When Tweets Get Viral - STIL 2023

A Deep Learning Approach for the Stance Analysis of Covid-19 Vaccines Tweets of Brazilian Political Elites

Authors: Lorena Barberia, Pedro Schmalz e Norton Trevisan Roman

DCP-USP, EACH-USP

Contents of the Presentation

Universidade de São Paulo

Political Background (2020)

- In 2020, The Federal Government opposed lockdowns and restrictions during the pandemic and constantly attacked Covid-19 vaccines;
- Congress, Supreme Court and local elites in the opposition contested its policies regarding lockdown and Covid-19 Vaccination.
- In turn, there was wide politicization around the Lockdown and Vaccination subjects.

Universidade de São Paulo

Political Events and Discourse Change (2021)

- April 2021 - Opposition calls a congressional hearing (*CPI da Covid*)
- June 2021 - Covaxin Scandal;
- These events led to important discourse changes in the government over the subject of Covid-19 vaccines.

Universidade de São Paulo

Motivation

- Brazilian Congress operated remotely during most of 2020 and 2021;
- Increased usage of social media by political candidates and politicians;
- We aimed to understand how politicians positioned themselves in what concerns Covid-19 vaccines, on social media.

Universidade de São Paulo

Research Questions

- How did Brazilian Mayoral Candidates drive the COVID-19 vaccine agenda on X (former Twitter)?
- Did the Data Drift phenomenon affect the model's results, given the changes in the politicians positions and discourses between 2020 and 2021?

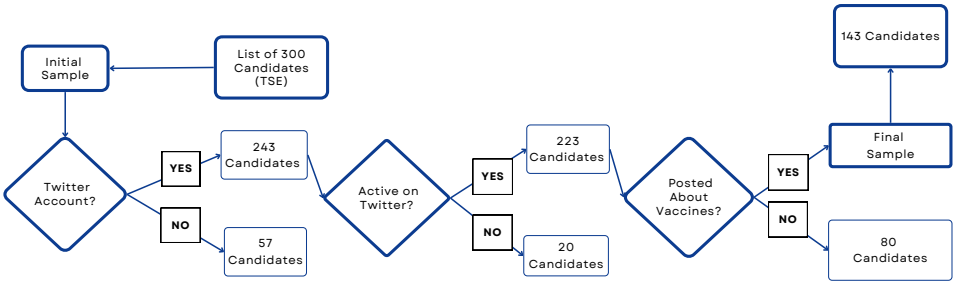
Universidade de São Paulo

Research Design

- Barberia, Schmalz et al. (2022) analyzed Covid-19 Vaccines and Vaccination tweets (now posts) from mayoral candidates (local elites) posted from 2020 to 2021 on X (former Twitter).
- This training dataset is now used to train BERTimbau with three different datasets to predict Covid-19 Stances: one for 2020, one for 2021 and one with posts from both years.

Universidade de São Paulo

Building the Corpus



Universidade de São Paulo

Building the Corpus - Keywords and Annotation

- We have chosen a keyword search as our primary corpus selection strategy.
- A training data set was created by manually annotating vaccine-related tweets.
- Annotation - **Relation to Covid-19 Vaccines** and **Stance** towards them.

Universidade de São Paulo

Corpus Expansion and Model

- To expand the time period of our analysis, we added tweets from the same individuals for 2021.
- To classify tweets, we used the Brazilian Portuguese pre-trained model, *BERTimbau*[1], which is based on BERT[2].

[1] Souza, Nogueira, and Lotufo 2020.
[2] Devlin et al. 2019.

Universidade de São Paulo

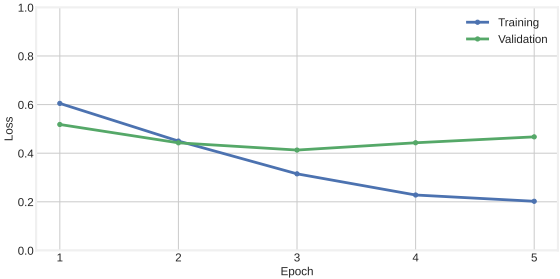
Corpus Distribution

Table: Distribution of Classes

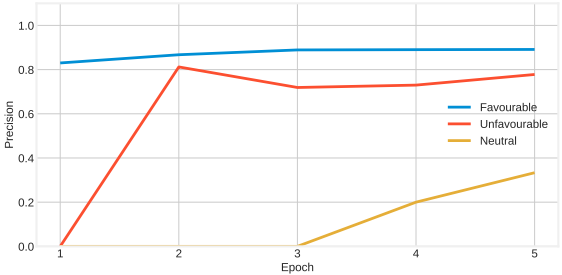
Class	2020	2021	Combined
Favorable	1,319	2,682	4,001
Neutral	82	2,085	2,167
Unfavorable	188	64	252
Total	1,589	4,831	6,420

Universidade de São Paulo

Validation Results (2020)



(a) Avg. Training and Validation Loss

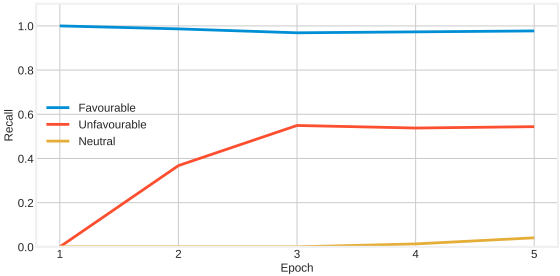


(b) Avg. Precision per Class

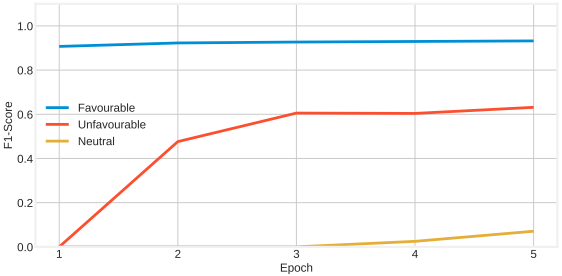
Figure: Validation Metrics (2020)

Universidade de São Paulo

Validation Results (2020)



(a) Avg. Recall per Class

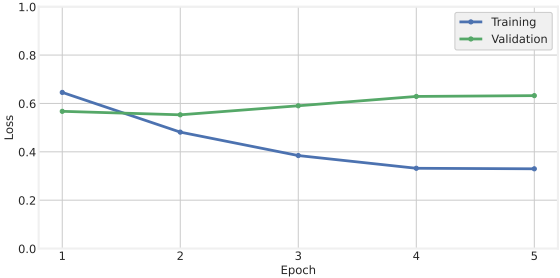


(b) Avg. F1-Score per Class

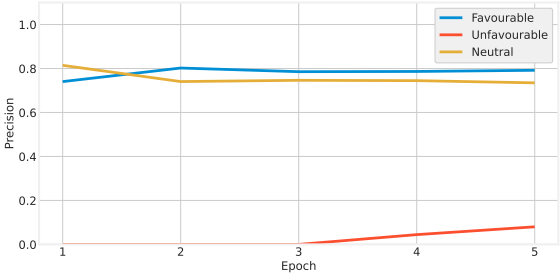
Figure: Validation Metrics (2020)

Universidade de São Paulo

Validation Results (2021)



(a) Training and Validation Loss

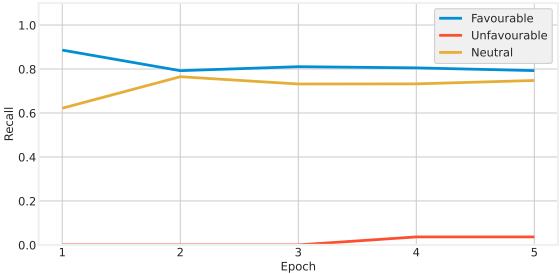


(b) Precision per Class

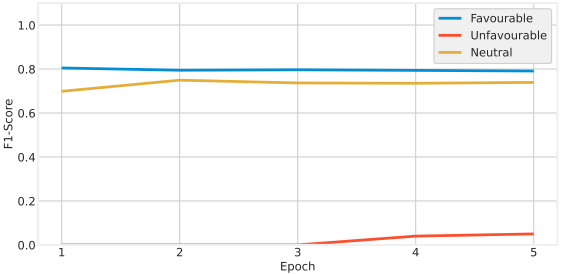
Figure: Validation Metrics (2021)

Universidade de São Paulo

Validation Results (2021)



(a) Recall per Class

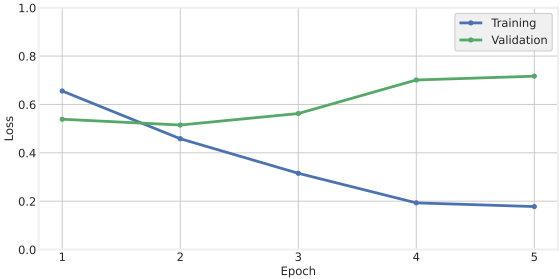


(b) F1-Score per Class

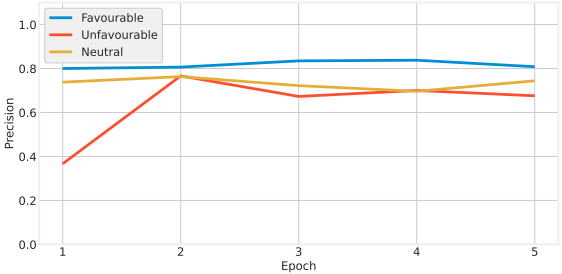
Figure: Validation Metrics (2021)

Universidade de São Paulo

Validation Results (2020 and 2021)



(a) Training and Validation Loss

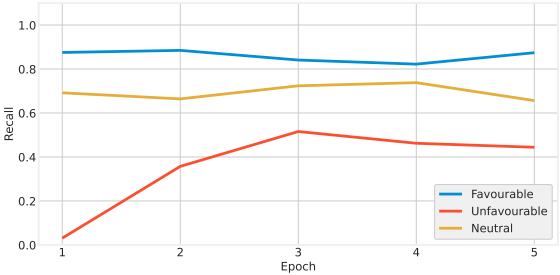


(b) Precision per Class

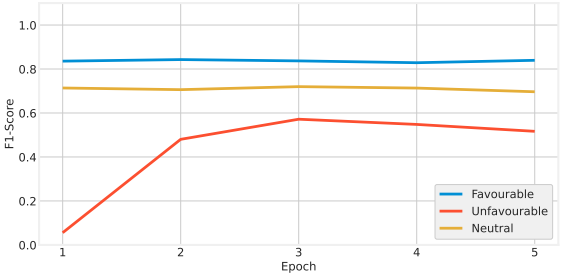
Figure: Validation Metrics (2020 and 2021)

Universidade de São Paulo

Validation Results (2020 and 2021)



(a) Recall per Class



(b) F1-Score per Class

Figure: Validation Metrics (2020 and 2021)

Universidade de São Paulo

Epoch 2 Results*

Table: Average of Folds' Validation Results (Epoch 2)

Data Set	Train Loss	Val. Loss	Val. Accuracy	F1-Micro	F1-Macro
2020	0.45 (0.43:0.46)	0.44 (0.40:0.49)	0.86 (0.84:0.88)	0.86 (0.84:0.88)	0.51 (0.44:0.57)
2021	0.48 (0.48:0.49)	0.55 (0.53:0.58)	0.77 (0.76:0.78)	0.77 (0.76:0.78)	0.67 (0.66:0.69)
Both	0.46 (0.45:0.47)	0.52 (0.50:0.53)	0.79 (0.78:0.80)	0.79 (0.78:0.80)	0.67 (0.65:0.69)

*Notes: 95% Confidence Intervals in Brackets.

Universidade de São Paulo

Conclusions

- Important changes between 2020 and 2021 made the classification task more difficult;
- The imbalance in the data set affected the results on the unfavorable class for 2021 and in the neutral class for 2020;
- Using both years, the model performs relatively well. However, there will be an underestimation of unfavorable tweets.

Universidade de São Paulo

Next Steps

- Compare the results obtained by the BERTimbau model with other models (e.g. SVMs, RNNs, CNNs) or other different portuguese pre-trained BERTs - e.g. BERTabaporu, Albertina (PT-Br), etc.;
- Expand the annotation to include sentiment analysis;
- Minimize the impact of class imbalance on the classification task.
- Use the model to classify tweets from other political elites (e.g. Congresspersons).

Universidade de São Paulo

Important Links

- Tweet Corpus - `https://github.com/PedroSchmalz/covid19-tweets-brazilian-mayoral-candidates.git`
- Replication Files - `https://github.com/PedroSchmalz/when-tweets-get-viral-replication-files.git`
- Contact - `https://www.linkedin.com/in/pedro-schmalz/`

Universidade de São Paulo

Thank you for you attention!

Any questions?

pedrosantanaschmalz@usp.br

Universidade de São Paulo