

Avaliação do desempenho comportamental de um algoritmo de mineração de texto em conexões com internet nos municípios de Conceição do Araguaia e Redenção - PA
Evaluation of the behavioral performance of a text mining algorithm on internet connections in the municipalities of Conceição do Araguaia and Redenção - PA
Evaluación del desempeño conductual de un algoritmo de minería de texto en conexiones a internet en los municipios de Conceição do Araguaia y Redenção – PA

Recebido: 03/01/2023 | Revisado: 09/01/2023 | Aceitado: 10/01/2023 | Publicado: 12/01/2023

Wilker José Caminha dos Santos

ORCID: <https://orcid.org/0000-0002-5265-583X>
Universidade do Estado do Pará, Brasil
E-mail: wilkercaminha@uepa.br

Renato Ferreira Carr

ORCID: <https://orcid.org/0000-0002-2719-607X>
Universidade do Estado do Pará, Brasil
E-mail: renatocarr@uepa.br

Rogério Santiago Lopes

ORCID: <https://orcid.org/0000-0002-1954-6842>
Faculdade Integrada Carajás, Brasil
E-mail: rogeriosantiago08@gmail.com

Daveny Jales da Silva Junior

ORCID: <https://orcid.org/0000-0002-5157-1503>
Faculdade Integrada Carajás, Brasil
E-mail: davenyjuniior@gmail.com

William Gabriel Barbosa dos Santos

ORCID: <https://orcid.org/0000-0002-3502-9896>
Faculdade Integrada Carajás, Brasil
E-mail: williamgbs@hotmail.com

Edgard Augusto Nascimento Ribeiro

ORCID: <https://orcid.org/0000-0002-7325-1199>
Faculdade Integrada Carajás, Brasil
E-mail: edgard.augusto.n.ribeiro@gmail.com

Armando José de Sá Santos

ORCID: <https://orcid.org/0000-0001-8242-6571>
Universidade do Estado do Pará, Brasil
E-mail: armando.santos@uepa.br

Resumo

Este estudo teve como objetivo geral avaliar o desempenho comportamental de um método de mineração de texto escrito em *Python* com atributos voltados para redes bayesianas em conexões de baixa velocidade nas cidades de Conceição do Araguaia e Redenção-PA localizadas no interior do estado do Pará. Dentre os autores pesquisados para a constituição conceitual deste trabalho, destaca-se Caminha *et al* (2017), Conte *et al* (2016), Santana *et al* (2016), Almeida *et al* (2016) e Mata *et al* (2013). A metodologia utilizada se encontra (Caminha *et al*, 2017), que por sua vez, se enquadra como qualitativa-analítica, em virtude de buscar trabalhar no aprofundamento da avaliação de desempenho comportamental com base de textos minerados da rede social *twitter*. As conclusões mais relevantes deste trabalho foram decorrentes da comparação dos resultados obtidos na cidade de Conceição do Araguaia e Redenção, onde foi possível observar inconformidades no desempenho do tráfego de dados de ambas as cidades, tanto para aspectos positivos quanto para negativos apesar da igualdade na velocidade do link em ambas as cidades nos anos de 2016, 2017, 2021 e 2022, a fim de comparar a existência de uma evolução no seu comportamento.

Palavras-chave: Mineração de texto; Avaliação de desempenho; Estado do Pará.

Abstract

The general objective of this study was to evaluate the behavioral performance of a text mining method written in Python with attributes aimed at Bayesian networks in low-speed connections in the cities of Conceição do Araguaia and Redenção-PA located in the interior of the state of Pará. Among the authors researched for the conceptual constitution of this work, Caminha *et al* (2017), Conte *et al* (2016), Santana *et al* (2016), Almeida *et al* (2016) and Mata *et al* (2013) stand out. The methodology used is found (Caminha *et al*, 2017), which, in turn, is classified as

qualitative-analytical, as it seeks to work on deepening the evaluation of behavioral performance based on texts mined from the social network twitter. The most relevant conclusions of this work resulted from the comparison of the results obtained in the city of Conceição do Araguaia and Redenção, where it was possible to observe nonconformities in the performance of data traffic in both cities, both for positive and negative aspects despite the equality in speed. of the link in both cities in the years 2016, 2017, 2021 and 2022, in order to compare the existence of an evolution in its behavior.

Keywords: Text mining; Performance evaluation; State of Pará.

Resumen

El objetivo general de este estudio fue evaluar el desempeño conductual de un método de minería de texto escrito en Python con atributos dirigidos a redes bayesianas en conexiones de baja velocidad en las ciudades de Conceição do Araguaia y Redenção-PA ubicadas en el interior del estado de Paraca. Entre los autores investigados para la constitución conceptual de este trabajo, se destacan Caminha et al (2017), Conte et al (2016), Santana et al (2016), Almeida et al (2016) y Mata et al (2013). Se encuentra la metodología utilizada (Caminha et al, 2017), que, a su vez, se clasifica como cualitativa-analítica, ya que busca trabajar en la profundización de la evaluación del desempeño conductual a partir de textos extraídos de la red social twitter. Las conclusiones más relevantes de este trabajo surgieron de la comparación de los resultados obtenidos en la ciudad de Conceição do Araguaia y Redenção, donde fue posible observar no conformidades en el desempeño del tráfico de datos en ambas ciudades, tanto para los aspectos positivos como negativos a pesar de la igualdad en velocidad del enlace en ambas ciudades en los años 2016, 2017, 2021 y 2022, con el fin de comparar la existencia de una evolución en su comportamiento.

Palabras clave: Minería de texto; Evaluación del desempeño; Estado de Pará.

1. Introdução

De fato, há tempos, observa-se que o brasileiro tem gastado boa parte de seu tempo conectado à internet, seja pelo uso de dispositivos como o computador, *tablet* ou *smartphone*, sendo este último citado o mais utilizado no momento (CGI, 2016).

O objetivo geral é avaliar o desempenho comportamental de um método de mineração de texto escrito em *Python* com atributos voltados para redes bayesianas em conexões de baixa velocidade nas cidades de Conceição do Araguaia e Redenção-PA localizadas no interior do estado do Pará. O método de mineração texto utilizado nas conexões localizadas no interior do estado do Pará é proveniente do *Naive Bayes*, capaz de executar as tarefas de extração e manipulação de textos pré-definidos em uma base de dados acentuada. Durante os procedimentos de extração utilizamos um termo para eventuais testes que envolveu atividades de mineração, sendo escolhido o termo “*rio2016*”. O resultado da análise após a mineração foi visualizado em forma gráfica para melhor compreensão dos resultados decorrentes da sua manipulação.

Diante a este fato, essa pesquisa justifica-se que, usuários conectados, as redes sociais digitais estão gerando volume de dados de grande valia sendo utilizadas para mineração e processamento de texto, suscitando informações que possam garantir resultados eficazes e compreensíveis do conteúdo contribuindo para tomada de decisão importante em pesquisas científicas ou em outras atividades com fins específicos, como aponta Gusmão (2014).

Ainda convém salientar que para que seja possível a mineração de tal conteúdo contido nas redes sociais, faz necessário a utilização de técnicas de mineração de texto, umas das formas mais utilizadas são as fermentas de mineração de dados, texto e conteúdo. Normalmente são algoritmos que exigem elevadas bandas de conexão para que possam ser utilizadas de forma produtiva, segundo Machado (2010). Entretanto, grande parte das cidades brasileiras não dispõem de serviços de internet de alta velocidade. O que ocasionou na hipótese de avaliar um método de mineração de texto em condições extremas de conexão.

A metodologia utilizada se encontra em Caminha *et al*, (2017), que por sua vez, se enquadra como qualitativa-analítica, em virtude de buscar trabalhar no aprofundamento da avaliação de desempenho comportamental com base de textos minerados da rede social *twitter*, tendo como coleta de dados relacionados a pesquisa o levantamento bibliográfico e análise de dos dados coleados para possíveis conclusões. (Lakatos; Marconi, 2003; Gil, 2010).

2. Metodologia

De fato, a base de dados do *twitter* possui uma *API (Application Programming Interface)* que possibilita a interação entre desenvolvedores e o conteúdo contido em sua base de dados em tempo real, possibilitando a criação de aplicações, onde a mesma, por sua vez, possa utilizar das informações contidas e disponíveis dentre os diversos usuários portadores de uma conta ativa na rede social. A requisição pode ser configurada e executada em tempos determinados para experimentos de extração (Santana *et al* 2016).

Todo o esquema de desenvolvimento, configuração e adaptação do algoritmo para iniciar o processo de extração consiste na modificação e inserção de permissões para acesso a base de dados do *Twitter*. Os conceitos, técnicas e metodologias de desenvolvimento e configuração estão contidas em Caminha *et al.* (2017). Portanto não serão abordados detalhes referentes ao desenvolvimento do método, apenas na sua pertinente avaliação em ambientes que oferecem condições extremas referentes a velocidade de conexão com a internet.

A presente crise política do Brasil, uma economia abatida e a resultante contração nos recursos públicos tornarão muito difícil para o governo alcançar seus objetivos de elevar a velocidade média da banda larga do Brasil para 25 Mb/s até 2018, o que exerce fundamento relevante para a realização de pesquisa para mensurar o desempenho de aplicações que utilizam a internet para seu funcionamento, objetivo e propósito (Rafael B. C. Henrique L. B. *et al.* 2018)

O método avaliativo teve base na utilização das seguintes etapas para alcançar os objetivos propostos, sendo primeiramente, a escolha do método de mineração de texto a qual foi realizado todas as atividades que envolva: avaliar e diagnosticar. A opção do termo a ser minerado, que definiu o termo “rio2016, para que dessa forma pudesse gerar conteúdo significativo para a mineração. Ainda convém advertir que o método é decorrente de aprendizado de máquina supervisionado, adjunto de um tutor supervisor para atividades definidas. (Conte *et al*, 2016).

A segunda etapa se deu no estabelecimento do tempo de execução do método, que por sua vez, foi definido que ele estaria ativo em um período de três horas de execução contínua em tempo real. Os atributos que contribuíram para a decisão e definição do tempo se deram na perspectiva a qual a mineração pudesse ser supervisionada pelo autor.

Conceição, Santos e Nascimento *et al*, 2021 aponta que, dados minerados, utilizando a internet pode ocasionar variações que precisam ser acompanhadas de perto, a supervisão das tarefas realizadas na pesquisa é de fato uma característica comum na tecnologia da informação, dessa forma, é possível contribuir para resultados íntegros.

A terceira teve o intuito de efetivar a velocidade dos *links* para a avaliação do método, as velocidades disponíveis no ecossistema se deram em um, dois e dezesseis megabytes de velocidade, sendo assim, os efetivos *links* para a realização da eventual mineração. Por fim foi executada a etapa de quantificação dos dados, sendo escolhidos o texto do *tweet* como parâmetro principal para a construção de gráficos para melhor visualização dos resultados.

Ainda convém lembrar que a metodologia de avaliação foi criada pelo próprio autor com base em pesquisas anteriores, que podem ser conferidas em Caminha *et al*, (2017). Após uma série de levantamentos afim de que pudesse ser executada de maneira limpa e ao mesmo tempo concreta obteve-se inicialmente a justificativa de escolha do método, que se deu em: Escolher um método que fosse constituído de princípios de usabilidade, com o propósito de facilitar sua manipulação adaptação, conexão com a base de dados, inserção do termo a ser minerados e dentre outros. Dessa forma o tempo de execução da pesquisa seria investido em termos mais relevantes tal como os objetivos propostos citados anteriormente. (Almeida V 2014)

A escolha do termo a ser minerado foi propositalmente definida a fim de gerar amplas quantidades de textos, já que os eventos realizados não apenas no Brasil, mas em outros países é de fato alvo de uma gama de opiniões diversificadas que geram grades quantidades de dados que podem ser utilizados de forma dedicada em diversas áreas do conhecimento, isso inclui utilizar desse meio para mensurar o desempenho de algoritmos de mineração, com base no tempo de execução e velocidade de

host. Entretanto o ponto chave de extrema relevância para o desenvolvimento e conclusão da avaliação é a velocidade a qual o algoritmo trafega, e a sua total disponibilidade, o que contribuiu para escolha dos *links* utilizados na avaliação.

A pesquisa foi realizada em cidades do interior do estado, o que dificultou a disponibilização de diversificadas opções de uso, dessa forma foram levantadas quais velocidades estariam disponíveis para a utilização em ambas as cidades, e que ao mesmo tempo não possuíssem elevada monetização, estando dentro dos padrões cabíveis e aceitáveis para o autor. Foram estabelecidas três distintas alternativas com diferente singularidade, enfatizando que a escolha dos *links* com base na disponibilidade da mesma e valores monetários, descartando a possibilidade de ações aleatórias, para que não pudesse influenciar nos resultados propostos na avaliação do algoritmo. Após o levantamento dos requisitos e esquematização da metodologia de avaliação deu-se início ao processo de obtenção dos resultados adaptado com base em Cezar, G (2017) e Castanheira (2008).

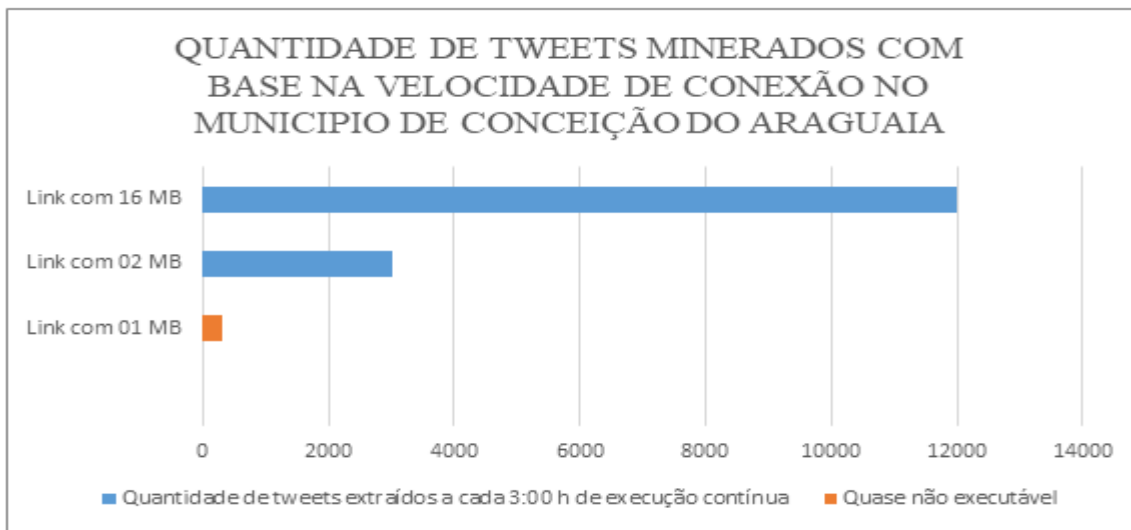
A execução do algoritmo de mineração concedeu no período de 05 de agosto até 21 de agosto de 2016 e 01 de novembro a 10 novembro de 2017 realizando a coleta de *tweets* relacionados ao evento dos jogos olímpicos do Rio, utilizando o termo “rio2016” sendo observado que a quantidade de parâmetros coletados durante o período. O mesmo processo foi realizado no ano de 2021 A execução do algoritmo de mineração concedeu no período de 05 de julho até 23 de outubro de 2021 e 05 de janeiro a 10 de abril de 2022 realizando a coleta de *tweets* relacionados ao Covid 19, utilizando o termo “covid” sendo observado que a quantidade de parâmetros coletados durante esse período.

A principal fonte de dados, a qual foram extraídos utilizando o método foi a rede social *Twitter*, onde ela abriga centenas e milhares de *Post*, inseridos a cada dia. Conceição, Santos e Nascimento et al, 2021, afirma, que ambientes computacionais interativos possibilitam produção de informações diárias, sendo considerável satisfatória para utilização de pesquisas de opinião e mensuração de desempenho de aplicação computacionais na rede. Redes sociais abrigam possibilidades diversificadas, principiante no que se refere a análise de dados e descoberta de conhecimento com base e mineração de dados e textos. Foram constatados que durante a coleta, *tweets* com *hashtags*, *links* e *trending topics* relacionados ao evento que foram quantificados de forma individual para as cidades de Conceição do Araguaia e Redenção guardados em um arquivo *JSON*.

3. Resultados

A seguir são apresentados os resultados do método nas Figuras 1 e 2, que dispõe da quantidade de dados extraídos com o termo “rio2016” no ano de 2016 e 2017, em um período de 3 horas de execução contínua do método juntamente com a comparação do desempenho em ambas as cidades conforme as Figuras 3 e 4. Já a Figura 5 apresenta a quantidade de dados retornados de forma global em ambas as cidades. As Figuras 6, 7 e 8 apresentam o mesmo procedimento de mineração, entretanto o processo sucedeu entre os anos de 2021 e 2022.

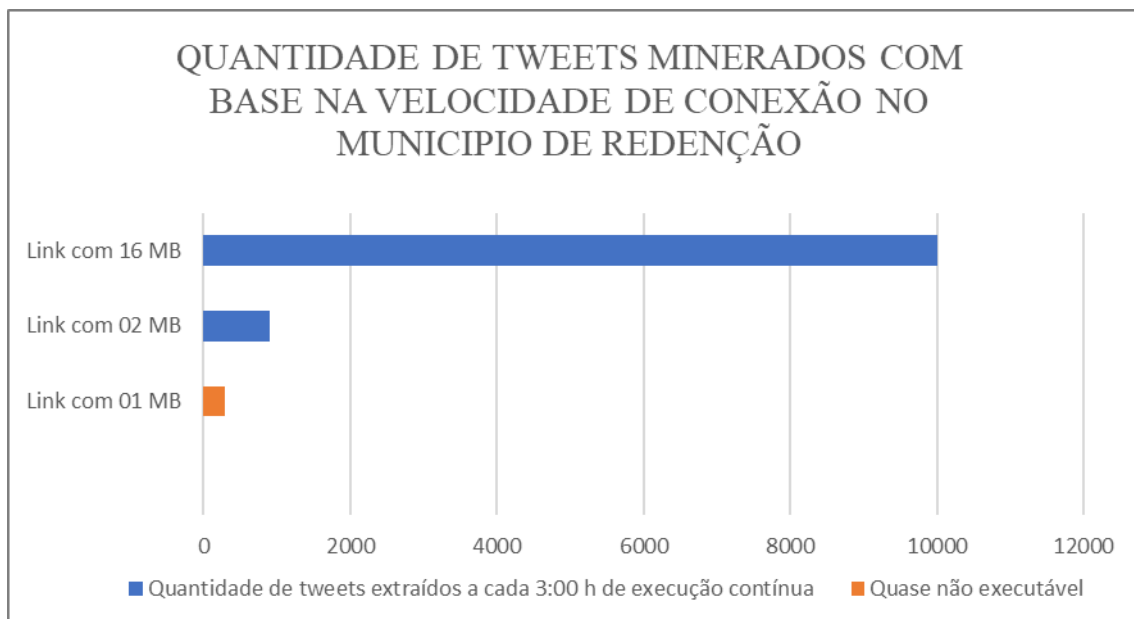
Figura 1 – Quantidade de *tweets* minerados por tempo de execução e velocidade do *link* no município de Conceição do Araguaia.



Fonte: Autores (2017).

Como observado na Figura 1, quanto maior a velocidade do *link* utilizado melhor será o retorno da análise dos dados minerados. Os testes com *links* de conexão acima de 2MB possuem suficiência se levarmos em consideração tempo de execução. Testes realizados com 2MB de conexão mostram que o algoritmo consegue minerar 3000 *tweets* em um período de 3 horas de execução contínua, já com 16MB a capacidade aumenta para 12000 *tweets*. Algumas conexões de fato não são relevantes para a execução do algoritmo, por não está dentro dos requisitos mínimos para um cenário em tempo real como é o caso de uma simples conexão de 1MB onde a quantidade de dados minerados em um longo período de tempo não se torna relevante como espaço amostral. Conforme a Figura 2, os dados apresentaram desempenho similar aos dados apresentados na figura 01, com exceção das quantidades de *tweets* minerados. Apesar dos testes terem sido realizados em conexões com a mesma capacidade de tráfego os valores retornados no município de Redenção foram inferiores, se comparados com os dados da cidade de conceição do Araguaia.

Figura 2 - Quantidade de *tweets* minerados por tempo de execução e velocidade do *link* no município de Redenção.



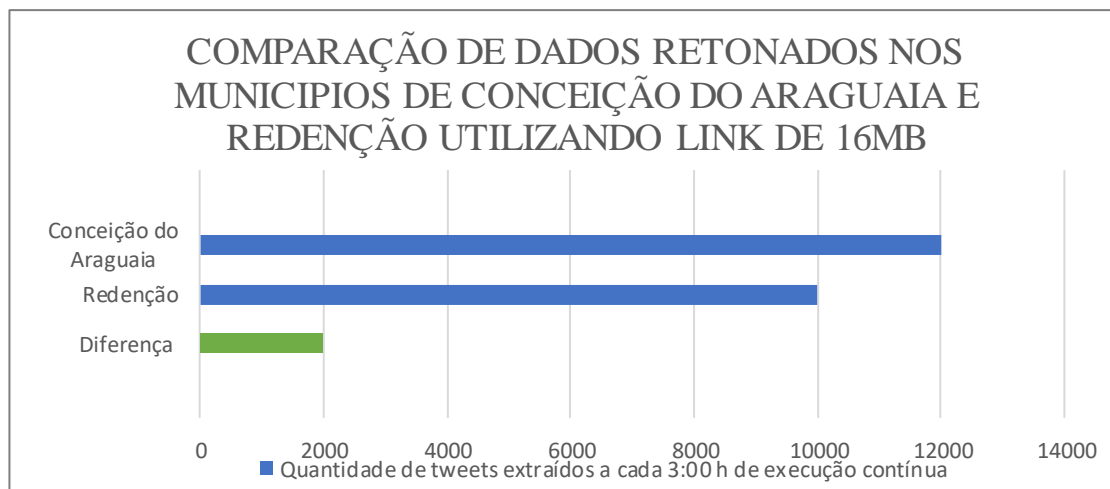
Fonte: Autores (2017).

Como observado acima a quantidade de *tweets* minerados por tempo de execução e velocidade do *link* no município de Redenção, inicialmente com o *link* com velocidade de 02MB de conexão, que por sua vez, conseguiu extrair no tempo proposto aproximadamente 900 *tweets*, enquanto, o desempenho consegue alcançar avanços significativos quando a velocidade do *link* é estabelecida em 16MB, contabilizado aproximadamente 10000 *tweets*. Assim como na Figura 1 os resultados alcançados a velocidade de 1MB não atingiu valores significativos para ser considerado com espaço amostral durante a pesquisa.

Salientando que o enredamento da informação solicitada para a extração é de fato ator de extrema influência nos resultados já que é levado em consideração a quantidade de dados que forem extraídos e o nível de complexidade do caractere solicitado. Ou seja, o tamanho de caractere que forma uma palavra e/ou simbologia a ser minerada interfere no volume de *tweets* e no tempo de processamento.

Por fim podemos observar que existe de fato uma divergência com a real velocidade de conexão disponível para sua utilização, apesar dos testes realizados em parâmetros de igual estatura com relação ao *link* de internet o algoritmo sofreu oscilações no fluxo de dados e valores finais retornados, contribuindo para essa hipótese conforme a Figura 3 e 4.

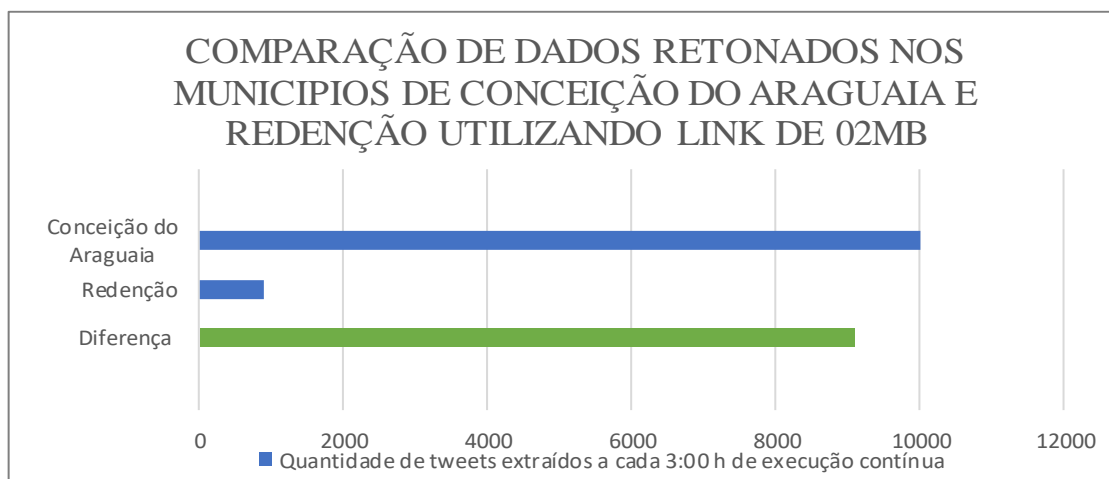
Figura 3 – Comparação de Dados retornados nos municípios de Conceição do Araguaia e Redenção utilizando link de 16MB.



Fonte: Autores (2017).

Conforme observado na Figura 3 os dados retornados na cidade de Conceição do Araguaia foram superiores aos dados retornados na cidade de Redenção, apesar dos testes realizados em velocidades de igual valor, no município de Conceição do Araguaia o algoritmo conseguiu minerar aproximadamente 12000 *tweets*, enquanto o município de Redenção obteve 10000 *tweets* objetivando uma diferença de aproximadamente 2000 *tweets*, utilizando o *link* de 16MB. O Algoritmo também apresentou oscilações com o *link* de 02MB no município de Redenção, retornando notórios 900 *tweets*, enquanto com a utilização em Conceição do Araguaia retornou valores equivalente a 10000 *tweets*, acentuando uma diferença de 9100 *tweets* conforme a Figura 4.

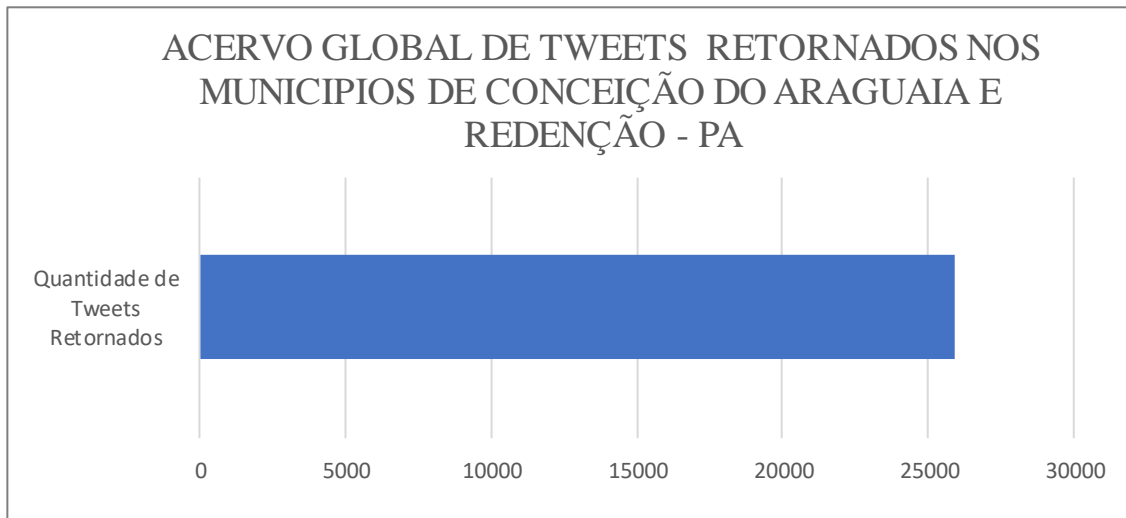
Figura 4 – Comparação de dados retornados nos municípios de Conceição do Araguaia e Redenção utilizando *link* de 02MB.



Fonte: Autores (2017).

Em uma escala global obteve-se a quantidades de textos minerados em ambas as cidades, de forma condensada conforme a Figura 5 expõe.

Figura 5 – Acervo global de *tweets* retornados nos municípios de Conceição do Araguaia e Redenção – PA.



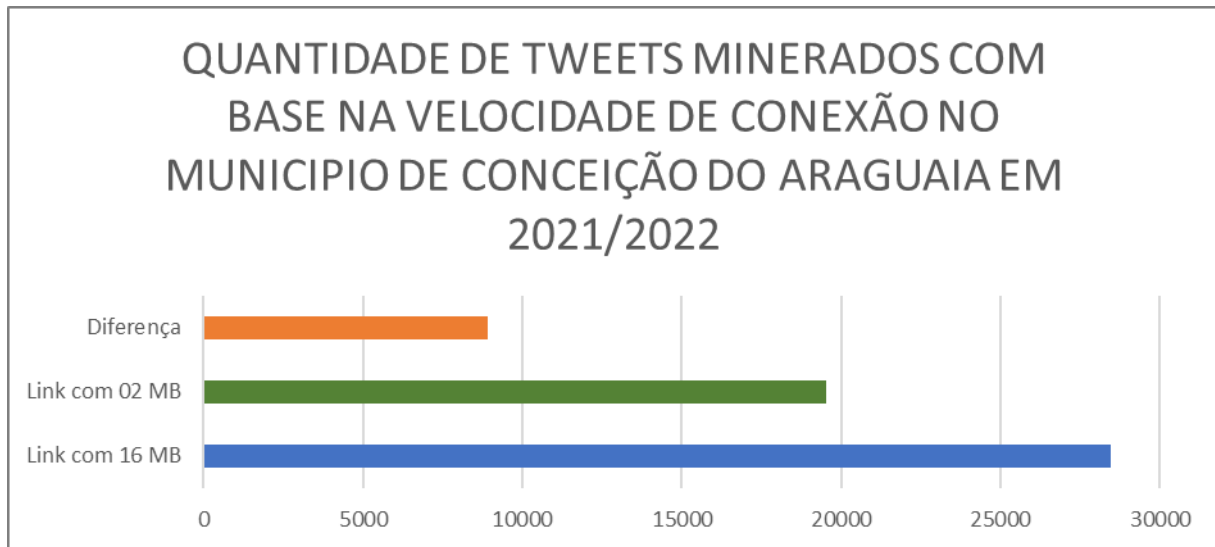
Fonte: Autores (2017).

Conforme a Figura 5 apresenta, a quantidade de *tweets* contendo o termo rio2016 de forma global, ou seja, a junção do conteúdo minerado na cidade de Conceição do Araguaia e Redenção-PA de forma condensada. A quantidade de texto minerados deu-se em aproximadamente 25900 *tweets*, um número considerável se levado em consideração o tempo a qual ele levou para retornar os valores nas duas cidades.

Os resultados coletados no ano de 2022, utilizando o termo covid como base para extração de textos minerados no *twitter*, também consegui alcançar resultados satisfatórios, se levados em consideração o período entre os dois procedimentos e o avanço da tecnologia que permite a conexão com internet nas cidades de Conceição do Araguaia e Redenção.

A Figura 6 apresenta, a quantidade de *tweets* contendo o termo covid de minerado na cidade de Conceição do Araguaia de forma condensada. A quantidade de texto minerados com conexão de 02 MB sucedeu em aproximadamente 18000 *Tweets*, enquanto a conexão de 16 MB retornou 25000 *Tweets* um número considerável se levado em consideração o tempo a qual ele levou para retornar os valores e ao avanço das tecnologias de conexão com a internet. A diferença entre os retornos foi de aproximadamente 7000 *Tweets*.

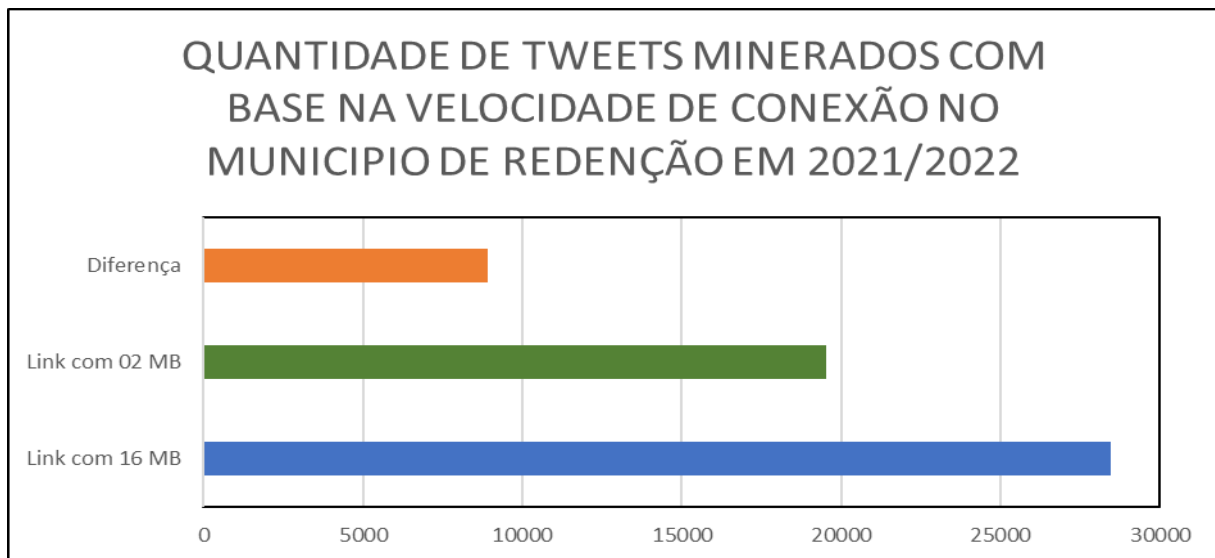
Figura 6 – Quantidade de *Tweets* Minerados utilizando Conexões com internet no Município de Conceição do Araguaia– PA em 2021/2022.



Fonte: Autores (2022).

Ainda convém lembrar que a Figura 7 apresenta, a quantidade de *tweets* contendo o termo covid de minerado na cidade de Redenção. A quantidade de texto minerados com conexão de 02 MB sucedeu em aproximadamente 19550 *Tweets*, enquanto a conexão de 16 MB retornou 28450 *Tweets* com uma diferença entre os retornos de aproximadamente 8900 *Tweets*.

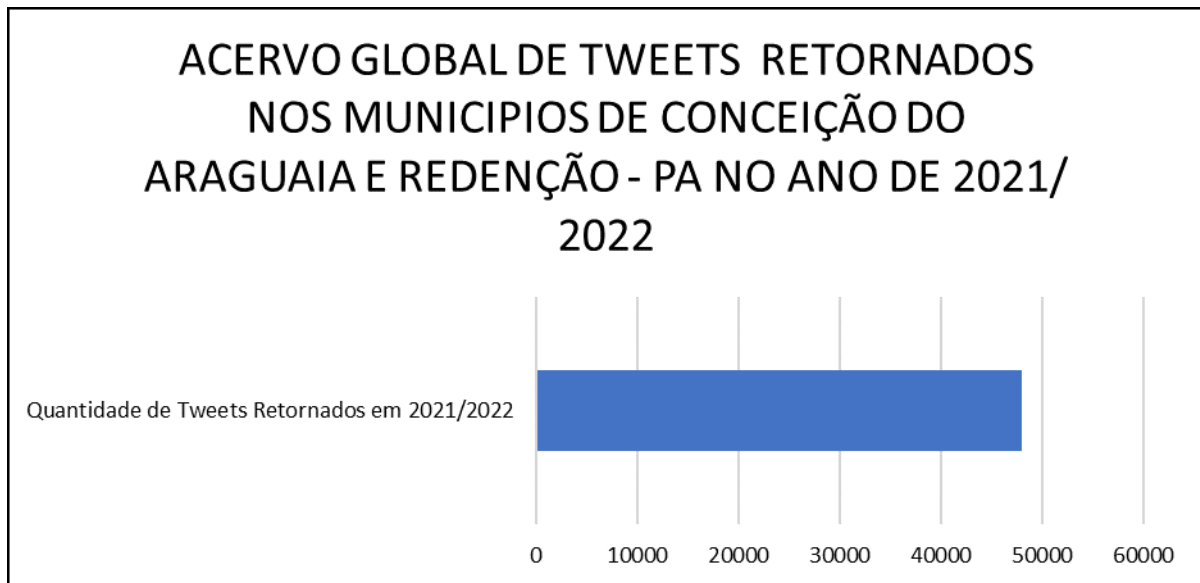
Figura 7 – Quantidade de *Tweets* Minerados utilizando Conexões com internet no município de Redenção – PA em 2021/2022.



Fonte: Autores (2022).

Em uma escala global obteve-se a quantidades de textos minerados em ambas as cidades, de forma condensada conforme a Figura 8 expõe.

Figura 8 – Acervo global de *tweets* retornados nos municípios de Conceição do Araguaia e Redenção – PA em 2021/2022.



Fonte: Autores (2022).

Conforme a Figura 8 apresenta, a quantidade de *tweets* contendo o termo covid de forma global, do conteúdo minerado na cidade de Conceição do Araguaia e Redenção-PA. A quantidade de texto minerados deu-se em aproximadamente 48000 *tweets*, um número considerável no processo de mineração.

De modo geral o algoritmo de mineração de texto se comportou de forma convincente e válida, ainda que seja em velocidades de conexão extremas se comparadas aos grandes centros urbanos o algoritmo testado consegue ser funcional e eficaz, realizando as atividades a qual ele o programado, o que não desqualifica a metodologia de utilização em ecossistemas desse porte. Entretanto algo que deve ser levado em consideração é o tempo escolhido para deixá-lo em atividade pois ele é um atributo que possivelmente influencia no valor dos dados retornados, dessa forma a melhor opção de fato para trabalhar com algoritmos que dispõe de grandes taxas de processamentos de informações em tempo real na *web* é a capacidade de tráfego do *host* utilizado.

4. Conclusão

Durante a pesquisa foram criadas e utilizadas medidas tais como levantamento bibliográfico para a criação de uma metodologia de avaliação que contribuiu para a tomada de decisões que consequentemente aferiu os resultados esperados durante a análise. Dessa forma foi possível concluir os objetivos propostos na pesquisa de forma satisfatória, O objetivo geral visou avaliar o desempenho comportamental de um método de mineração de texto escrito em *Python* com atributos voltados para redes bayesianas em conexões de baixa velocidade nas cidades de Conceição do Araguaia e Redenção-PA localizadas no interior do estado do Pará. uma vez que a avaliação do método de mineração de texto em conexões de internet no interior do estado do Pará foi concluída com sucesso.

Algo que deve ser mencionado de forma disjunta e que foi essencial dentro do seu respectivo papel na elaboração da pesquisa foi a escolha do termo a ser minerado. O termo rio2016 consequentemente foi responsável por causar a disseminação do conteúdo no algoritmo de mineração, proporcionado em uma escala global aproximadamente 25900 *tweets* englobando o conteúdo minerado nas cidades de Conceição do Araguaia e Redenção-PA.

Durante a pesquisa obteve-se notáveis resultado referentes ao comportamento do método em distintas velocidades de conexão, possibilitando a avaliação com base na quantidade de dados retornados em um período de três horas de execução

continua em tempo real. Através da comparação dos resultados obtidos na cidade de Conceição do Araguaia e Redenção foi possível observar uma inconformidade no desempenho do tráfego de dados de ambas as cidades, tanto para aspectos positivos quanto para negativos apesar da igualdade na velocidade do link em ambas as cidades.

Ainda convém advertir que apesar de condições extremas de conexão se comparados aos grandes polos urbanos, o método conseguiu realizar as atividades a qual foi designado, minerando textos decorrentes da escolha do termo rio2016 de maneira satisfatória. No município de Conceição do Araguaia utilizando 16MB de velocidade de tráfego, a maior banda de velocidade disponível, o método conseguiu aferir aproximadamente 12000 *tweets*, enquanto que no município e redenção a capacidade foi reduzida para 10000 *tweets*, ambas com a mesma capacidade de tráfego de dados, o que ocasionou na hipótese de inconformidade no retorno do conteúdo no município de Redenção, pois a diferença entre a quantidades de textos retornados se comparadas ao município de Conceição do Araguaia torna-se pertinente para a classificação do método.

Foi perceptível a mesma variância no retorno dos textos utilizando o *link* de 02MB onde o Município de Conceição do Araguaia conseguiu contabilizar através do método aproximadamente 10000 *tweets*, enquanto o município de Redenção obteve 900 *tweets*, uma diferença de aproximadamente 9100 *tweets*.

Os textos minerados entre os anos de 2021 e 2022 contendo o termo covid de minerado na cidade de Redenção com conexão de 02 MB sucedeu em aproximadamente 19550 *Tweets*, enquanto a conexão de 16 MB retornou 28450 *Tweets* com uma diferença entre os retornos de aproximadamente 8900 *Tweets*. De forma global, do conteúdo minerado na cidade de Conceição do Araguaia e Redenção-PA deu-se em aproximadamente 48000 *tweets*. Contudo, sabemos que a constante evolução na forma como nos conectamos com a internet favorece de maneira satisfatória a maneira como nos conectamos com os dados que estão disponíveis nas diversas bases de dados que conhecemos.

De fato, é notório a percepção de alguns acontecimentos durante todo o esquema de planejamento esquematização e formulação da pesquisa, que por sua vez, acarretaram pequenas dificuldades que contribuíram para que houvesse mais atenção dentro dos padrões estabelecidos na metodologia, tal como, as conexões disponíveis há qual os testes fossem realizados, o que colaborou para a seleção de apenas três conexões distintas.

Portanto, com base nos dados gerados ao decorrer da pesquisa pode-se observar de forma especial detalhes consideráveis sobre o método. Ainda que o algoritmo utilizado para a mineração de texto seja aproveitado em *links* com a mesma proporção de velocidade e tempo pré-estabelecidos, os valores retornados podem sofrer oscilações, quedas, baixas e inconsistências involuntariamente, o que não compromete de forma ampla sua qualidade quanto software, o que qualifica o algoritmo como elemento adaptável a condições desfavoráveis de desempenho. Ainda convém lembrar que houve uma mudança significativa em relação a forma a qual o algoritmo de mineração trabalha. Com a evolução na forma como utilizamos as conexões com a internet possibilitou seu melhor aproveitamento após um espaço de tempo deste os testes iniciais de extração de tweets entre os anos de 2016 e 2022.

Referências

- Alexander, P., & Naroubek, P. (2010) Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In Proceedings of the Seventh conference on International Language Resources and Evaluation
- Almeida, G. R. T., et al. (2017). *Improving Social CRM through electronic word-of-mouth: a case study of ReclameAqui*. In: XXIII Simpósio Brasileiro de Sistemas Multimídia e Web, Porto Alegre-RS: SBC.
- Almeida, V. (2014). A governança global da internet e o papel do Brasil. Estadão.com. <http://www.estadao.com.br/noticias/impreso,a-governanca-global-da-internet-e-o-papel-do-brasil,1151269>
- Arnaldo, D. (2017). *O Brasil e o Marco Civil da Internet - O Estado da Governança Digital Brasileira*: Instituto Igarapé a think and do tank.
- Caminha, W. J. S., et al. (2017) Método de Mineração de Texto com Foco no Português do Brasil com Redes Bayesianas. São José do Rio Preto- SP: DINCON 2017- Conferencia Brasileira de Dinâmica Controle e Aplicações.

- Castanheira, L. (2008) Aplicação de Técnicas de Mineração de Dados em Problemas de Classificação de Padrões. Departamento de Engenharia Elétrica - UFMG"
- César, G. (2017) Análise de dados: como ela ajuda na tomada de decisão em sua empresa <http://www.buscarid.com/artigos/analise-de-dados-como-ela-ajuda-na-tomada-de-decisao-na-sua-empresa>
- Conte, T. N. M., et al. (2016). *Classificação de Gêneros Musicais Utilizando Aprendizado de Máquina*. Belém: Universidade Federal do Pará. In Programa de pós-graduação em Engenharia Elétrica. PPGEE.
- Conceição, R. M., et al. (2021). Método de aprendizagem supervisionada para a identificação de rastros de cyberbullying Lisboa: Proceedings of the IADIS Ibero American Conference Applied Computing.
- CGI. (2016). Pesquisa sobre o uso das tecnologias da informação e comunicação no Brasil: TIC domicílios e empresas http://www.cetic.br/media/docs/publicacoes/2/TIC_DOM_EMP_2013_livro_eletronico.pdf
- Domingos, P. (2012) A few useful things to know about machine learning. Commun.in ACM, New York: *Computer Science and Computer Engineering*.
- Fernanda, E. (2006) redes neurais e sua aplicação em sistemas de recuperação de informação, São Paulo, Universidade de São Paulo USP. <http://www.scielo.br/pdf/ci/v35n1/v35n1a03.pdf>.
- Gil, A. C. (2010) *Como elaborar projetos de pesquisa*. São Paulo: Atlas, 184.
- Gusmão, G. (2017). Dados de redes sociais podem ser alternativa a estudos <http://exame.abril.com.br/tecnologia/dados-de-redes-podem-ser-alternativa-a-estudos-diz-pesquisa>
- Lakatos, E. M., & Marconi, M. de A. (2003). *Fundamentos da Metodologia Científica*. São Paulo: Atlas.
- Machado, A. P., et al. (2021). Mineração de texto em redes sociais aplicada à educação a distância. *Revista Digital da CVA*, 1-21.
- Mata, A. E. C., et al. (2013). Improvements in Communication Process in Real Projects using Social Network. In: *WORLDCOMP'13 - World Congress in Computer Science, Computer Engineering, and Applied Computing*. Las Vegas: Computer Science and Computer Engineering.
- Rafael B. C., Henrique L. B., et al. (2018). Mineração da web para estimação e monitoramento de índices de preços de imóveis de Florianópolis. *Anais do Congresso Internacional de Conhecimento e Inovação*.
- Santana, A. C., et al. (2016). Proposta de Arquitetura de Software de API de Extração, Armazenamento, Pré-processamento e Fusão de Dados advindos de Redes Sociais. Belém: *II Conferência de Social CRM*.
- Ronnie C. T., Nunes, Márcio C., Alexandre L. G., et al (2022). Utilização de bases de conhecimento de domínio público: um estudo de caso sobre a linguagem RDF.Ciki: *Anais do Congresso Internacional de Conhecimento e Inovação*.