MESA-REDONDA - PAINEL DE FERRAMENTAS BASEADAS EM INTELIGÊNCIA ARTIFICIAL PARA APOIO À METODOLOGIA DE PESQUISA

CEFET/RJ

USO DE LLM PARA APOIAR A ESCRITA E REVISÃO DE ARTIGOS

Eduardo Ogasawara
eogasawara@ieee.org
https://eic.cefet-rj.br/~eogasawara

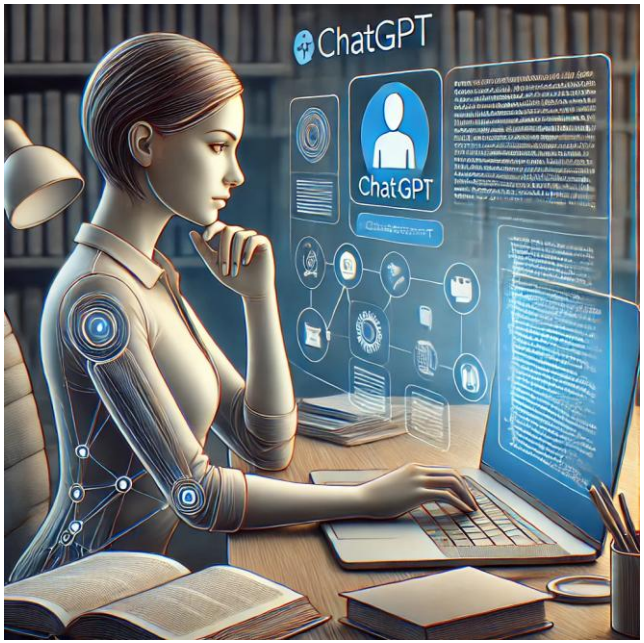
SEMANA NACIONAL DE CIÊNCIA E TECNOLOGIA 2024
SEMANA DE ENSINO, PESQUISA E EXTENSÃO 2024 - CEFET/RJ

O que são Grandes Modelos de Linguagem (LLMs)?

- Definição
  - Modelos de linguagem treinados com grandes volumes de dados textuais que compreendem e geram linguagem natural
- Exemplos
  - GPT, BERT, ChatGPT
- Capacidades
  - Responder a perguntas, completar textos, sugerir melhorias de estilo, resumir conteúdo, entre outros

LLMs no Processo de Escrita de Artigos

- Auxílio na geração de ideias
  - Brainstorming de tópicos e organização de ideias iniciais
- Estruturação
  - Sugestões para estrutura de seções
- Geração de rascunhos
  - Escrever trechos ou parágrafos com base em instruções
- Correção gramatical e estilística
  - Melhorias em fluidez, gramática e coesão

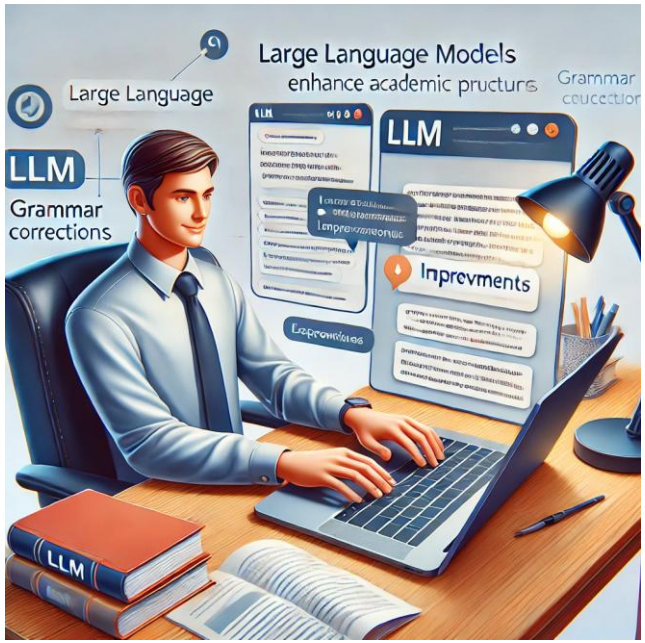


LLMs no Processo de Revisão de Artigos

- Revisão de gramática e estilo
  - Sugestões de melhorias de linguagem, gramática e estilo acadêmico
- Identificação de inconsistências
  - Verificação de coerência entre diferentes seções
- Sugestão de melhorias
  - Recomendações para clareza e coesão
- Revisão "crítica"
  - Auxílio na identificação de pontos fracos no argumento ou na metodologia

Benefícios do Uso de LLMs

- Aumento de produtividade
  - Redução de tempo na escrita e revisão
- Feedback instantâneo
  - Sugestões imediatas sem a necessidade de revisões humanas
- Auxílio na organização de ideias complexas
  - Clareza e estruturação de argumentos
- Padronização do estilo acadêmico
  - Consistência conforme padrões acadêmicos
- Revisão automática inicial
  - Identificação de erros antes da submissão
- Suporte à análise de literatura
  - Resumos automáticos de artigos relacionados



Apoio na organização das ideias

Monte uns slides powerpoint sobre uso do LLM para apoiar a escrita e revisão de artigos.

Para criar slides sobre o uso de Modelos de Linguagem Grande (LLM) para apoiar a escrita e revisão de artigos, sugiro a seguinte estrutura, que você pode adaptar conforme necessário:

Escreva um resumo de um paragrafo para um painel de "Painel de Ferramentas Baseadas em Inteligência Artificial para Apoio à Metodologia de Pesquisa"

O painel "Ferramentas Baseadas em Inteligência Artificial para Apoio à Metodologia de Pesquisa" explora como tecnologias de IA estão transformando o campo da pesquisa científica, oferecendo soluções para otimizar o processo de coleta, análise e interpretação de dados. As ferramentas de IA proporcionam suporte em diversas etapas da pesquisa, desde a revisão de literatura automatizada até a análise preditiva e criação de insights, permitindo aos pesquisadores economizar tempo e aumentar a precisão de seus estudos. O painel discutirá as principais ferramentas disponíveis, suas aplicações práticas e os desafios éticos e metodológicos envolvidos no uso dessas tecnologias emergentes no contexto acadêmico.

Apoio a construção de taxonomia

time series data preprocessing

Time series data preprocessing is an important step in analyzing and modeling time series data. It involves preparing the data for analysis by performing various operations on the data such as cleaning, transforming, and filtering.

Here are some common steps in time series data preprocessing:

1. Cleaning the data: This involves handling missing values, outliers, and errors in the data. Missing values can be filled in using techniques such as interpolation or imputation. Outliers can be detected using statistical methods and removed or replaced.
2. Resampling the data: Depending on the frequency of the data, it may need to be resampled to a higher or lower frequency to match the desired analysis time frame. This can be done by aggregating or interpolating the data.
3. Normalizing the data: This involves scaling the data to a common range or unit so that it can be compared or combined with other data. Common normalization techniques include standardization, min-max scaling, and log transformation.
4. Feature engineering: This involves creating new features or variables from the existing data to capture additional information or relationships. Common feature engineering techniques include lagging, differencing, rolling windows, and Fourier transforms.
5. Splitting the data: It is important to split the data into training, validation, and testing sets to evaluate the performance of the model and avoid overfitting.

Desenvolvendo e refinando a taxonomia

What is a good reference for that?

There are several good references for time series data preprocessing. Here are a few:

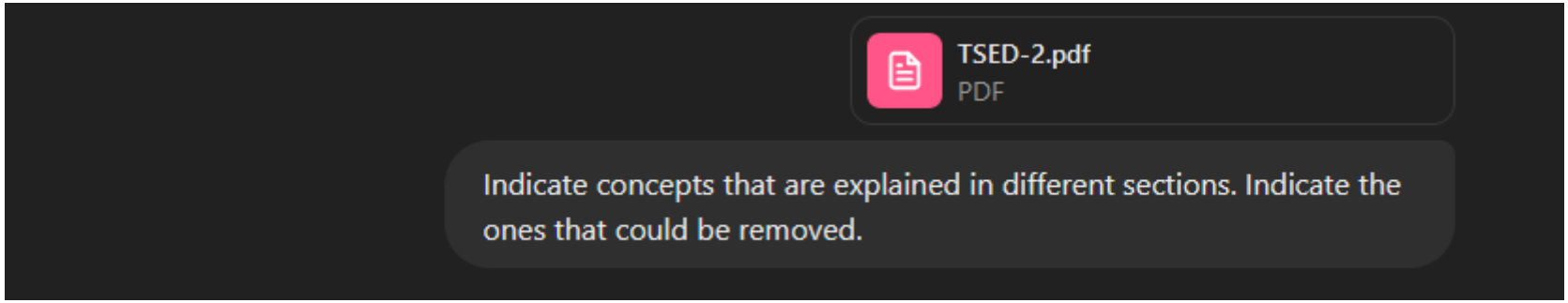
Explain Splitting the data for time series

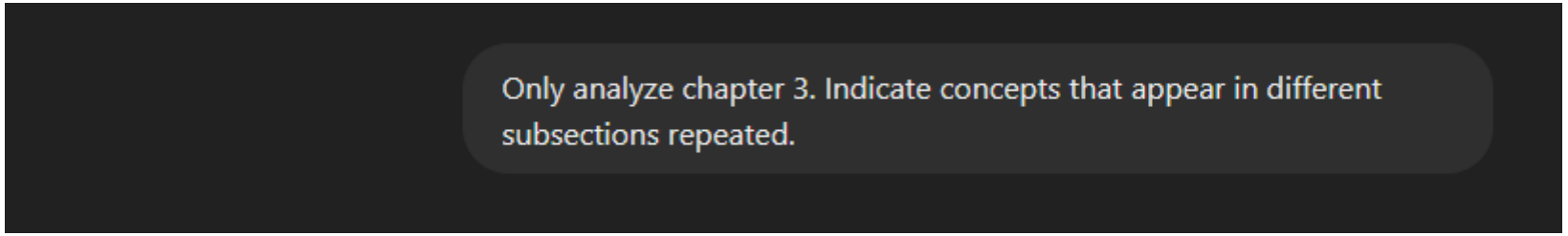
Splitting the data is an essential step in time series analysis and modeling. It involves dividing the available time series data into different sets for training, validation, and testing purposes.

in what category of time series preprocessing does percentage change transform is part of?

Percentage change transform is a data transformation technique used in time series data preprocessing, specifically in the category of feature engineering techniques.

Reorganização de seções





Repeated Concepts Across Subsections

Supervised, Semi-Supervised, and Unsupervised Learning

- Discussed in 3.1.5 as types of learning models.
- These concepts also inherently appear within 3.2 (especially in 3.2.1 Regression, 3.2.2 Classification, and 3.2.3 Clustering) as methods are categorized by how they learn from data.

Reorganização de seções

Are some methods explained twice? Where?

Yes, some methods appear to be explained more than once across different subsections. Here is an analysis of where these methods are discussed:

Repeated Methods

Regression Methods

- **Section 3.2.1 Regression:**
  - This section is dedicated to discussing regression methods in detail, including their application in anomaly detection.

Suggestions to Avoid Repetition

1. **Consolidate Discussions on Supervised, Semi-Supervised, and Unsupervised Methods:**
   - Introduce these learning types in one place (either in 3.1.5 or as an introduction to 3.2 Methods) and reference them in the detailed methods sections.
2. **Unify Statistical Methods:**
   - Detailed discussions in 3.2.4 can be referenced in other sections to avoid redundancy.
3. **Combine Spectral Methods Descriptions:**
   - Keep the detailed explanation in 3.2.5 and reference it in 3.3.2 to avoid repeating the same content.
4. **Graph-Based Methods:**
   - Focus the detailed explanation in 3.3.3 and ensure any other mention points back to this section for detailed reading.

Ajuste de estilo

Consider the following guidelines. I will send you a text for you to adjust.

guideline #1: You have many small sentences of the form « subject verb complement » but when they are related (same point, same idea), it is better to group them together, e.g., with « and » or « which » ... Note that good writing consists of a mixture of short and long sentences. Use short sentences to have punch and emphasize important points, longer sentences for descriptions.

guideline #2: remove unnecessary Adjectives, Adverbs, Adjectival nouns

guideline #3: remove "common", "commonly", "generally" and other words that would give the impression that the text is not sure.

guideline #4: sentences with less than 20 words should be connected to the following using connections "and", "which" etc.

guideline #5: do not remove latex commands.

guideline #6: convert sentences in the past to present.

Understood! Feel free to send me the text, and I'll adjust it according to the provided guidelines.

Apoio na inclusão de citações

Revisando citações

Check if the citations in the following chapter are appropriate. Indicate if some are wrong, in the sense that it is not related to the provided text, or if we have any citation missing.

\myabstract{Motif discovery refers to identifying recurring sequences in a time series. An important property regarding motifs is that the

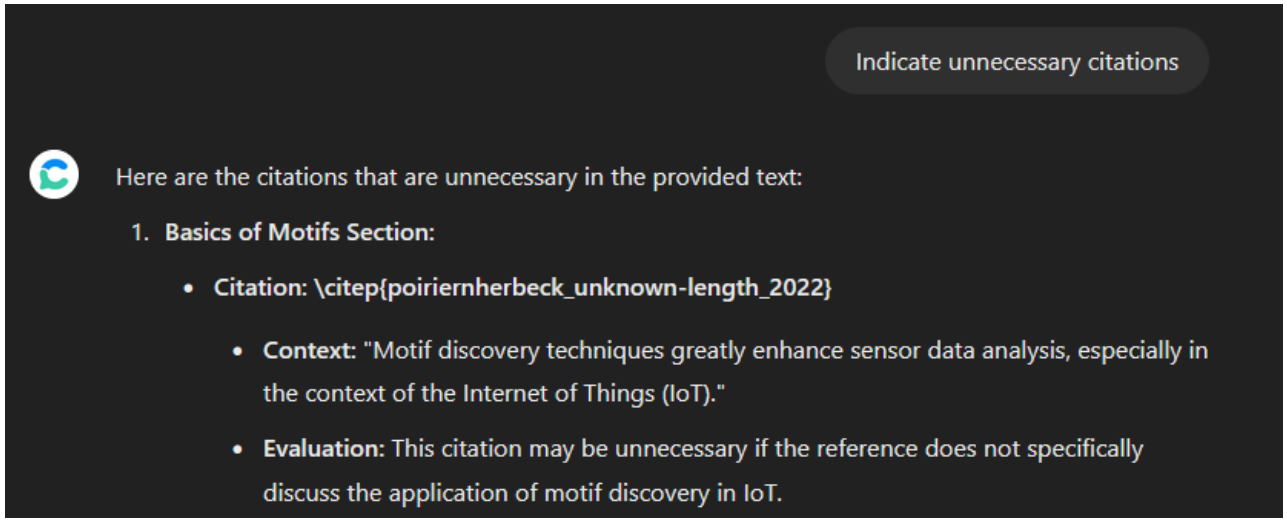
Just report the ones that are not appropriate

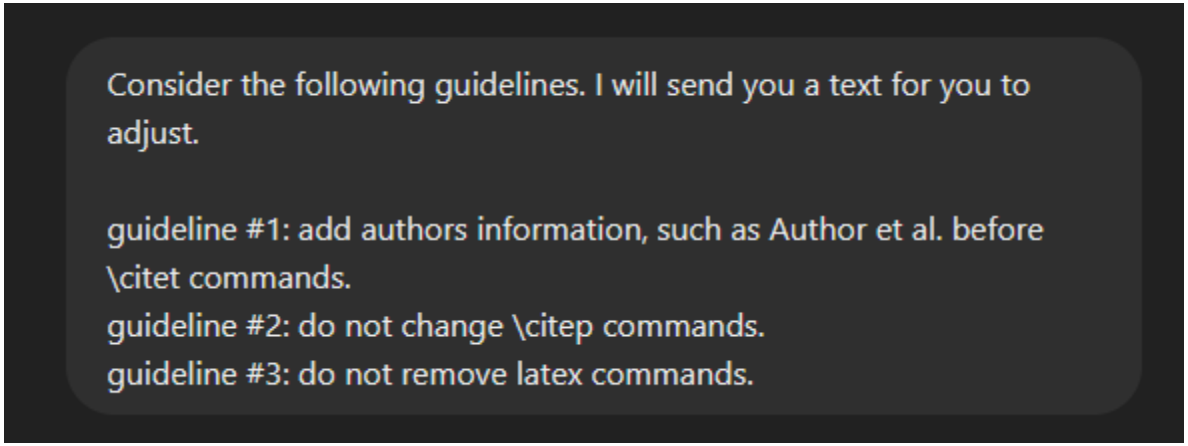
Here are the citations that are not appropriate in the provided text:

1. **Basics of Motifs Section:**
   - **Citation:** **\citep{poiriernherbeck_unknown-length_2022}**
   - **Evaluation:** This citation is not directly related to the context of motif discovery in IoT and sensor data analysis.
2. **Time Series Indexing Section:**
   - **Citation:** **\citep{ogasawara_neural_2009}**
   - **Evaluation:** This citation is not directly related to the discussion of Z-Score normalization in the context of motif discovery.

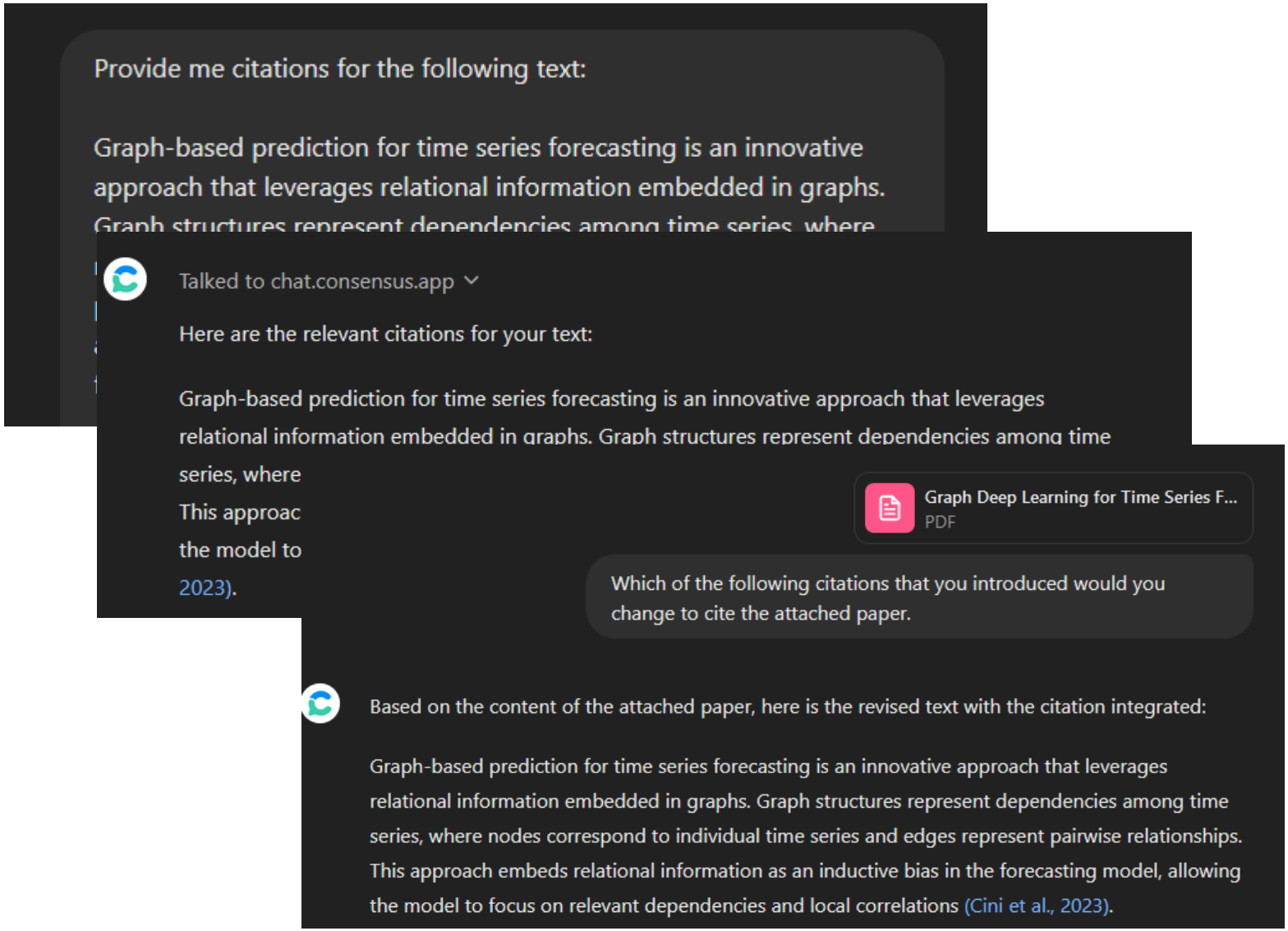
These citations should be reviewed and potentially replaced with more relevant sources.

Revisando citações





Inclusão de citações



Revisando o artigo

- Considere este artigo:
- <texto ou PDF>
- Qual é a definição do problema que ele aborda?
- Qual é a contribuição científica?
- Qual é o método que resolve o problema?
- O método está bem formalizado?
- A seção de resultados avalia bem a resposta da definição do problema e explora bem o método proposto?
- A conclusão clarifica o trabalho?
- Há alguma sugestão para a escrita da introdução?
- Há alguma sugestão para a escrita dos trabalhos relacionados?
- Há alguma sugestão para a escrita dos fundamentos?
- Há alguma sugestão para a escrita dos métodos?
- Há alguma sugestão para a escrita dos resultados?
- Há alguma sugestão para a escrita das limitações?
- Há alguma sugestão para a escrita das conclusões?

Avaliando o artigo

- Forneça-me uma análise dos pontos fortes e fracos deste artigo.
- Caracterize os pontos fortes e fracos do artigo.
- Analise o resumo fornecido, está faltando algo importante?
- Considere a introdução. Ela conduz bem à definição do problema, objetivo e contribuições?
- A introdução omite algum ponto importante?
- Considere o referencial teórico. O referencial apoia o método? Está faltando algo importante? Está mencionando algo desnecessário?
- Os trabalhos relacionados estão bem caracterizados e apoiam a novidade do artigo? Está faltando algo importante?
- O método caracteriza bem uma solução para a definição do problema? Ele sustenta bem os resultados? Está faltando algo importante?
- Os resultados respondem às questões de pesquisa? Eles avaliam o método? O artigo faz uma boa contribuição? Está faltando algo importante?
- A conclusão está bem integrada com o restante do artigo? Está faltando algo importante?

Analise-o na régua do fórum

- Agora, analise este artigo: (PDF)
- Escreva um único parágrafo com uma avaliação geral do artigo.
- Explique por que o artigo deve ser rejeitado, use esse raciocínio para as solicitações a seguir.
- Preencha este formulário:
- Relevância para o <evento>:
  - 5: Forte Aceitação
  - 4: Aceitação
  - 3: Aceitação Fraca
  - 2: Rejeição Fraca
  - 1: Rejeição
- Originalidade:
  - 5: Forte Aceitação
  - 4: Aceitação
  - 3: Aceitação Fraca
  - 2: Rejeição Fraca
  - 1: Rejeição
- Qualidade Técnica:
  - 5: Forte Aceitação
  - 4: Aceitação
  - 3: Aceitação Fraca
  - 2: Rejeição Fraca
  - 1: Rejeição
- Impacto das ideias e resultados:
  - 5: Forte Aceitação
  - 4: Aceitação
  - 3: Aceitação Fraca
  - 2: Rejeição Fraca
  - 1: Rejeição
- Apresentação:
  - 5: Forte Aceitação
  - 4: Aceitação
  - 3: Aceitação Fraca
  - 2: Rejeição Fraca
  - 1: Rejeição
- Avaliação Geral:
  - 5: Forte Aceitação (trabalho excepcional, publicável como está)
  - 4: Aceitação (bom trabalho, publicável)
  - 3: Aceitação Fraca (revisões menores)
  - 2: Rejeição Fraca (revisões maiores)
  - 1: Rejeição (não publicável)
- Potencial para ser o Melhor Artigo Curto:
  - 6: Muito Alto
  - 5: Alto
  - 4: Acima da Média
  - 3: Abaixo da Média
  - 2: Baixo
  - 1: Muito Baixo
- Experiência do revisor no tópico:
  - 4: Especialista
  - 3: Conhecedor da Área
  - 2: Ciente do tópico
  - 1: Teve que usar o senso comum
- Principais contribuições (até um parágrafo):
- Aspectos positivos (por favor, numere-os P1, P2,...) :
- Aspectos negativos (por favor, numere-os N1, N2,...):
- Comentários detalhados (por favor, numere-os C1, C2, ...):