**Cronograma de Execução**

| **Etapas/Mês** | **Abril** | **Maio** | **Junho** | **Julho** | **Agosto** | **Setembro** | **Outubro** | **Novembro** | **Dezembro** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Definição do Escopo do Trabalho | X | X |  |  |  |  |  |  |  |
| Pesquisa por Trabalhos Relacionados | X | X | X | X | X | X |  |  |  |
| Montagem/Coleta da Coleção de Dados |  | X | X | X | X |  |  |  |  |
| Pré-Processamento do texto |  | X | X | X | X |  |  |  |  |
| Classificação Binária |  |  |  | X | X | X |  |  |  |
| Qualificação do TCC |  |  |  |  |  |  | X |  |  |
| Envio do Texto para Banca |  |  |  |  |  |  |  | X |  |
| Escrita do texto Final de TCC | X | X | X | X | X | X | X | X | X |
| Defesa |  |  |  |  |  |  |  |  | X |

**Reunião 19/05/2025 -**

* Estiveram presentes……
* Analisamos os trabalhos relacionados coletados até o momento e verificamos quais são os padrões de métricas e modelos para definirmos nossa metodologia
* Falamos sobre definir o escopo do trabalho
  + TCC
  + Quais métodos vamos utilizar?
  + Vamos utilizar uma coleção de dados (agrupada) ou várias? - Decidimos utilizar as coleções de dados separadas, e usar uma coleção de dados baseada na união de uma ou mais coleções de dados (para tratar o desbalanceamento de dados)
  + Falamos sobre o pipeline de classificação e quais são as etapas que vamos tomar
    - Pré-processamento disponível em <https://journals-sol.sbc.org.br/index.php/reic/article/view/2696/1954>
    - geração de word cloud
* Orientador vai enviar as coleções de dados disponíveis
* Orientandos vão continuar trabalhando nos Trabalhos Relacionados (preenchimento de tabela)
* Quais categorias podemos trabalhar?
* Técnicas de pré-processamento

26 de maio de 2025

* Definição de quais coleções de dados serão utilizados (hateBR, Offcom2, OLID-BR e ToLD)
* Definição de função de pré-processamento. Os estudantes devem analisar a coleção de dados para definir quais serão as funções.
* Anotação da coleção de dados HateBR
  + Definição de legenda (2-racismo, 3-sexismo , 4- homofobia, 5-xenofobia)
* No dia 02/06

**02/06**

* Definiu que vamos re-rotular a coleção de dados OffComBr-3
  + Definição de legenda (2-racismo, 3-sexismo , 4- homofobia, 5-xenofobia)

- Começar a escrita da Fundamentação teórica e Introdução

- Exemplo de Introdução <https://sol.sbc.org.br/index.php/ercas/article/view/9052/8954>

-Reprodução do tutorial : <https://github.com/jiayao99/tfidf-text-classification/blob/main/resume_classification_tfidf.ipynb>

**09/06/2025**

* Offcom-Br3 -> rotulação baseado nas 6 classes
* HateBR -> exemplo de coleção de dados balanceada

Resultados :

* OffcomBR2 (binária)
* OffcomBR3 (binária)
* OffcomBR3 (reclassificada)
* HateBR (binária e balanceada)
* OLID (binária)
* ToldBR (somente binário - Converter para binária)
* ALL -> (standby)

**16/06**

**Pautas** :

* Vocês precisam trabalhar no texto! Isso é nossa urgência
* ~~Como está a rotulação das coleções de dados? Ok!~~
* ~~Melhor deixar a escrita da Metodologia pra depois, quando já tivermos tudo mais bem definido (Matheus tinha falado sobre escrever a parte de pré-processamento de texto)~~
* ~~Tutorial Overleaf~~

**Encaminhamentos** :

- **Introdução :** Lázaro

- **Trabalhos relacionados :** Izabelly, Ana Clara, Edkelly

- **Reproduzir tutorial :** Matheus

~~-~~ **~~Rotulação da base de dados~~**

**- (Orientador) - Olhar a coleção de dados re-rotulada . Colocar o label final**

**Passo-a-passo para inserir citação**

* Busca o nome do artigo em Google Scholar
* Clique em “Citar” e copie o código Bibtex
* Cole o código no arquivo refbase.bib
* Para fazer a citação, use \cite{codigo}, onde código é a chave

**Exemplo :** Ao pesquisar o artigo “Using digital game as clinical screening test to detect color deficiency in young children” no Google Scholar, cliquei em citar e em seguida em “bibtex”. Foi gerado o código abaixo :

@inproceedings{nguyen2014using,

title={Using digital game as clinical screening test to detect color deficiency in young children},

author={Nguyen, LinhChi and Lu, Weiquan and Do, Ellen Yi-Luen and Chia, Audrey and Wang, Yuan},

booktitle={Proceedings of the 2014 conference on Interaction design and children},

pages={337--340},

year={2014}

}

Copiei todo o código em vermelho e colei no arquivo refbase.bib (no overleaf). Após isso, sempre que eu quero fazer uma citação a esse artigo eu uso o comando \cite{nguyen2014using}

**03/07/2025**

* Como está o andamento da escrita do texto?
  + Introdução está sendo escrita
  + **Precisamos avançar nos Trabalhos Relacionados**
* Desenvolvimento do software
  + Conseguiram realizar a reprodução do tutorial?
  + Como está a definição do pré-processamento?
    - Vamos definir agora um modelo de pré-processamento mais básico
      * lowercase, remoção de nomes de usuários, remoção de urls, lematização (manter o lema das palavras)
* Iniciar classificação binária. Para cada coleção de dados :
  + Gerar métricas (acc, f1)
  + gerar wordcloud
  + gerar matriz de confusão

**Encaminhamentos :**

* **Finalizar a escrita dos trabalhos relacionados**
* **Desenvolvimento da função de pré-processamento (vide arquivo TOLD\_.ipynb)**
* **Pesquisar quais classificadores queremos incluir (NB, Random Forest, SVM Linear (LinearSVC()), Regressão Logística, KNN, ensemble(**[**https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingClassifier.html**](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingClassifier.html)**))**
* **explicação do código**

**14/07/2025**

* Verificar andamento dos Trabalhos Relacionados
* Definição dos modelos a serem utilizados.
  + Verificar andamento do desenvolvimento do código

**Geração de Tabelas**

* <https://www.tablesgenerator.com/>
* Utilizar esse site para gerar tabelas, clicar em “Gerar” e colar no texto

**04/08/2025**

* Terminar a escrita da seção de Trabalhos Relacionados;
* Começar a escrita do Capítulo 3
  + pré-processamento
  + coleção de dados
  + métodos de classificação (<https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html>)
* Desenvolvimento de código