde sua meta foi analisar e classificar o sentimento, tanto positivo quanto negativo, de Tweets relacionadas a quatro grandes empresas do cenário da tecnologia. Os tweets foram classificados em três grupos: Positivo, Negativo e Neutro, que eram baseados em palavras contidas em um dicionário. As avaliações com nota superior a zero foram classificadas como “Positivo”, notas inferiores a zero como “Negativo” e por fim notas, iguais a zero como “Neutro”. Foi utilizada a Linguagem R e o seu conjunto de bibliotecas de mineração de textos, como pré-processamento e análise de frequência de termos. Ao final do projeto, durante seis dias, foram coletados por volta de um total de 120.000 tweets. Uma coleta diária tinha um média de extração de 5.000 tweets por marca.

O estudo de Oliveira (OLIVEIRA, 2012), elenca os principais desafios a serem enfrentados no reconhecimento de entidades no Twitter. São seis pontos citados: o grande volume de dados, falta de formalismo, dependência do idioma, ambiente dinâmico, falta de contextualização das mensagens e a orientação ao fluxo de dados. Neste trabalho, o autor optou por usar uma abordagem denominada Filter Stream named Entity Recognition, a FS-NER (OLIVEIRA, 2012). Que se configura no emprego de filtros para solucionar a tarefa de reconhecimento de entidades, nesse projeto foram utilizados cinco filtros e três coleções de dados do twitter. A combinação de filtros melhora o desempenho e aumenta a efetividade do reconhecimento. O autor observou que a abordagem FS-NER foi capaz de obter uma melhora de aproximadamente 3% quando comparada a abordagens fundamentadas em Conditional Random Fields (CRF).

3. JUSTIFICATIVA

Durante a última década o uso das redes sociais aumentou exponencialmente. Com isso, uma infinidade de dados e informações circulam diariamente pela rede. Segundo o estudo de Sonia Vermelho e Ana Paula Velho (VERMELHO et al., 2014), a crença de que as redes sociais são ferramentas eficientes de comunicação está em uma crescente diária. Uma vez que esses instrumentos são formados pela poderosa circulação de informação.

Um grande desafio para as organizações é realizar a coleta e a curadoria desses dados e transformá-los em vantagem competitiva para alavancar seus negócios. A partir dessas informações coletadas é possível saber os pontos fracos e positivos dos serviços ou produtos tanto de sua própria empresa quanto de seus concorrentes de mercado.

Baseando-se nesse cenário e visando solucionar os problemas de coleta, filtragem e identificação de dados em redes sociais, esse trabalho foi desenvolvido. A partir de técnicas de reconhecimento de entidades será possível mapear o serviço ou produto utilizado, o problema/opinião dos usuários e a organização bancária envolvida.

4. MATERIAIS E MÉTODOS

A base de dados utilizada neste trabalho foi construída a partir de sentenças coletadas na rede social Twitter por meio do *streaming* orquestrado pelo uso de uma Application Programming Interface (API) pública da própria rede social. A linguagem Python¹ foi utilizada para construir o código da aplicação de coleta e armazenamento das mensagens em um banco de dados não-relacional de documentos na nuvem da empresa MongoDB². Um filtro foi aplicado à coleta de mensagens, restringindo os tweets coletados a um grupo de instituições mencionadas no corpo da mensagem. O código da aplicação de coleta e filtros utilizados pode ser acessado no GitHub³. A aplicação de *streaming* utilizou uma

¹ <https://www.python.org/>

² <https://www.mongodb.com/>

³ <https://github.com/financetcc/dumpTweets>

máquina virtual criada no ambiente Azure da Microsoft⁴. As configurações da máquina criada foram 500 Mb de memória RAM, uma única CPU e 4GB de armazenamento.

A etapa de pré-processamento das postagens removeu os *retweets* e exportou os dados para o padrão CSV, para posterior utilização na tarefa de classificação manual das postagens. O código desta etapa pode ser acessado no Google Colab⁵.

O software Doccano⁶ foi a ferramenta escolhida para a classificação manual das postagens. Dois tipos de entidades foram definidas: Produto e Problema. A primeira refere-se aos produtos ou serviços oferecidos pelas instituições bancárias. Já a segunda refere-se aos problemas que os usuários enfrentam ao utilizar um produto ou serviço da instituição bancária. Foram analisadas aproximadamente vinte mil postagens, dentre elas, 3.439 sentenças apresentaram as entidades produto e problema presentes no corpo da postagem e foram marcadas manualmente pelos autores do projeto. As postagens marcadas com as entidades produto e problema foram exportadas segundo o formato JSON⁷, disponível na ferramenta Doccano. O arquivo pode ser acessado no diretório GitHub⁸.

As postagens classificadas manualmente foram transformadas do formato JSON para CONLL2003 (SANG; DE MEULDER, 2003). O código responsável por esta transformação pode ser acessado no Google Colab⁹. O formato CONLL2003 é amplamente utilizado no desenvolvimento de algoritmos de reconhecimento de entidades.

O algoritmo utilizado para o treinamento do reconhecimento de entidades foi o Conditional Random Fields (CRF) (LAFFERTY; MCCALLUM; PEREIRA, 2001). A partir da leitura do arquivo no formato CONLL2003, o algoritmo CRF é capaz de classificar as entidades previamente determinadas, por meio de abordagens estatísticas que levam em consideração o sequenciamento das palavras nas postagens. A implementação do modelo pode ser acessada no Google Colab¹⁰.

⁴ <http://portal.azure.com/>

⁵ <https://bit.ly/2U2AhgA>

⁶ <https://doccano.herokuapp.com/>

⁷ <https://www.json.org/>

⁸ <https://bit.ly/3rYigMT>

⁹ <https://bit.ly/37mf8ke>

¹⁰ <https://bit.ly/37mf8ke>

As medidas de precisão, revocação e F-score foram aplicadas aos resultados do treinamento e teste alcançados pelo classificador construído neste projeto. Precisão corresponde a quanto os dados do projeto são relevantes, Revocação (recall) é quão completo os dados estão e o F-score ou F-measure corresponde a média harmônica entre as métricas de desempenho.

A etapa de validação do classificador utilizou aproximadamente 60 mil tweets. Nesta etapa, o algoritmo construído foi capaz de identificar 4374 postagens com a presença das entidades Produto e Problema simultaneamente. O código construído para a validação pode ser acessado no Google Colab¹¹.

Uma aplicação web foi desenvolvida para mostrar a visualização dos tweets classificados na etapa de validação. O framework-web Flask¹² e a biblioteca vis.js¹³ foram utilizados para a construção de visualizações dinâmicas e interativas dos dados. A navegação no grafo foi construída a partir de métodos da biblioteca NetworkX¹⁴, que contém funções de redes complexas. O código da aplicação web pode ser acessado no GitHub¹⁵.

A visualização dos dados foi apresentada por meio de um grafo formado por nós representados pelas entidades identificadas automaticamente pelo classificador construído neste projeto, Produto e Problema, e pelos IDs dos Clientes, que realizaram as postagens na plataforma Twitter. As arestas do grafo conectam as entidades e Clientes e facilitam a compreensão da rede e estrutura dos dados, bem como as relações estabelecidas entre produto, problema e cliente.

A aplicação, primeiramente, foi publicada no GitHub e posteriormente hospedada na plataforma de nuvem denominada Heroku¹⁶. A Tabela 1 mostra as tecnologias utilizadas para a construção da aplicação web.

Tabela 1. Lista de tecnologias utilizadas

¹¹ <https://bit.ly/3isrbD4>

¹² <https://flask.palletsprojects.com/en/2.0.x/>

¹³ <https://visjs.org/>

¹⁴ <https://networkx.org/>

¹⁵ <https://github.com/financetcc/ner-visualization-of>

¹⁶ <https://www.heroku.com>

Tecnologia	Camada/Subsistema	Justificativa
Git	Infraestrutura	Versionamento de código distribuído entre todos os desenvolvedores. Necessário pelas ferramentas de hospedagem escolhidas
Virtual Machine - Azure	Infraestrutura	Hospedagem e execução do código responsável por realizar o stream de tweets.
MongoDB	Servidor	SGBD não-relacional de licença gratuita.
Python	Desenvolvimento	Linguagem de programação utilizada no projeto para coleta e classificação das informações.
HTML e CSS	Desenvolvimento	Linguagem de marcação/estilização utilizadas para desenvolvimento da página web.
JavaScript	Desenvolvimento	Linguagem de programação utilizada para estruturar a visualização e comportamento do grafo.
Flask	Servidor	Framework utilizado para criação de rotas e templates back-end.
Vis.js e NetworkX	Servidor	Framework utilizada para visualização dinâmica dos dados em formato de grafo.

5. RESULTADOS E DISCUSSÃO

A Tabela 2 mostra o desempenho do classificador construído neste projeto. O treinamento e teste foi medido por meio da precisão, revocação e f1-score aplicado a uma base com 3.439 tweets classificados manualmente. Os problemas relatados pelos Clientes (usuários) na rede social identificados na tabela como B-PROBLEMA e I-PROBLEMA, apresentam uma porcentagem menor. Este resultado pode estar associado à composição do texto relacionado a esta classe que, na maioria dos casos, é formado por duas ou mais palavras, por exemplo: “não está funcionando” e “sem acesso”. Esta condição dificulta a tarefa do classificador em identificá-los. Os produtos e serviços identificados pelo classificador receberam a nomenclatura B-PRODUTO E I-PRODUTO e alcançaram valores de desempenho superiores às outras classes.

Pela classificação manual dos tweets, foi alcançado um total de 733 problemas, onde a expressão B-PROBLEMA representa o total de primeiras palavras dos problemas e 675 produtos, bem como descrito para os problemas o B-PRODUTO significa o total de primeiras palavras dos produtos.

Tabela 2. Desempenho do classificador

Label	Precision	Recall	F-score	Support
B-PROBLEMA	0.447	0.179	0.255	733
I-PROBLEMA	0.452	0.237	0.311	1707
B-PRODUTO	0.709	0.677	0.692	675
I-PRODUTO	0.565	0.309	0.399	324
micro avg	0.544	0.318	0.401	3439
macro avg	0.543	0.350	0.415	3439
weighted avg	0.512	0.318	0.382	3439

Fonte: Autores.

A aplicação web¹⁷, construída neste projeto, mostra uma visualização utilizando conceitos de grafo para ilustrar a ligação dos Clientes (usuários), produtos e problemas. A partir deste grafo é possível chegar a algumas conclusões, por exemplo: um único problema pode estar relacionado a mais de um produto; um usuário pode ter mais de um produto registrado, tendo em vista que uma pessoa pode ter mais de uma reclamação a fazer; é possível obter o usuário que mais relatou problemas, bem como identificar o produto que mais recebeu reclamações.

A Figura 2 mostra a visualização do grafo após a construção. Os nós são unificados por arestas que correlacionam os problemas, produtos e usuários. Os nós de cor **vermelha** representam diversos produtos ofertados pelos bancos. Já os nós de cor **amarela** identificam os problemas relatados pelos usuários, cujo *username* está representado pela cor **azul** do grafo.

Figura 2. Visualização Geral do Grafo



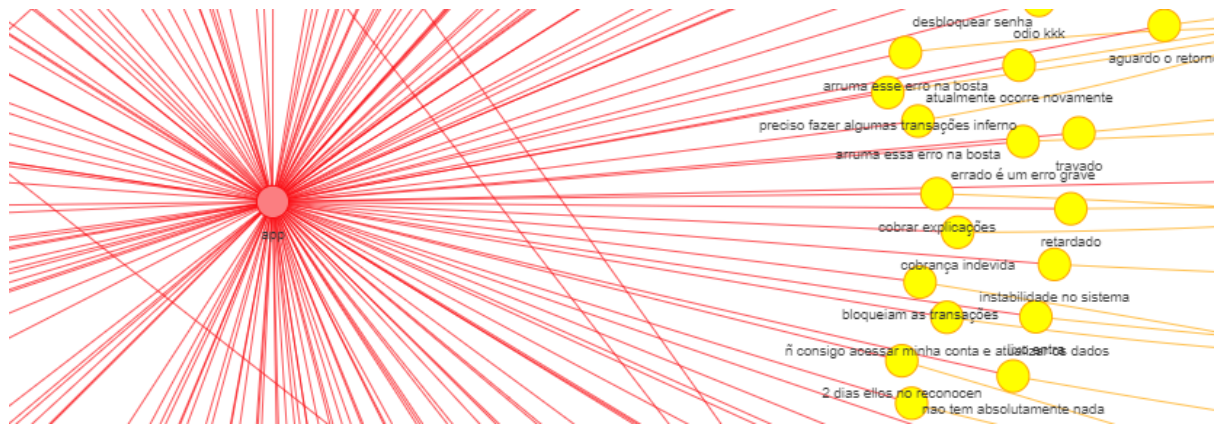
Fonte: Autores.

A Figura 3 mostra as correlações criadas entre produto, problema e usuário conforme foram definidas na aplicação do projeto. Cada produto possui uma

¹⁷ <https://ner-visualization.herokuapp.com>

correlação direta com os problemas mencionados pelos usuários facilitando a interpretação do grafo e das informações.

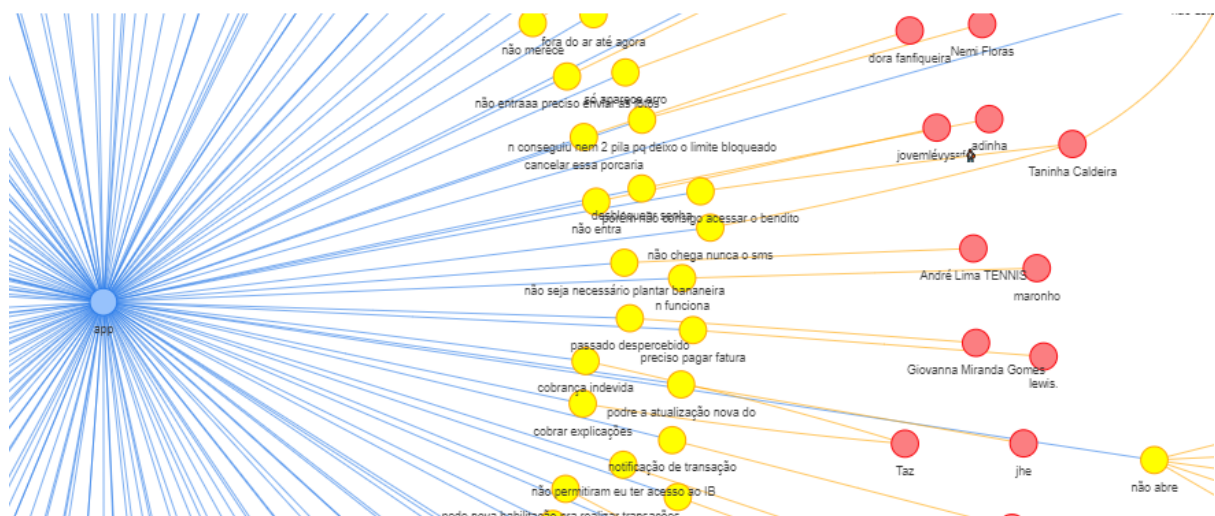
Figura 3. Visualização da relação Produto-Problema



Fonte: Autores.

A Figura 4 mostra a relação dos grupos de entidades a partir de um filtro que possibilita a visualização de produtos, problemas e usuários relacionados somente a um produto. Neste caso, por exemplo, é possível analisar todos os problemas mencionados somente no produto “app” oferecido pelos bancos. Para acessar este recurso do grafo, basta clicar 2 vezes no ponto central do nó de um serviço ofertado e o grafo será reestruturado.

Figura 4. Visualização das relações de um único Produto do Grafo



Fonte: Autores.

Outras técnicas de reconhecimento de entidades podem ser exploradas e aplicadas aos dados deste projeto, como Bi-LSTM (HUANG; XU; YU, 2015), ou ainda ampliar a base de postagens classificadas manualmente para posterior utilização no treinamento do algoritmo e melhora nos índices de desempenho.

6. CONCLUSÃO

Através das pesquisas realizadas neste projeto, foi concluído que a aplicação de mecanismos e ferramentas tecnológicas nas redes sociais, sobretudo, a aplicação de conceitos de análise e ciência de dados, podem contribuir e agregar valor para as empresas interessadas em oferecer melhores serviços para o usuário final. Embora muitas postagens não tenham relevância alguma, a solução apresentada no projeto é capaz de filtrar os dados importantes e descartar outras postagens que não se fazem necessárias.

Para que uma maior eficiência seja atingida, ainda é necessário que o banco de dados com *tweets* seja expandido, para que o algoritmo possa interpretar melhor e adapte-se às diferentes postagens. O resultado até o momento é razoável e já serve de grande ajuda e material de pesquisa para aperfeiçoamentos futuros.

Por fim, os resultados do projeto foram satisfatórios e foi possível adquirir uma boa base para pesquisas futuras que envolvam Inteligência Artificial e *Machine Learning*. Houve também o aprendizado de diversas outras ferramentas e linguagens como: Python, Collab, Zotero, etc. para que este artigo fosse desenvolvido e há a esperança de que o mesmo possa ser usado como base para futuras pesquisas que tenham como foco os temas citados anteriormente.

7. REFERÊNCIAS

- ABRANTES, F. A. C. **Processos e ferramentas de análise de Big Data : a análise de sentimento no twitter**. masterThesis—[s.l.] Instituto Superior de Economia e Gestão, out. 2017.
- FINKEL, J. R.; GRENAGER, T.; MANNING, C. **Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling**. Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05). **Anais...** In: ACL 2005. Ann Arbor, Michigan: Association for Computational Linguistics, jun. 2005. Disponível em: <<https://aclanthology.org/P05-1045>>. Acesso em: 4 ago. 2021
- GONZALEZ, M.; STRUBE DE LIMA, V. **Recuperação de Informação e Processamento da Linguagem Natural**. 2003.
- HUANG, Z.; XU, W.; YU, K. **Bidirectional LSTM-CRF Models for Sequence Tagging**. **arXiv:1508.01991 [cs]**, 9 ago. 2015.
- JUNQUEIRA. **A Utilização das Redes Sociais para o Fortalecimento das Organizações.**, 2014. Disponível em: <<https://www.aedb.br/seget/arquivos/artigos14/22020181.pdf>>
- LAFFERTY, J. D.; MCCALLUM, A.; PEREIRA, F. C. N. **Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data**. Proceedings of the Eighteenth International Conference on Machine Learning. **Anais...**: ICML '01. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 28 jun. 2001. . Acesso em: 16 abr. 2021
- MONARD, M. **Conceitos sobre aprendizagem de máquina.**, 2003. Disponível em: <<https://dcm.ffclrp.usp.br/~augusto/publications/2003-sistemas-inteligentes-cap4.pdf>>
- MUELLER, J. P.; MASSARON, L. **Machine Learning For Dummies**. [s.l.] John Wiley & Sons, 2021.
- OLIVEIRA, D. M. DE. **Uma abordagem baseada em fluxo de filtros para o reconhecimento de entidades em mensagens do twitter**. 26 out. 2012.
- PIVETA, M. **TCC. Detecção de Ligação entre Histórias com Reconhecimento de Entidades no Twitter.**, 2013. Disponível em: <<http://www.uel.br/cce/dc/wp-content/uploads/PlanoTCC-MateusPiveta.pdf>>
- SANG, E. F. T. K.; DE MEULDER, F. **Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition**. **arXiv:cs/0306050**, 12 jun. 2003.
- VERMELHO, S. C. et al. **Refletindo sobre as redes sociais digitais**. **Educação & Sociedade**, v. 35, n. 126, p. 179–196, mar. 2014.
- VIESBA, L. M. et al. **Anais do I Encontro Nacional Movimentos Docentes Volume - II**. [s.l.] V&V Editora, 2020.