Critérios de Avaliação do Miniprojeto:

• Entrega de documento escrito, contemplando:

o Artigo escolhido (título, autores, nome do periódico, ano de publicação), o qual

deve ter sido publicado de 2021 a 2025, em um periódico com fator de impacto

(Web of Science), envolvendo aplicação de qualquer técnica de inteligência

computacional relacionada à disciplina

o Detalhes do miniprojeto:

▪ Descrição da base de dados utilizada

▪ Tamanho da população (em se tratando de inteligência de enxames)

▪ Percentual de dados utilizados para treinamento e teste

▪ Uso de validação cruzada?

▪ Algoritmo(s) considerado(s) / implementado(s)

▪ Critérios de escolha dos hiperparâmetros (Grid Search?)

▪ Uso de transfer learning?

▪ Uso de técnicas de regularização? Dropout?

▪ Uso de aumento de dados?

▪ Como lidar com desbalanceamento de classes? SMOTE? Tomek links?

▪ Figuras de mérito utilizadas para a avaliação de desempenho. Acurácia,

precisão, recall...

o Informação das contribuições de cada integrante da equipe para o

desenvolvimento do miniprojeto (implicação na nota individual)

o Informação de como foi feita a implementação

o Fontes consideradas para fins de implementação (githut, sites de internet, tais

como página de grupo de pesquisa...)

o Modificação realizada na técnica considerada e/ou avaliação de aspectos não

considerados no artigo

• Entrega dos slides

• Tempo de apresentação da equipe: o tempo mínimo individual é de 5 minutos –

implicação na nota individual. Em seguida, será realizada a etapa de perguntas e

comentários

• Domínio do conteúdo

• Qualidade dos slides

• Segurança / desempenho nas respostas às perguntas

Apresentação:

1. Mostrar artigo
2. Explicar motivação do artigo
3. Mostrar as acurácias e modelos testados
4. Explicar pq o tempo de treinamento do modelo com acurácia maior demora tanto e pq não trabalhamos com ele- O que é BERT e RoBERT
5. Falar das dificuldades antes de começar o projeto
6. De onde vieram os dados
   1. Requisitos para extração
   2. Ferramentas usadas para extrair
   3. Tamanho dos dados total (2 bases)
   4. O que tem de errado com o dado bruto que pode atrapalhar o processamento
7. Parte teórica: tratamento de dados
   1. O que precisou ser tratado
   2. Mostrar documentação do GitHub
   3. Falar do tamanho dos dados
   4. Como os dados estão divididos (proporção de cada classe)
   5. Está balanceado? NÃO! Mais pra frente explicar SMOTE
8. Falar das métricas usadas no artigo
9. Modelos usados no artigo:

\*para cada modelo falar rapidamente como funciona e mostrar os dados que o artigo disponibiliza sobre cada um (f1 score)

* 1. Bert model
  2. Random Forest
  3. Suport vector machine
  4. Redes neurais

1. Mostrar o modelo Bert e explicar por que optamos por não utilizar ele
2. Guia do artigo de pré-processamento e tratamentos com dados pra treinamento (K-fold e GridSearch)
3. Melhorias: balanceamento
4. Parte prática – Mostrar o código
   1. Bibliotecas (que o artigo também usa, tem a do twitter, Re, SMOTE e stopwords que precisa explicar brevemente o funcionamento)
   2. Pré-processamento do texto
   3. Explicar processo de tokenizacao e vetorização
   4. Mostrar divisão de dados (assim como no artigo 70/30)
   5. CV o que é e para que serve
   6. GridSearch – parâmetros + escolha de parâmetros
      1. Diferença dos parâmetros dados
   7. Mostrar que foi usado o SMOTE em um código mas abaixou a acurácia (explicar pq isso acontece)
5. Análise dos resultados:
   1. Tempo de compilação do gridSearch
   2. Tempo de treinamento
   3. Análise dos resultados de cada modelo, passos para cada:
      1. Mostrar resultado obtido no artigo
      2. Mostrar resultado obtido na implementação com dados do artigo
      3. Mostrar resultado obtido na implementação com dados obtidos no meu GidSearch
      4. Comparativo entre modelos e dados(2 modelos e 2 datasets)

//quantas vezes foi citado?

//procurar se ele foi citado em algum lugar e esse lugar teve resultados parecidos com os nossos

//está balanceado? Não

//organização dos dados

Antes do SMOTE: Counter({'Aprobación/Empatía/Confianza': 490, 'Desaprobación': 345, 'Indeterminado': 205, 'Decepción/Tristeza': 196, 'Interés/Anticipación/Hype': 185, 'Enfado/Ira': 129})

Antes do SMOTE: Counter({'Negativo': 715, 'Positivo': 674, 'Indeterminado': 161})

//buscar outras melhorias