

Dados da Bolsa**Tipo de Bolsa:** ☒ IC ☐ PqEP ☐ TCC ☐ ME ☐ DO ☐ PD**Nome do Orientador/Supervisor:** Anna Helena Reali Costa**Nome do Projeto:** Chatbot Q&A multi-agente**Período da Bolsa:** 01/02/2021 a 01/02/2022**Relatório:** ☐ Final ☒ Parcial**Período do Relatório:** 25/05/2021 a 25/06/2021**Descrição das Atividades de Pesquisa do Projeto****Descrição das atividades acadêmicas:**

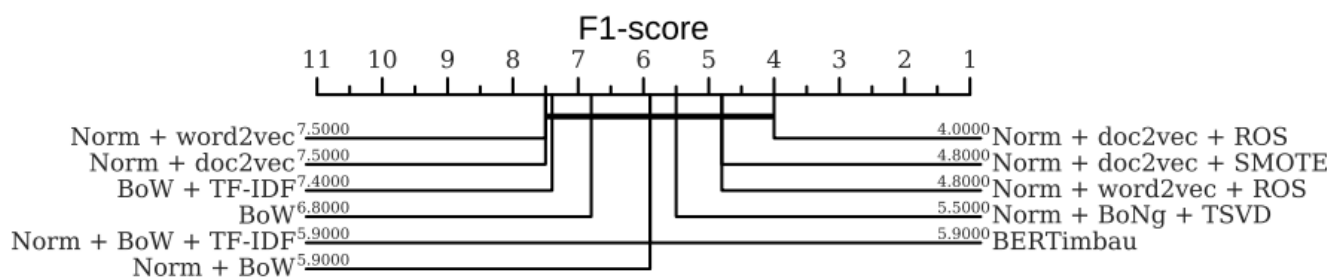
- PSI3211 - Circuitos Elétricos I
- PCS3115 - Sistemas Digitais I
- PSI3212 - Laboratório de Circuitos Elétricos
- PEF3208 - Fundamentos de Mecânica das Estruturas
- PSI3321 - Eletrônica I
- MAP3121 - Métodos Numéricos e Aplicações
- 4323301 - Física Experimental C
- SCC0633 - Processamento de Linguagem Natural

Descrição das atividades de pesquisa:

Durante o último mês o foco majoritário foi finalizar nosso artigo que foi submetido para o BRACIS 2021. O foco do artigo foi um teste de ablação de múltiplos classificadores e também de múltiplas condições dos dados para a tarefa de classificar uma interrupção ou continuidade das falas do Presidente do Diário da Assembleia da República Portuguesa, com a finalidade de separar blocos de falas a respeito de um mesmo assunto/tema.

Como dados, usamos as falas do Presidente em diversas configurações, usando diferentes técnicas de extração para features de textos como o Bag of Words, TF-IDF, Bag of N-Grams, Truncated SVD, word2vec, doc2vec, Random Over Sampling e SMOTE, bem como as técnicas de normalização de textos como tokenização, remoção das stopwords e também lematização.

No artigo defendemos que apesar do BERTimbau (versão para português do Bidirectional Encoder Representations from Transformers [BERT]) ser uma poderosa técnica no âmbito de NLP, podemos adquirir classificadores estatisticamente equivalentes com modelos muito mais simples, com menor tempo de treinamento e gasto computacional. Para isso testamos KNN, Linear Regression, Random Forest, SGD, SVM e XGBoost. Com uma correta busca de hiperparâmetros, conseguimos, para todas as configurações de dados, modelos que são estatisticamente equivalentes ao BERTimbau utilizando uma análise par a par de Wilcoxon-Holm, conforme ilustra a figura a seguir.



Houve alteração no cronograma original: () Sim (x) Não

Justifique em caso positivo:

Apreciação Circunstanciada do Orientador/Supervisor sobre as Atividades do Bolsista

Apreciação:

Etapas cumpridas no relatório: (x) Ótimo () Bom () Regular () Fraco

Programação para a próxima etapa: () Ótimo (x) Bom () Regular () Fraco

Resultados em relação às expectativas iniciais: () Acima (x) Dentro () Aquém () Muito aquém

Previsão de conclusão no prazo: (x) Sim () Não

Justifique em caso negativo:

Comentários: o bolsista está bastante engajado no projeto, que tem motivação e aplicação reais. Em função do processamento dos textos das atas, um bot será treinado para informar o eleitor sobre os acontecimentos na Assembleia da República Portuguesa. Vale ressaltar que os modelos desenvolvidos podem ser aplicados em quaisquer atas de quaisquer reuniões que tenham debates com um moderador. O aluno tem ido muito bem e deve terminar seu projeto no prazo inicialmente estipulado.

Protocolo

Data: 25/06/2021

Nome Completo do Bolsista: Enzo Bustos Da Silva