

Data Discover

Sociedade da Informação: O Papel do Motor de Busca na Pesquisa e aprendizado Acadêmico

Henrique Oliveira NEVES (Aluno)

Faculdade de Tecnologia de Santana do Parnaíba – CEETEPS

henrique.neves6@fatec.sp.gov.br

Vinicius Naziozeno Santoro do RIO (Aluno)

Faculdade de Tecnologia de Santana do Parnaíba – CEETEPS

vinicius.rio@fatec.sp.gov.br

Rosineide Aparecida De Lira VOLGARIN (Orientadora)

Faculdade de Tecnologia de Santana do Parnaíba – CEETEPS

rosineide.volgarin@fatec.sp.gov.br

RESUMO

O resumo tem por princípio, dar a compreender de forma sucinta e clara, tudo que é abordado no estudo, deve se atentar para a essencialidade de se retratar aquilo que o texto aborda; nenhum elemento a mais deve ser incorporado em seu conteúdo e sua totalidade gira em torno de 120 palavras, o resumo não possui parágrafo ou espaçamentos, sendo uma construção textual contínua e linear.

Palavras-chave:

ABSTRACT

O mesmo texto do resumo em outro idioma. O idioma deve atender as diretrizes do veículo de divulgação.

Keywords:

INTRODUÇÃO

Na classificação dos períodos históricos da sociedade, o Mundo Antigo ou Antiguidade é o intervalo de tempo entre a criação da escrita, por volta de 4.000 a 3.500 a.C, e o declínio do Império Romano do Ocidente, em 476 d.C. Até então, a aprendizagem estava intimamente ligada a transmissão oral do conhecimento. Civilizações como a egípcia, mesopotâmica, grega e romana valorizavam a educação, embora esta fosse restrita a uma elite privilegiada. As escolas, vinculadas a templos ou academias, visavam formar cidadãos instruídos em áreas como retórica, filosofia, matemática e ciências naturais (Ribeiro, 2019).

Contudo, a partir das necessidades humanas de armazenar informações de forma longa, e superar as limitações da difusão do conhecimento de forma oral, a escrita passou a ocupar um posto de destaque em relação ao compartilhamento de conhecimento e acesso à informação. Nas épocas da humanidade seguintes a Idade Antiga, a partir de novas necessidades humanas no que se refere o alcance da informação, outras tecnologias surgiram para esse fim.

Na idade contemporânea, mais especificamente no século XX, duas novas tecnologias ocasionaram mudanças importantes no que diz respeito não somente ao acesso à informação, como também a disseminação do conhecimento. No caso a criação da internet e a popularização dos computadores pessoais. Anteriormente, a obtenção de informações era limitada por barreiras geográficas, com acesso restrito a bibliotecas e instituições de ensino (Cunha, 2009). No entanto, a internet democratizou o acesso ao conhecimento, tornando possível para quase qualquer pessoa, independentemente do lugar que esteja, conectar-se a uma vasta quantidade de informações por meio de uma conexão à rede. Os computadores pessoais proporcionaram às pessoas as ferramentas necessárias para explorar esse conhecimento, o que lhes permitiu buscar, acessar e compartilhar informações de forma facilitada.

Todavia, a facilidade de acesso à informação proporcionada por essas tecnologias, também fez com que surgissem desafios relacionados a qualidade e confiabilidade das informações disponíveis nos meios digitais. Em um cenário em que a quantidade de conteúdo disponível só aumenta, torna-se cada vez mais difícil discernir entre informações verídicas e úteis daquelas que são imprecisas, tendenciosas ou até mesmo falsas. Esse problema é especialmente preocupante para estudantes de ciência de dados, em que a precisão e a credibilidade das fontes de informação, são essenciais para o desenvolvimento do conhecimento, estudos e pesquisas. (Calazans, 2008).

Excluído: rápida, eficiente e acessível.

Excluído:

Excluído: e

Excluído: no contexto acadêmico

Excluído:

Excluído: e a produção de trabalhos acadêmicos de qualidade ...

Dessa forma, é de extrema importância que existam ferramentas que possibilitem o acesso a bases de dados, revistas e artigos científicos. Instrumentos desse tipo desempenham um papel fundamental na promoção da aprendizagem e capacitação dos estudantes em relação a prática da pesquisa e desenvolvimento do pensamento crítico.

Excluído: ,

Diante do exposto, este trabalho propõe relatar o desenvolvimento de um Motor de Busca direcionado à estudantes de Ciência de dados, que funciona por meio da indexação de artigos acadêmicos e base de dados para estudos científicos. A partir de uma entrada fornecida pelo usuário, o mecanismo de pesquisa tem como proposta retornar informações como autores e ano de publicação dos artigos. Por meio de indicadores, será possível que o estudante explore informações analíticas dos trabalhos retornados. Além da funcionalidade de indexação de artigos, a ferramenta disponibiliza a inserção de bases de dados para serem avaliadas, dessa forma, é possível que o discente selecione a mais adequada para a utilização em seu projeto estudantil. Complementarmente, a partir da base de dados e informações adicionais fornecidas pelo usuário, a ferramenta indica quais algoritmos de aprendizado de máquina são mais adequados para seu estudo.

Excluído: a

Objetivando uma solução que propicie o acesso a informações de qualidade e fomento o aprendizado prático por meio da pesquisa e experimentação, o que se busca responder é: **Como um motor de busca pode auxiliar na eficácia da pesquisa e aprendizado em Ciência de Dados?**

Excluído: frente

Excluído: investigação científica

Excluído: acadêmico

REFERENCIAL TEÓRICO

A informação é um conceito fundamental no campo da ciência da informação e desempenha um papel essencial em diversos contextos acadêmicos e práticos. Em sua essência, a informação pode ser definida como um conjunto de dados organizados e estruturados de tal forma que tenham significado e relevância para um determinado contexto ou objetivo. (Oleto, 2006). Isso implica que a informação não é simplesmente um conjunto de dados brutos, mas sim uma representação selecionada e processada de dados que pode ser utilizada para compreender, tomar decisões, resolver problemas e comunicar conhecimento.

Excluído:

Segundo Bukland (1991 como citado em Oleto, 2006), a informação pode ser categorizada em três tipos, informação como processo, informação como conhecimento e informação como coisa. No que diz respeito em sua aplicação como processo, a informação é utilizada para transmitir conhecimento. O ato de relatar, seja em meios físicos ou digitais

caracteriza o processo de informar algo. A informação como conhecimento está relacionada a exposição do ser humano a algo que, até então era desconhecido, e a internalização do que lhe foi informado. Dessa forma, é algo abstrato, intangível e individual à cada pessoa. A informação como coisa está ligada aos instrumentos que, de alguma forma, são tidos como informativos.

Excluído: indivíduo

No meio de todas as modificações sucedidas nos meios tecnológicos que ocorreram após a Segunda Guerra Mundial, destaca-se a proliferação das tecnologias da informação. Estas ocasionaram uma mudança significativa em vários aspectos da sociedade, como relações políticas, sociais e econômicas. Além disso, houve uma mudança intransigente nos sistemas de informação habituais (Tomaél et al., 2001).

Dado o contexto da época, informação era invariavelmente associada a meios físicos e impressos, como livros. Contudo, com o advento da internet, as fontes de referência, começaram a ser relacionadas as mídias digitais. Adicionalmente, a proliferação de computadores em universidades, instituições de ensino e empresas, corroborou ainda mais para a digitalização de uma variedade de fontes de informação. Outro ponto foi a quebra das barreiras relacionadas ao compartilhamento de informação. Dessa forma, foi visto um aumento veemente na disponibilidade de informação (Oleto, 2006).

Entretanto, alguém que navegue na internet inevitavelmente se depara com uma variedade de conteúdos, que vão desde informações úteis até aquelas de qualidade duvidosa, devido à natureza aberta do sistema, em que qualquer pessoa pode publicar qualquer tipo de informação sem prévia avaliação. O aumento do volume de informações irrelevantes ressalta a importância de implementar filtros que possibilitem a recuperação de informações de qualidade e com maior precisão. Isso é de extrema importância, principalmente para quem utiliza essa rede global de comunicação para pesquisa e aprendizado (Edwards, 1998 como citado em Tomaél et al., 2001).

No âmbito acadêmico, é de suma importância avaliar cuidadosamente a fonte de informações, o que implica na identificação da pessoa ou organização responsável pela publicação. Isso pode ser feito por meio da investigação sobre quem escreveu uma obra e validar os títulos que demonstram sua proficiência para escrever sobre algo, o que engloba: a relevância de ser reconhecido na área, citações em outros trabalhos e estabelecer uma relação direta entre sua expertise e o conteúdo abordado.

Excluído: Investigar

Além da confiabilidade e credibilidade das fontes de informação, é igualmente importante que as informações sejam apresentadas de forma clara, simples e compreensível.

Nesse contexto, motores de busca que simplificam o acesso a informações de cunho acadêmico e possibilitam que seja feita uma análise da responsabilidade intelectual do conteúdo retornando, contribuem tanto para o êxito das pesquisas quanto no aprendizado e pensamento crítico de estudantes e pesquisadores.

METODOLOGIA

Com o intuito de solucionar os problemas supracitados, este trabalho foi desenvolvido por meio de uma pesquisa aplicada, de modo que fosse possível compreender, os problemas relacionados a dificuldade de acesso a informações de qualidade para fins acadêmicos, e desenvolver uma solução verossímil para esta situação.

Para fins de embasamento em relação ao objeto de estudo, inicialmente foi feita uma pesquisa bibliográfica. Dessa forma, foram revisados artigos acadêmicos e publicações, tanto sobre motores de busca quanto de pesquisa e acesso a informações no meio acadêmico. A partir desta pesquisa inicial, foram levantadas hipóteses de solução e os meios necessários para desenvolvê-las, o que envolveu a análise de requisitos funcionais e não funcionais do motor de busca.

Com o levantamento de requisitos em mãos e a fundamentação teórica inicial, foi iniciado o desenvolvimento do motor de busca com a utilização da linguagem de programação Python. Ela foi escolhida pelo fato de ser uma linguagem versátil que pode ser usada em uma ampla variedade de aplicações, desde desenvolvimento web e científico até automação de tarefas, análise de dados e inteligência artificial. Para a funcionalidade de busca e indexação de artigos, foi necessário realizar uma pesquisa exploratória em relação a como esse processo iria ser desenvolvido. Diante disso, foi optado pela utilização da técnica conhecida como web scraping, ou, raspagem de dados, que consiste em extrair dados de sites da web de forma automatizada. Isso é feito por meio análise do código em HyperText Markup Language (HTML), linguagem de marcação de texto utilizada para estruturação das páginas da web - para identificar os elementos que contêm os dados desejados e extrair esses dados de forma programática.

Entretanto, para aplicar essa técnica, foi utilizado o pacote BeautifulSoup, que já possui nativamente métodos voltados para a análise de conteúdo HTML, o que simplifica o processo de desenvolvimento e fez com que fosse possível se dedicar em outros pontos do projeto. Como

fonte de dados para a ferramenta de pesquisa, no caso, o site em que foi feito o processo raspagem de dados, foi escolhido o Google Acadêmico.

Além da indexação de artigos científicos, a solução tem como proposta retornar ao usuário métricas relacionadas a pesquisa. Desse modo, para a criação de tal recurso foi utilizado a biblioteca Pandas, um acrônimo para Python Data Analysis Library, que realiza uma análise exploratória nos dados retornados na pesquisa. Adicionalmente, para o recurso de inserção e análise de arquivos, também foi a Pandas, porém, em conjunto com a biblioteca LazyPredict. Essas duas ferramentas em conjunto realizam uma avaliação na base e retornam indicadores descritivos, como por exemplo, a quantidade de valores nulos e, caso o usuário deseje utilizá-la para alguma aplicação de aprendizado de máquina, também retorna o algoritmo mais adequado para isso.

Comentado [HO1]: É outra biblioteca

Para a criação da interface visual da aplicação, foi escolhido a ferramenta streamlit, uma biblioteca de código aberto em Python que permite criar aplicativos web interativos e visualizações de dados de forma simplificada. Com a utilização deste recurso, não foi necessário escrever um número considerável de linhas de código em HTML e Cascading Style Sheets (CSS), voltadas para marcação e estilização do conteúdo.

Com o objetivo de controlar e, gerenciar as diferentes versões dos arquivos de código durante o desenvolvimento do projeto, foi utilizado o software de versionamento Git, que foi fundamental para o rastreamento de erros e reversão de alterações indesejadas. Como metodologia de controle de versão, foi optado pelo GitFlow, que define um conjunto de regras e convenções que auxiliam no gerenciamento do fluxo de trabalho. Em relação ao gerenciamento do projeto como um todo, foi utilizado o aplicativo Microsoft Planner.

Como resultado, foi confeccionado um software que não apenas torna possível o acesso a artigos, revistas e publicações científicas de relevância acadêmica, bem como fomenta o aprendizado e aquisição do conhecimento por meio da pesquisa. Todavia, é válido destacar que a melhoria contínua do mecanismo de pesquisa é fundamental para que ela continue sendo relevante ao longo do tempo. Ao passo de que novas tecnologias surgem, e mais dados ficam acessíveis, esta ferramenta precisa evoluir para continuar a amparar as necessidades de estudantes e pesquisadores. O que significa que este projeto é apenas o início de um processo de estudo e aperfeiçoamento nessa alçada tecnológica.

Comentado [HO2]: Substituir por outra palavra.

ANÁLISE E DISCUSSÃO

APRESENTAÇÃO DE FORMA GERAL DOS RESULTADOS

DISCUSSÃO DOS RESULTADOS:

- Apresentar cada interface detalhadamente e suas funcionalidades;
- Interpretação dos resultados: discorra como os resultados se relacionam com as hipóteses iniciais ou os objetivos do estudo (Como os resultados se aplicam aos problemas levantados?);
- Implicações práticas: como o motor de busca pode ser utilizado na prática? Quais são os benefícios potenciais para os estudantes de ciência de dados?
- Possíveis limitações do estudo: reconhecer as limitações da ferramenta;

Formatado: Fonte: Negrito

Formatado: Fonte: Negrito

CONTRIBUIÇÕES DO ESTUDO:

- Contribuições acadêmicas: em que aspectos seu trabalho avança o conhecimento já existente;
- Contribuições práticas: como o motor de busca pode beneficiar os estudantes de ciência de dados?

Formatado: Fonte: Negrito

APRESENTAÇÃO DOS RESULTADOS:

A partir dos métodos supracitados, foi possível desenvolver uma ferramenta especializada para estudantes de ciência de dados. Como resultado, o motor de busca demonstrou ser um instrumento eficaz para facilitar o acesso a informações de qualidade e possibilitar que os estudantes selecionem os artigos e pesquisas mais apropriadas para realização de estudos de ciência de dados. O sistema foi desenvolvido com funcionalidades específicas para atender as necessidades dos estudantes, o que inclui a capacidade de inserir e avaliar bases de dados, e a possibilidade de escolher entre mais de 30 algoritmos de aprendizado de máquina o mais adequado para o seu projeto, sejam elas tarefas de regressão ou classificação.

Fato é que já existe uma ampla diversidade de ferramentas de busca e análise de dados disponíveis, contudo, a abordagem deste estudo se ressaí pela orientação específica para estudantes acadêmicos de Ciência de Dados, o que proporciona uma experiência de aprendizado prático por meio da análise e avaliação de bases de dados e a experimentação de diferentes algoritmos de machine learning. A união desses dois recursos em um único mecanismo de pesquisa, representa um diferencial relevante em relação a outras ferramentas convencionais de pesquisa.

Desta forma, foram desenvolvidas cinco interfaces, cada uma com propósitos distintos.

A primeira, definida como “Home”, foi criada com o objetivo de ser o ponto de partida para utilização do Data Discover, nome dado ao motor de busca. Como pode ser visto na figura 1, para fins informativos, há uma explicação inicial do que se trata o trabalho.

FIGURA 1

Para melhor organização da página inicial, foi feita a divisão de seu conteúdo em três sessões, Recursos principais, Informações sobre o Projeto e Como ele foi desenvolvido. Como pode ser observado na figura 2, ao clicar sobre o nome de cada um dos recursos do motor de busca, é retornado um texto explicativo sobre suas funcionalidades.

FIGURA 2

A sessão de Informações sobre o projeto trás de forma sucinta o contexto de seu desenvolvimento, os alunos responsáveis e o docente orientador, como demonstrado na figura 3:

FIGURA 3

Adicionalmente, no conteúdo final da página há uma sessão dedicada para elencar quais foram as tecnologias utilizadas e suas respectivas aplicações. As ferramentas podem ser observadas na figura 4:

FIGURA 4

A próxima página do sistema foi desenvolvida para a busca de publicações acadêmicas. Nela existe uma barra de pesquisa para que os usuários insiram o tema que desejam pesquisar. A partir dessa entrada de dados fornecida pelo utilizador, o mecanismo realiza a extração e indexação de artigos. Como pode ser visto na figura 5, é possível selecionar entre uma e quatro páginas para que seja feito o processo de webscrapping.

FIGURA 5

É possível observar na figura 6 que, como resultado, é retornado o título dos artigos, ano de publicação, pessoa autora, número de citações, site de publicação e o link do respectivo documento. Caso a pessoa estudante queira utilizar os artigos publicado para uma análise mais criteriosa, é possível baixar os resultados no formato csv.

Complementarmente ao recurso de download, abaixo dos resultados são criados dois gráficos, um com a distribuição de artigos publicados por ano e outro com a porcentagem de publicações por site de busca. Com o objetivo de permitir que fosse feita uma avaliação que leve em conta além de duas dimensões, no caso, index e ano de publicação, foi optado pela utilização do gráfico de bolhas para a distribuição dos artigos, que faz com que também seja possível considerar a quantidade de publicações. Para a porcentagem de publicações por ano, foi escolhido o gráfico de pizza. Ele permite visualizar rapidamente e de forma eficaz a distribuição percentual de cada página da internet. As duas visualizações podem ser vistas na figura 6.

FIGURA 6

Os artifícios de download de arquivos e visualização gráfica da distribuição dos documentos, juntos propiciam que o estudante realiza a análise das publicações e selecione o mais adequado para o seu projeto.

Na figura 7, temos a página de Análise Exploratória de dados, uma etapa importante durante a criação de um modelo preditivo. Nesta etapa é feita a investigação do conjunto de dados que será utilizado para treinamento, no intuito de descobrir padrões, identificar anomalias e entender a estrutura dos dados.

FIGURA 7

Para utilizar essa funcionalidade é necessário que o usuário insira um arquivo de até 200 mb em formato csv ou xlsx. Entretanto, com a finalidade de proporcionar uma melhor compreensão do que se trata a Análise Exploratória de Dados, e permitir que o usuário experimente a ferramenta mesmo que não tenha uma tabela para ser introduzida, foi deixado um conjunto de dados de exemplo.

Com o upload da tabela feito, a ferramenta realiza uma exploração sistemática e retorna as informações da tabela em seis seções diferentes. Conforme apresentado na figura 8, a primeira seção elucida uma visão geral em relação aos dados. Ela é dividida em três subseções diferentes, em que cada uma contém informações específicas a partir da análise realizada. A subseção “overview” é responsável por abordar valores quantitativos do dataset. Essas informações são importantes para que o estudante possa avaliar as características do conjunto de dados em que está trabalhando, como por exemplo o número de variáveis e observações.

Caso alguma coluna possua apenas valores únicos, esta situação é sinalizada em “Alerts”. No processo de criação de um modelo preditivo, é necessário que os dados sejam preparados para a etapa de treinamento e, uma coluna ou conjunto de dados com apenas valores únicos é um ponto de atenção. Caso seja optado pela utilização de um algoritmo que dependa da variação dos dados, essas colunas podem prejudicar o desempenho do modelo. Dessa forma, remover essas colunas pode reduzir a dimensionalidade do conjunto de dados e simplificar o treinamento. A subseção “Alerts” pode ser vista na figura 8.

FIGURA 8

Conforme demonstrado na figura 9, em “Reproduction” são retornadas informações a cerca do processamento da base de dados, entre elas o tempo demorado.

Na seção “Variables”, para que a pessoa estudante possa ter informações específicas a cerca de cada coluna, é possível selecioná-las individualmente. Como cada coluna representa um aspecto específico dos dados, compreendê-las de forma minuciosa é essencial para que possíveis outliers ou valores inconsistentes sejam encontrados. Como demonstrado na figura 10, inicialmente os usuários poderão encontrar informações gerais a respeito da coluna como a média e a quantidade de valores distintos. Além disso, a ferramenta também apresenta um histograma para caso o usuário prefira uma análise visual.

Contudo, como pode ser observado nas figuras 10, 11, 13 e 14, caso recursos adicionais sejam necessários, é possível expandir a quantidade de informações ao selecionar a opção “more details”. Elas são divididas em “Statistics” - Estatísticas, “Histogram” - Histograma, “Common Values” – Valores comuns e “Extreme Values” – Valores Extremos.

FIGURA 10, 11, 13 e 14

A próxima seção, chamada “Interactions”, foi desenvolvida com objetivo de identificar as relações entre as variáveis. Conforme demonstrado na figura 15, é possível selecionar duas colunas distintas para que seja produzido um gráfico de dispersão, o que permite avaliar se existe uma relação linear ou não.

FIGURA 15.

Na seção seguinte, chamada “Correlations”, também é utilizado um recurso visual para identificar a correlação entre as variáveis, como demonstrado na figura 16. Neste caso, trata-se de um gráfico de calor que representa a matriz de correlação entre todas as variáveis da base de dados inserida pelo usuário. Além da correlação propriamente dita, o heatmap permite constatar se existe a presença de multicolinearidade, que ocorre quando duas ou mais variáveis são altamente correlacionadas entre si – um comportamento que pode afetar a precisão de modelos de machine learning. Conforme a figura 17, também é disponibilizado para o pesquisador, a opção de analisar a matriz de correlação por meio de uma tabela.

FIGURA 16 e 17.

A divisão “Missing Values”, representada na figura 18, oferece para o usuário informações a respeito da presença de valores faltantes entre as colunas da tabela. Identificar esse cenário é importante para validar a qualidade das informações que serão utilizadas, o que é feito por meio da análise da distribuição de dados faltantes. No exemplo da figura 18, todas as colunas possuem 100% dos registros.

Também é possível verificar se existem linhas vazias por meio de uma matriz, como pode ser visto na figura 18.

A última divisão da página de Análise Exploratória de Dados é denominada “Sample”. Ela permite que usuário verifique os 10 registros iniciais e finais da tabela de forma simples, como pode ser observado nas figuras 19 e 20 respectivamente. Dessa forma, é possível realizar uma verificação inicial da integridade dos registros, como por exemplo, se há a presença de caracteres especiais.

FIGURA 19, FIGURA 20.

A próxima página do mecanismo de pesquisa é a Relatório de Performance de Modelo. Semelhante ao processo que é realizado na página anterior, também é necessário carregar uma base de dados para avaliação e, para fins explicativos, há um conjunto de dados de exemplo. Contudo, neste caso o que será avaliado é o desempenho de 42 algoritmos de aprendizado de máquina em relação as suas capacidades de prever a variável definida como alvo. Conforme a figura 21, após a base de dados ser inserida ou ter sido selecionada a de exemplo, é necessário que o usuário defina o percentual dos dados que será utilizado para o treinamento dos modelos e os respectivos testes. Também é necessário selecionar um valor para a aleatoriedade dos modelos, que consiste em um número utilizado para definir como o algoritmo irá gerar os valores para seus parâmetros internos. Dessa forma, ao utilizar em outras execuções o mesmo valor para aleatoriedade, é possível reproduzir os resultados anteriores.

FIGURA 21

A divisão do conjunto de dados em treino e teste é uma etapa importante no processo de desenvolvimento de um modelo de machine learning. No caso, o conjunto de treino será utilizado para que o modelo aprenda os padrões presentes nos dados e seja capaz de realizar previsões. O conjunto de teste é utilizado para avaliar o desempenho do modelo com dados em que ele não foi exposto, ou seja, não teve seus padrões analisados. O que é essencial para avaliar a capacidade de generalização do modelo, que consiste em realizar previsões para valores além dos de treinamento.

Em sequência são retornadas informações em relação as dimensões da base de dados e suas variáveis, no caso a quantidade de linhas e coluna. Além disso, quais são as colunas utilizadas como variáveis preditoras e qual será a variável resposta. Por padrão a ferramenta considera a última coluna do conjunto de dados como variável dependente, a que se deseja prever. As outras colunas são utilizadas como variáveis independentes. As duas informações podem ser vistas na figura 22.

FIGURA 22

Após o processamento dos dados ter terminado, como pode ser visto nas figuras 23 e 24, por meio de duas tabelas, a pessoa estudante pode avaliar o desempenho dos algoritmos nos dados de treinamento e teste. Existem três métricas que representam como os modelos se saíram, e o tempo levado na computação em cada um dos algoritmos. Como a ferramenta é voltada para problemas de regressão, em que o objetivo é prever um valor numérico contínuo, as três métricas foram escolhidas para permitir que a pessoa estudante selecione o modelo mais adequado. Isso pode ser feito tanto pela análise individual quanto de forma conjunta, visto que, uma complementa a outra.

FIGURA 23

FIGURA 24

Além da tabela, existem dois gráficos que também são utilizados para representar a performance e o tempo de execução entre os diferentes algoritmos de machine learning. Os gráficos podem ser vistos nas figuras 25 e 26 respectivamente.

FIGURA 25

FIGURA 26

Prezando pelo aprimoramento contínuo da ferramenta, foi elaborada uma página que contém as instruções de como realizar o download do código-fonte da ferramenta de pesquisa. Diante disso, é possível que outros pesquisadores revisem e reproduzam o trabalho.

Como ilustrado nos procedimentos da figura 27, o código-fonte do projeto está armazenado em um repositório público do GitHub, uma plataforma de hospedagem de códigos que permite gerenciar e compartilhar projetos de forma gratuita. Portanto, basta que o estudante clone o repositório e instale os pacotes do motor de busca, para que ele consiga utilizá-lo localmente.

A ferramenta foi projetada para ser transparente e de código aberto, o que incentiva a colaboração e seu aperfeiçoamento constante. Desta maneira, a página seguinte nomeada “Fale Conosco” é a porta de entrada para que atualizações sejam sugeridas. Como demonstrado na figura 28, existem dois canais para se comunicar com os mantenedores da ferramenta, o GitHub para propor alterações diretamente no código, e o LinkedIn, uma rede social profissional em que é possível entrar em contato com os pesquisadores que desenvolveram o projeto.

Atentar para o fato de que um texto tem ordem cronológica, lógica e observa o grau de importância, impacto e a relação existente entre as informações.

Este tipo de cuidado permite que se apresente uma visão coerente dos fatos e contribui para o resultado da pesquisa.

Uma vez que o autor apresentou base teórica, fez todo processo de coleta de informações é hora de apresentar o que representam as informações coletadas, fazer análise de contribuições, efeitos ou mesmo ganhos obtidos por uma determinada abordagem

CONSIDERAÇÕES FINAIS

Quando estudo são qualitativos os estudos apresentam considerações finais, já os estudos quantitativos possuem ‘conclusão’, ambos servem para realizar o encerramento das ideias abordadas no estudo, apresentado tudo que houve de ganho em conhecimento.

REFERÊNCIAS

- Ribeiro, M. R. (2019). *Educação e política em Platão: reflexões a partir do livro VII d' A República* [Dissertação de Mestrado, Universidade federal de Alfenas]. Biblioteca Digital de Teses e Dissertações. <https://bdtd.unifal-mg.edu.br:8443/handle/tede/1485>
- Cunha, J. D. A. (2009). *Bibliotecas Digitais de Teses e Dissertações: uma estratégia de preservação da memória* [Dissertação de Mestrado, Universidade Federal da Paraíba]. Repositório Institucional da UFPB. <https://repositorio.ufpb.br/jspui/handle/tede/3989>
- Calazans, A. T. S. (2008). Qualidade da informação: conceitos e aplicações. *TransInformação*, 20(1), 29-45. <https://www.scielo.br/j/tinf/a/hfvRLR68SKzJrtDQ3DqGKLw/?lang=pt#>
- Oleto, R. R. (2006). Percepção da qualidade da informação. *Ciência da Informação* 35(1), 57-62. <https://doi.org/10.1590/S0100-19652006000100007>

Tomaél, M. I., Catarino, M. E., Valentim, M. L. P., de Almeida Júnior, O. F., & da Silva, T. E.

(2001). Avaliação de fontes de informação na Internet: critérios de qualidade.

Informação & *Sociedade*, 11(2).

<https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/293>