

Aplicação de NLP para análise de Tweets com menção à CEMIG, com foco no atendimento aos clientes

Trabalho final para a disciplina de Processamento de Linguagem Natural do curso de especialização em Ciência de Dados e Big Data da PUC-MG - Setembro de 2019.

Aluno: Carlos Alberto Rocha Cardoso
Professora: Keyla Macharet

Objetivo

O presente trabalho teve como objetivo aplicar e consolidar os conceitos e técnicas vistos na disciplina de processamento de linguagem natural. Para isso foi definido um caso de uso que consiste na coleta, processamento e análise de tweets com menção à empresa CEMIG - Companhia Energética de Minas Gerais, com foco na coleta de informações que auxiliem na identificação de reclamações de clientes em casos como falta de energia. A partir dessa identificação seria possível direcionar ações de atendimento junto a esses clientes, de forma mais ágil e direta.

Coleta dos dados

Para a coleta dos tweets foi utilizada a biblioteca tweepy. O tweepy utiliza a API do twitter para coleta de informações, para isso é necessário ter uma conta registrada como developer na plataforma do twitter e gerar as chaves para autenticação e uso da API. Após coleta os dados foram salvos em arquivo .CSV para análise.

Preparação dos dados

Nessa etapa foi construído um dicionário de palavras e polaridade utilizando um léxico de sentimentos para o português, chamado SentiLex-PT. Com ele foi possível verificar a polaridade dos tokens coletados nos tweets e analisar sentimento. Além disso, utilizando a biblioteca spaCy, foi feita a identificação de tags (verbos, substantivos, adjetivos) e entidades (nomes, localidades).

Análise dos dados

Nessa etapa foram realizadas análises relacionadas a quantidade de tweets por data de publicação; frequência e nuvem de palavras; frequência e nuvem de substantivos, verbos e adjetivos conforme polaridade; frequência e nuvem e entidades conforme polaridade; identificação de tweets contendo reclamações de falta de energia incluindo localização.

Conclusão

Esse trabalho permitiu, em um caso de uso real, a aplicação e consolidação dos diversos conceitos e técnicas vistos na disciplina de processamento de linguagem natural. Utilizando a biblioteca de nlp spaCy e um dicionário de polaridade de palavras baseado no SentiLex-PT02, foi possível realizar diversas análises a partir dos tweets coletados com menção à CEMIG.

Ao final cumpriu-se o objetivo de identificar de forma bastante precisa tweets relacionados à reclamação de falta de energia, a partir de termos coletados das análises das classes de palavras: substantivos, adjetivos, verbos, etc. É possível que a partir de aperfeiçoamentos, método semelhante possa ser aplicado como ferramenta no processo de atendimento ao cliente em companhias de energia elétrica, tornando a interação mais direta e ágil, em muitos casos automatizada. Um exemplo é que no caso de uma falta de energia por manutenção programada, o cliente poderia receber um tweet automático do canal oficial companhia, informando sobre o horário e detalhes da manutenção.