**Apresentação SIICUSP - 2021**

Meu nome é Enzo Bustos da Silva e vou apresentar para vocês o Democracia Aumentada: Um Sistema para Democracia Digital, meu trabalho de Iniciação Científica, sobre os dados do Diário da Assembleia da República Portuguesa.

Este trabalho foi orientado pela professora Anna Helena Reali Costa com colaboração do Thomas Palmeira Ferraz e do André Seidel Oliveira

Sobre a relevância temos que,

Atualmente, a internet alterou o conceito de espaço de debate, inserindo as redes sociais como meios válidos para a organização e engajamento político das pessoas.

Apesar dessa mudança, os órgãos públicos continuam conduzindo a política basicamente da mesma forma: através de longas reuniões e produzindo documentos que são extensos e de difícil interpretação.

E este descompasso que existe entre a população e o governo é muito prejudicial, pois contribui para a redução da confiança nas instituições democráticas.

A Democracia Digital surge, então, como uma forma dos indivíduos utilizarem as novas tecnologias de comunicação no âmbito político. E isso é feito através de três eixos principais: a informação, a discussão e a participação.

O trabalho de IC tomou como foco o eixo da informação, já que este é a base para a geração tanto de boas discussões quanto de uma participação política efetiva e, atualmente, conta com dois sérios agravantes, que são: o excesso de conteúdo e as fake news. Desse modo a proposta do Sistema de Democracia Aumentada visa utilizar técnicas de Inteligência Artificial e Processamento de Linguagem Natural para tornar os dados produzidos no escopo legislativo, mais acessíveis, concisos e fáceis de serem interpretados.

Sobre metodologia, temos que:

Para criação desse sistema, primeiramente foi necessário utilizar uma ferramenta de web scraping, para criação de um robô que automaticamente entra no website do Parlamento Português e faz o download das Atas.

Estes dados crus, então, foram estruturados, de modo a criar um banco de dados que é dividido pelas falas dos deputados, e também passaram por uma pré-segmentação que identifica qual iniciativa legislativa, está sendo discutida em um dado ponto

Temos então o esquema do sistema completo em que o usuário insere informações pertinentes, como período de tempo e assunto de interesse, e pode utilizar 3 funcionalidades especializadas, sumarização automática, modelagem de tópicos e análise de sentimentos.

Cada uma destas funcionalidades utiliza diferentes tipos de arquiteturas:

Para a sumarização, os textos passam primeiramente por uma tradução automática para inglês, utilizando o modelo M2M100 do Facebook, já que não existem bons modelos diretos em português. Após isso é utilizado o Pegasus-Large do Google para a etapa de sumarização abstrativa, isto pq é feito um parafraseamento do texto original para reduzi-lo. Finalmente o resumo em inglês é passado para o português, novamente com o M2M100.

Para a modelagem de tópicos, o conjunto de textos é segmentado em grupos de palavras de um mesmo assunto utilizando o BERTopic, que é um modelo derivado do BERT do Google a partir de um fine-tuning para a tarefa de extração de tópicos. Estes conjuntos de palavras passam por um modelo de classificação Zero-Shot do Facebook para serem afunilados em tópicos mais palpáveis, como “Economia”. “Cultura” ou “Trabalho”.

A Análise de Sentimentos é uma funcionalidade que ainda está em construção, mas planejamos utilizar um modelo de SVM para a classificação da polaridade do sentimento expresso pela fala, considerando os aliados e oposição do governo atual.

Falando de alguns resultados:

A sumarização automática é uma funcionalidade que recebe um conjunto de documentos e cria um resumo a partir deles, esta funcionalidade é usada para a sumarização de discussões inteiras dentro de um mesmo tópico ou ainda grandes falas de contextualização sobre uma nova pauta. Na imagem de exemplo o modelo Pegasus resumiu uma grande fala de quase 8000 caracteres em apenas um parágrafo sem grandes perdas no sentido.

A modelagem de tópicos é uma funcionalidade que determina automaticamente conjuntos de palavras que pertencem a um mesmo assunto, desse modo o usuário pode orientar sua busca para tópicos que são de seu interesse, por exemplo, Segurança ou Turismo e receber de volta quais discussões tratam destes temas. No exemplo podemos ver a saída do modelo BERTopic, em que conseguimos identificar claramente tópicos de Saúde, Economia e Meio Ambiente.

A análise de sentimento é uma funcionalidade que será usada para a construção de uma métrica de confiabilidade para partidos e deputados, considerando se o discurso proferido condiz, de fato, com as ações e votações em um dado item de pauta.

Por fim, podemos concluir que

a associação das técnicas de aprendizado de máquina para o âmbito das relações políticas apresenta um grande potencial para ser explorado, com aplicações que podem melhorar a qualidade final da informação recebida pelo usuário, tornando-a mais concisa e relevante, para que os cidadãos possam passar menos tempo filtrando as informações recebidas e mais tempo na atuação política.

Como trabalho futuro, pretendemos implementar uma interface amigável para o Democracia Aumentada que integre todas as funcionalidades em uma única aplicação, de modo que este trabalho sirva também como uma base para projetos que trabalhem os outros eixos, de discussão e participação política.

Finalmente, gostaria de passar aos agradecimentos

Primeiramente a professora Anna Reali, por ser minha orientadora neste projeto de IC

Bem como ao Thomas Palmeira e ao André Seidel por serem colaboradores

Além disso gostaria de agradecer especialmente ao Programa de Bolsas Itaú em conjunto com o C2D por terem financiado este trabalho de pesquisa.