Analysis of a brazilian electricity distributor based on population's tweets

Isabella Calfa Vieira Costa[1]
[1] Universidade Federal da Bahia

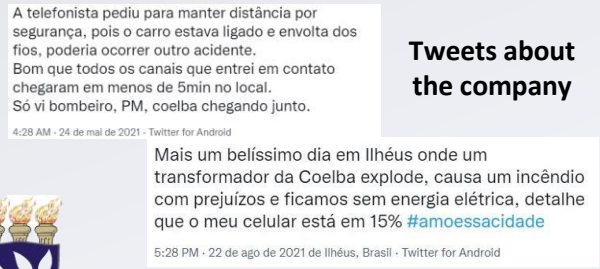
Abbreviated abstract: Electricity supply is an essential service, so it's expected that changes in amounts paid for the service and interruption in supply are commented on social media. This document analyzes tweets about a Brazilian distribution company of electric energy (extracted from Twitter's API) to understand which issues are most commented, perceptions about the company and understand the relationships with other company indicators, as well as, with moments, places and times of the posts. This kind of study is important to the company to work to improve the image and service provided. To understand clients perceptions were used machine learning to sentiment analysis and line and bar graphs, and word clouds to visualize the results.

4th Conference on
Statistics and
Data Science
Salvador, Brazil (online)
December 1-3, 2022

Problem and Data

Problem: Recognizing the importance of the electric power supply to the population and the hyper-connected world through social networks, we seek to understand **what is the perception that the customers of a certain electric power distributor in Brazil** have about the service provided. Also, what is its **image before society** and the **behavior of the messages about the company** on social networks.

Tweets about the company



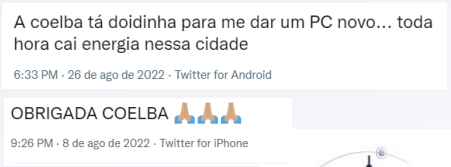
Solution: Analyze the tweets about the company between **2021 and 2022** to perform sentiment analysis (using ML) and find out what **times, places and people were commenting about the company** and, most importantly, **what are they commenting**.

Used Dataset: Extracted by **Twitter's API**.

Dataset:

- 13,559 tweets
- **Start**: Jan 1st, 2021
- **End**: Aug 26, 2022
- **Items**: id, text, date, user_id, georeference and language.

Examples of tweets in dataset:



Methods

Dataset Adjust:

Created at ➔ Converted to **GMT-3**

New Columns ➔ "**word_count**" and "**lenght**"

Text_process ➔ Removed accents, links, special characters, stopwords and unfrequent words; converted upper to lower case.

Comparison Indicators:

- **Source:** ANEEL
- **Indicators:** DEC (SAIDI), FEC (SAIFI), TMA and volume of complaints.

Sentiment Analysis:

- Topics:
  - ↑ **Positive**: Compliments and suggestions
  - ↓ **Negative**: Critics and complaints
  - ≈ **Neutral**: Everything else
- Steps:
  1. Manual classification
  2. Division in unigrams, bigrams and trigrams
  3. *Training Models*: Logistic Regression, SVM, Random Florest and Naive Bayes
  4. Evaluation metrics: F1-Score and accuracy

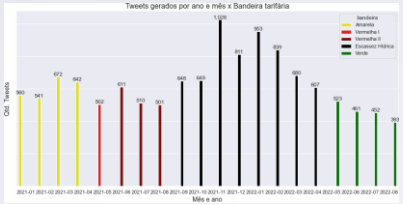
Manual Classification as Positive	A telefonista pediu para manter distância por segurança, pois o carro estava ligado e envolta dos fios, poderia ocorrer outro acidente. Bom que todos os canais que entrei em contato chegaram em menos de 5min no local. Só vi bombeiro, PM, coelba chegando junto. 4:28 AM · 24 de mai de 2021 · Twitter for Android

4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022

Results and Conclusions

- **Users**: No robots; most users tweeted once; user with more comments from Ilhéus and complains about the interruption of supply.
- **Date and time**: Most tweets at week days and between 10 AM - 2 PM and 6 PM - 7 PM; at night, many comments about interruption of supply; comments about bill close to payment day.
- **Location**: Salvador, Feira de Santana and Ilhéus.
- **Most frequent words**: energia, nao, luz, bahia, conta.
- **Sentiment Analysis**:
  - ***Final Model***: Naïve Bayes + Logistic Regression + SVM (soft)
  - ***Accuracy***: 87.6%
  - ***F1-Score***: 85.5%

- ***More tweets*** when bills are more ***expensive***;
- More **complaints in ANEEL** when **volume tweets** increased;
- "CPI" really commented and associated to bad service quality;
- Sentiment analysis results:
  - **70% Negative**
  - **25% Neutral**
  - **5% Positive**
- Negative topics: araras azuis-de-lear, expensive bill, bad service.
- Consumers quickly express their discontent when they think that the service is not being well provided;
- At weekends, 76% of tweets are negative; at weekdays is equal to 66%.



Volume of tweets x Taxes

Tweet correctly classificated

Wordcloud negative tweets

4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022