Lecture 30b:
Problem Set 5 Presentations

Big Data and Machine Learning for Applied Economics
Econ 4676

Ignacio Sarmiento-Barbieri

Universidad de los Andes

November 26, 2021

“A rose by any other name would smell as sweet” Juliet Capulet

- There is an adage that says, *“choose your words carefully.”* Words themselves may reveal far more than what we’re trying to say. There’s mounting evidence that our written words show who we are.
- The objective today is to predict to whom each tweet belongs. The training dataset contains around 7,000 tweets of four prominent Colombian politicians’ accounts: Claudia Lopez, Gustavo Petro, Alvaro Uribe, y Alejandro Gaviria. The test dataset contains 500 unlabeled tweets. We want to predict which account posted the tweets in the test set.

Results

Table 1

	Team	Rigth	Wrong
1	Prieto, Segura, Navarro	1,622	378
2	Acero, Pacheco, Saenz	1,617	383
3	Cortes, Rojas, Pena, Salazar	1,537	425
4	Agudelo, Cepeda, Cifuentes, Mosquera	1,442	558
5	Gonzalez, Rengifo	1,431	569
6	Rodriguez, Montero	1,251	749
7	Castro, Ramirez, Miranda	663	1,337

Results: Detail

	Team	Name	Rigth	Wrong
1	Prieto, Segura, Navarro	Gaviria	394	79
2	Acero, Pacheco, Saenz	Gaviria	401	78
3	Castro, Ramirez, Miranda	Gaviria	194	216
4	Prieto, Segura, Navarro	Lopez	443	69
5	Acero, Pacheco, Saenz	Lopez	458	96
6	Castro, Ramirez, Miranda	Lopez	145	299
7	Prieto, Segura, Navarro	Petro	357	78
8	Acero, Pacheco, Saenz	Petro	366	79
9	Castro, Ramirez, Miranda	Petro	108	237
10	Prieto, Segura, Navarro	Uribe	428	152
11	Acero, Pacheco, Saenz	Uribe	392	130
12	Castro, Ramirez, Miranda	Uribe	216	585

Next Week Presentations

- November 30th
  1. Startup Failure (Rodriguez y Montero)
  2. Covid-19 (Saenz)
  3. Precios Propiedades (Gonzalez)
- December 2nd
  1. Cambio Estructural (Rengifo)
  2. Booktopia (Agudelo, Cepeda, Cifuentes, y Mosquera)
  3. Rendimiento Educativo (Salazar, Cortes, Rojas, y Peña)
- December 3rd
  1. Pobreza Multidimensional (Miranda)
  2. Demanda de Energía (Ramírez y Castro)
  3. Basketball (Segura, Prieto y Navarro)

Final Submission

- Final submission is due on **December 10th at 6 pm**. Upload it to your Repo, the presentation should be there as well.
- The final document should not be longer than 5 (five) pages (not including the title page with abstract, and references).
- **This is worth 15% of your grade.**

Final Submission

- When writing up the document, it should contain the following:
  - Title
  - Abstract (200 words limit)
  - Introduction. It should contain at least: the problem/research question clearly defined, antecedents of your work, your value-added (i.e. why your project is interesting/novel/different), and a preview of results and takeaways.
  - Data. Treat this section as an opportunity to present a compelling narrative to justify or defend your data choices, walk the reader through your reasoning of how you think you got the right data for the task, describe it accordingly with descriptive stats, graphs, etc.
  - Model. Present the model you are using. Be sure to argue why this is the best model for your task. When writing this section think on the following questions: did you apply other models? Is this model the most accurate at predicting? Is this the only model that you can use? etc.
  - Results. Here you should present your results. I understand that a semester is a short time to have a full paper, so preliminary results are fine.
  - Conclusions and recommendations. In this section, you should state the main takeaways of your work.