**Resumo Apresentações 29/08/19**

Por fim, durante o último dia de apresentações dos projetos de Processamento de Linguagem Natural, um dos projetos apresentados considerou o artigo “NLP to the Rescue?” cujo objetivo é utilizar informações publicadas durante emergência para aplicar técnicas de Processamento de Linguagem Natural para extrair informações que aumentem a “Situational Awareness”.

Dessa forma, o grupo buscou dados relacionando alguns acontecimentos no Brasil para aplicar a ideia proposta pelo artigo. Assim, dois eventos foram selecionados: queda do avião Lamia 2933 com os integrantes da equipa de futebol Chapecoense em 29 de Novembro de 2016 e o rompimento da barragem de brumadinho em 25 de janeiro de 2019.

O grupo utilizou APIs do Twitter para coleta dos dados. Durante a classificação o grupo utilizou Unigrams e Bigrams após a aplicação de RegEx e Word Tokenize, além da remoção de StopWords. O algoritmo utilizado foi o Naive Bayes e o grupo alcançou uma acurácia próxima de 66%.

Na sequência, o projeto “Twitter as a Corpus for Sentiment Analysis and Opinion Mining” utilizou o artigo de mesmo nome publicado por Alexander Pak e Patrick Paroubek em 2010 na International Conference on Language Resources and Evalulation, Valleta, Malta. O artigo contém incríveis 2710 citações. Para a implementação, o grupo utilizou o dataset Sentiment140 disponibilizado publicamente pela Universidade Stanford com 1,6 milhões de Tweets rotulados em 0 a 4.

Na preparação, o grupo filtrou os dados e aplicou RegEx para remover nomes de usuários (palavras começadas por @), remoção de símbolos reservados do Twitter (como “RT” e “#”), além da remoção de URLs. Em seguida, o grupo utilizou a tokenização, remoção de sopwords, criação de n-gramas e POS-Tagging. No modelo, o grupo utilizou o algoritmo Naive Bayes e obteve uma acurácia de 72% com unigramas.