

Who Tweeted That?: Mexican Political Party Tweet Classifier

Jerry Montes, Data Scientist

01

Mexican Politics

And you thought American politics were messy...

Government Type



México

Federal presidential
republic



U.S.A.

Constitutional
federal republic

Political Fun Facts



71 years straight

Number of years that **PRI** held the presidential office



12%

Difference between poll results and election results after system **crashed**



Distilled Water

Was purchased instead of chemotherapy by governor who **embezzled** state funds



5

Number of presidents whose **alleged wrongdoing** would have been investigated following 40% participation in 2021 referendum

The Biggest Worries

“What is the primary problem the country is currently facing?”

1. Crime

*75.6% of the population considers their city to be
unsafe*



¡El nueve ninguna se mueve!

2. Corruption

*60% of the population believes that the **majority** of police are **corrupt***



Octubre 2009



Febrero 2017

Case Study: Rebsamen School and the 2017 Quake

3. Economy

44% of the population lives in poverty

Minimum wage in Mexico and border states in 2019



Vox

(Political) Party Time!

We will solve X by doing Y...

6 of the 7 current Political Parties



Action for Mexico



Revolutionaries



A new dawn



Green gets it done



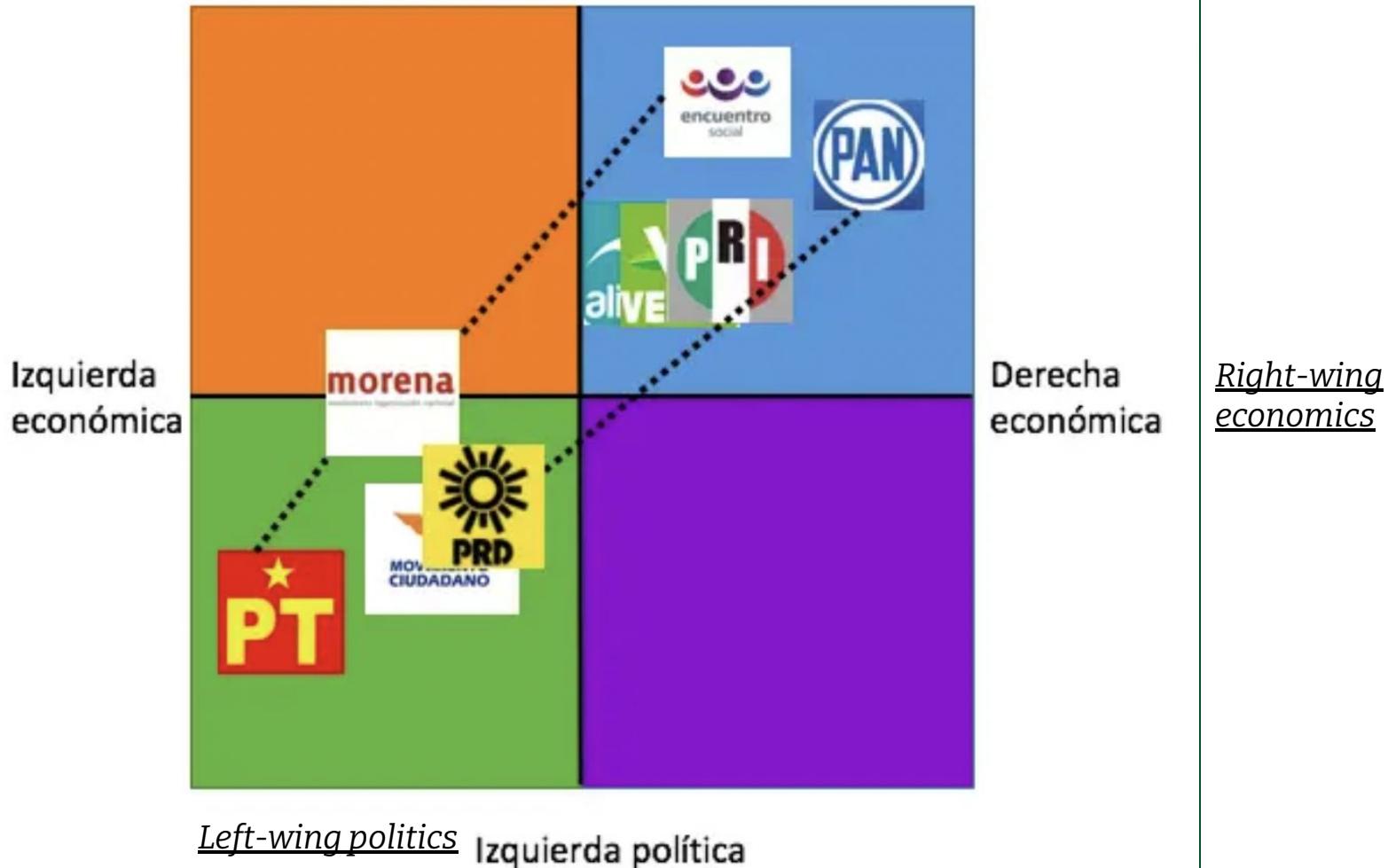
The future is orange



Morena

Mexico's hope

Derecha política Right-wing politics



02

Problem Statement

Making sense of each party's message

Who Done Said It?

Official Mexican Political Party Twitter Accounts

- @AccionNacional
- @PRI_Nacional
- @PRDMexico
- @partidoverdemex
- @MovCiudadanoMX
- @PartidoMorenaMx

I sought to build a multi-class classification model that, when presented with a tweet from one of the six official Twitter accounts associated with Mexican political parties, was capable of determining which party posted it.



Data

Twitter API v2: Elevated Access

Used to pull Twitter data

3,200 most recent tweets

Pulled from each of the 6 official Twitter accounts associated with
Mexican political parties

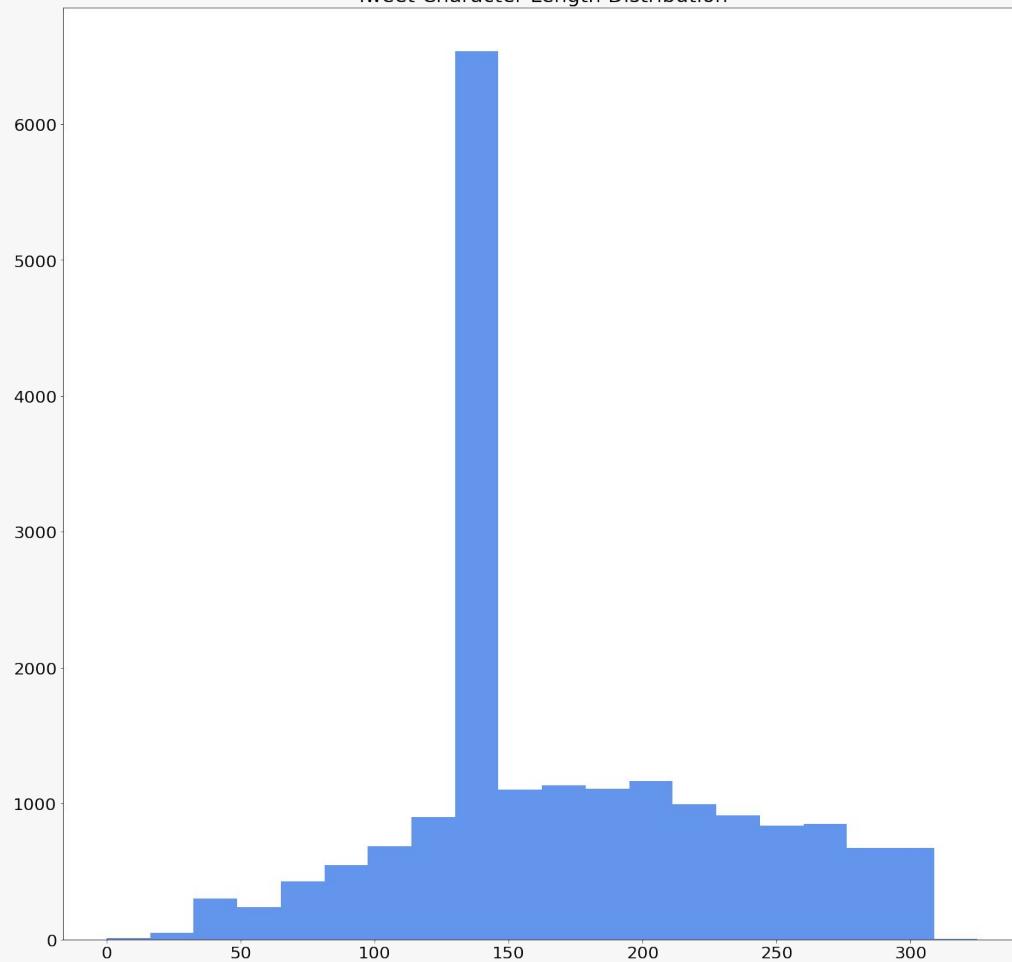
19,169 tweets

Total number of tweets obtained after removing duplicates

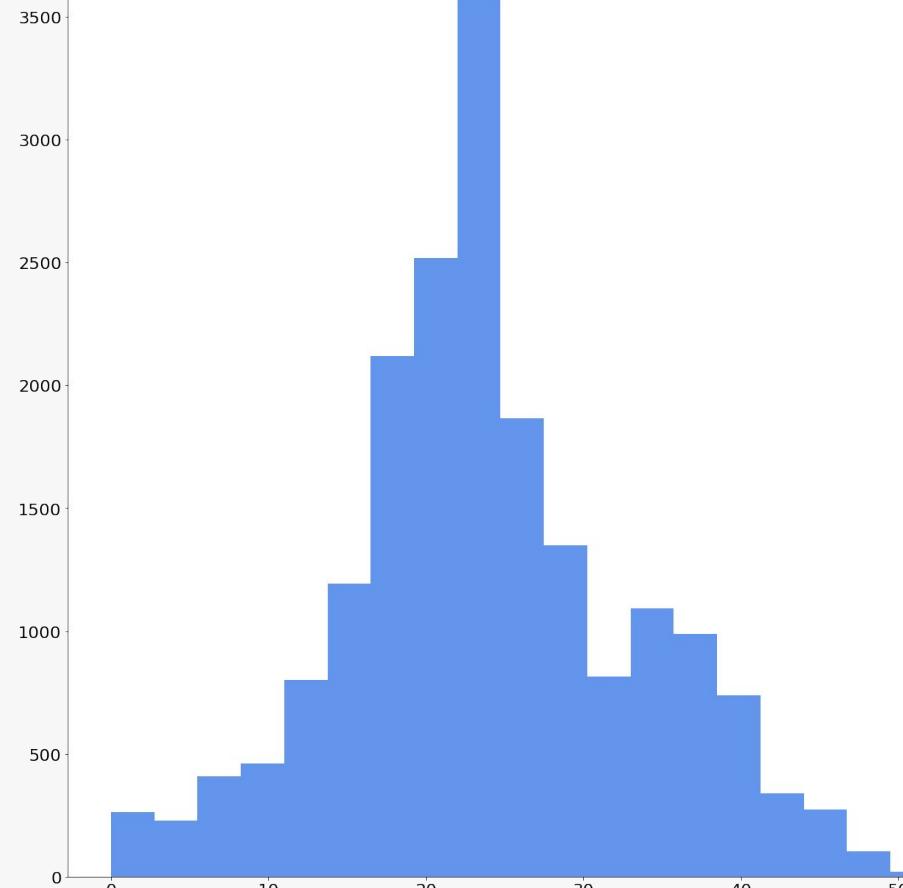
Exploratory Data Analysis



Tweet Character Length Distribution



Tweet Word Count Distribution



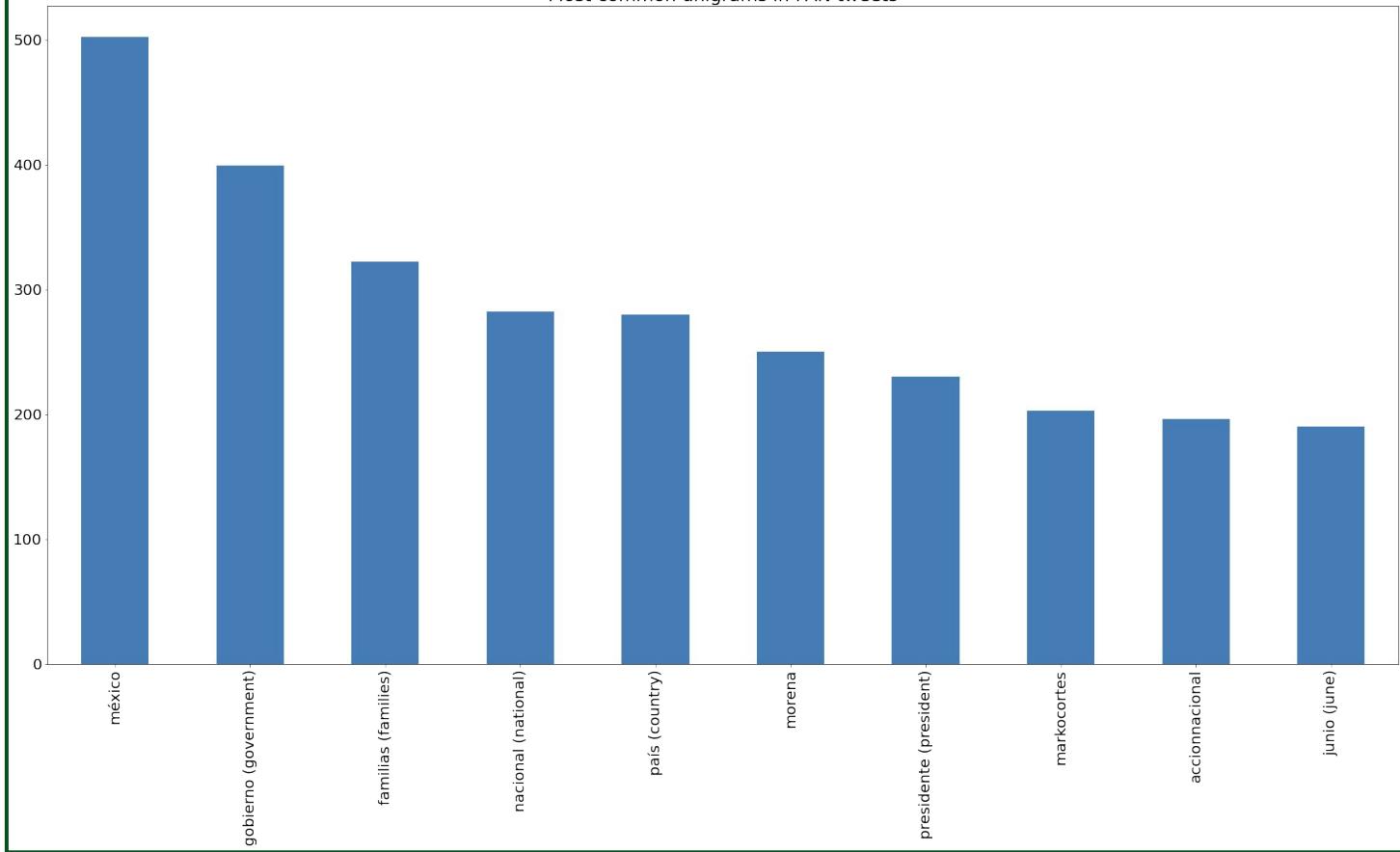
Stop Words List 1

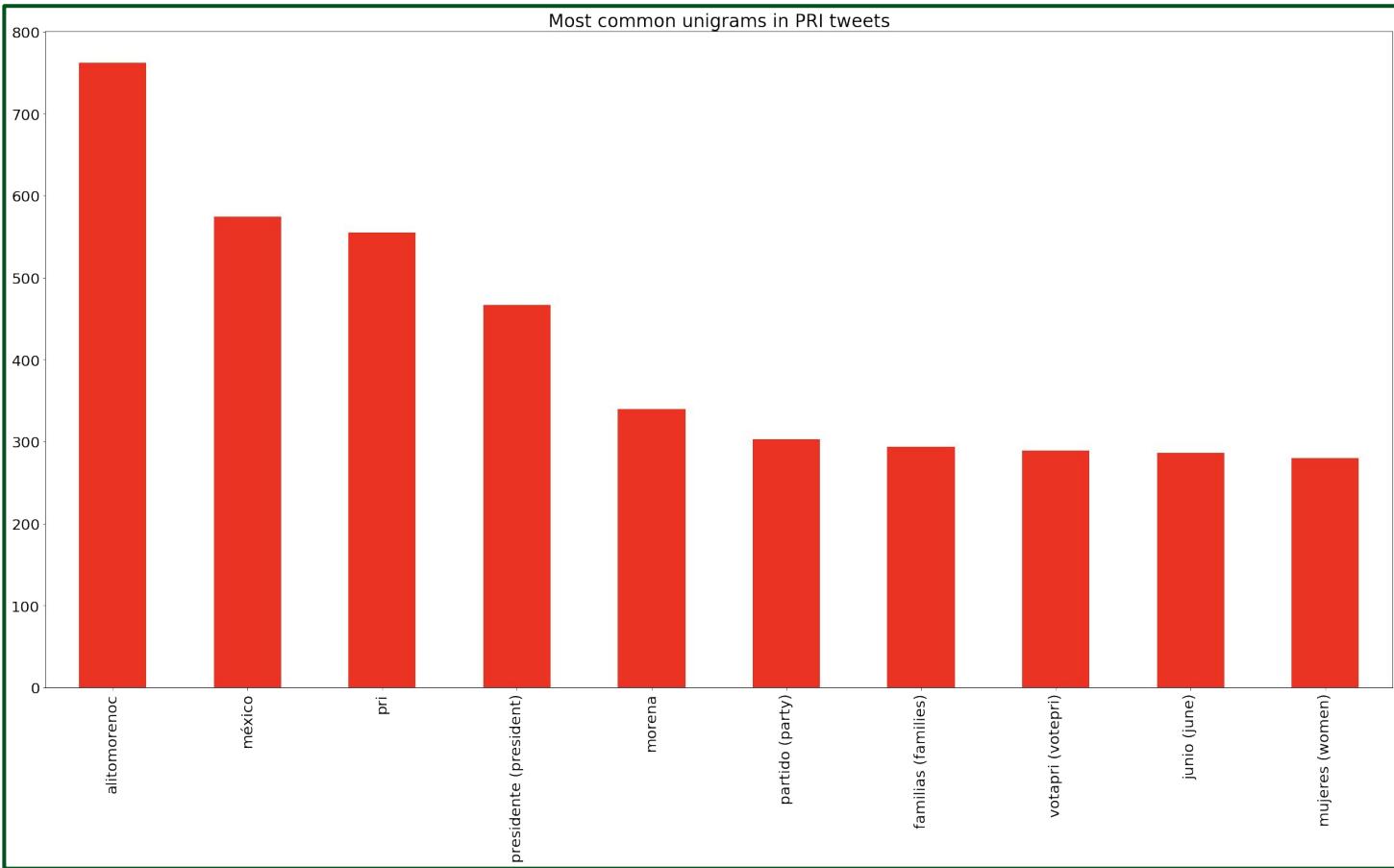
737 words

*Spanish + “rt” + “https” +
“www” + “co” + “com”*

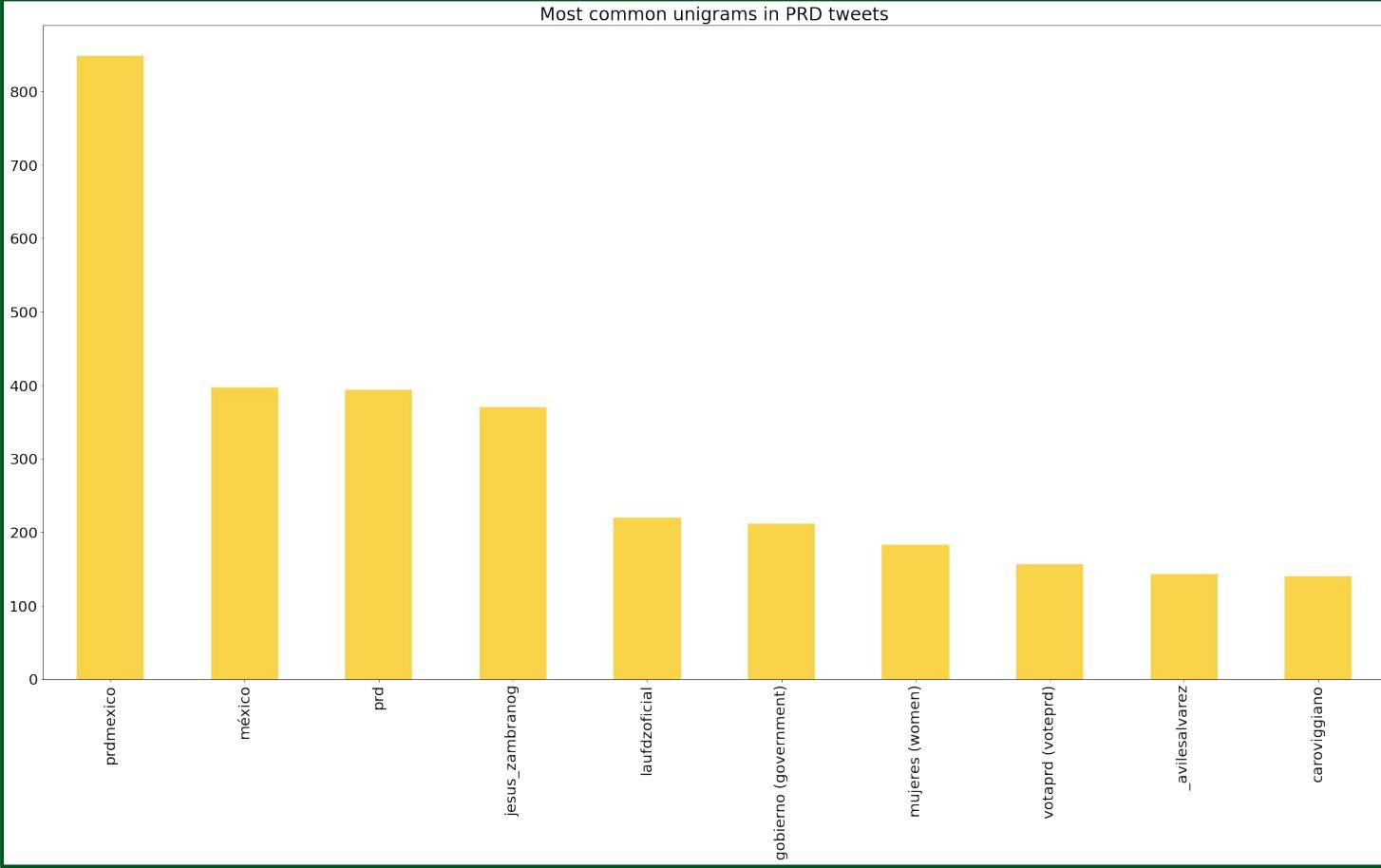


Most common unigrams in PAN tweets

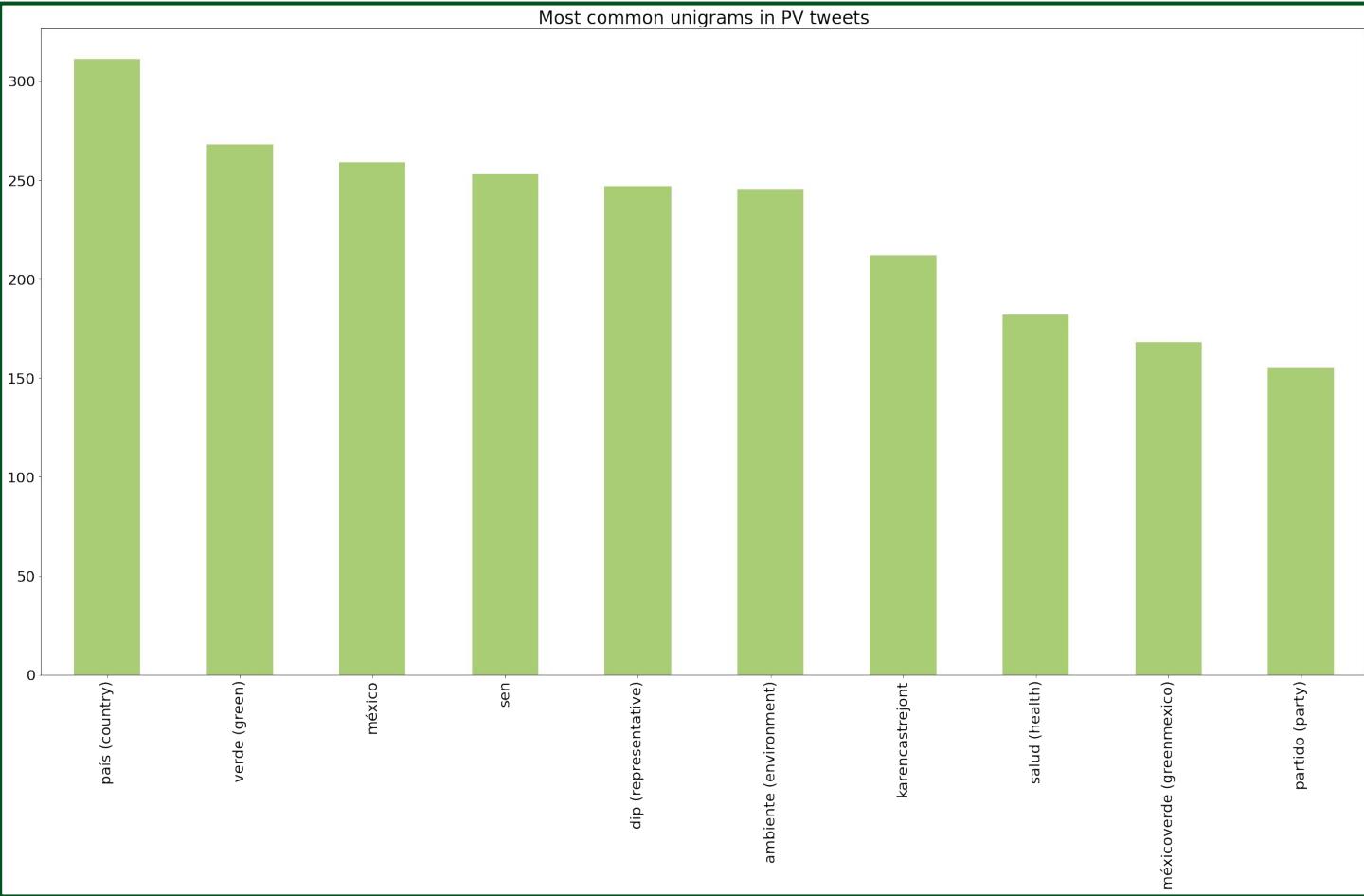




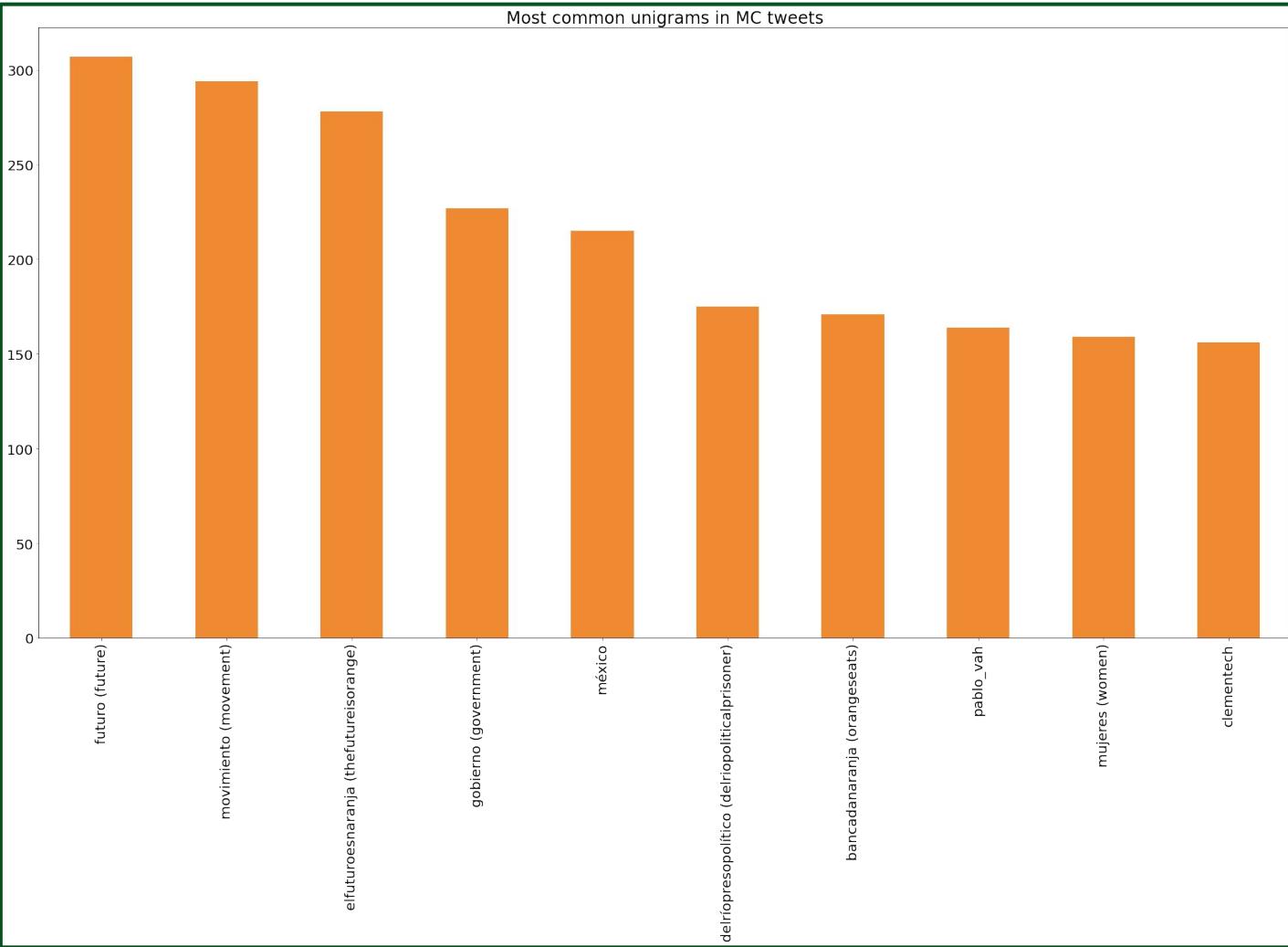
Most common unigrams in PRD tweets



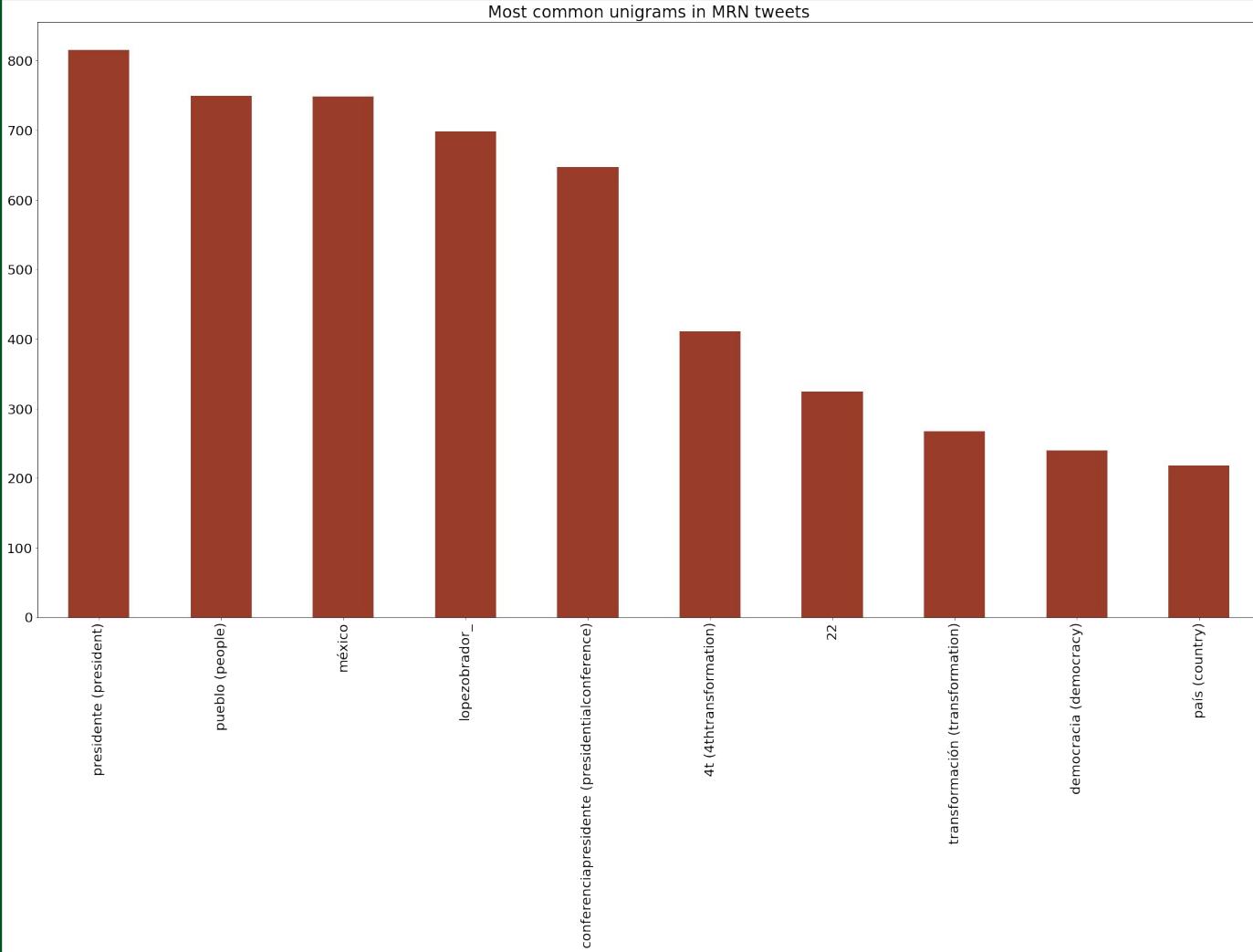
Most common unigrams in PV tweets



Most common unigrams in MC tweets



Most common unigrams in MRN tweets



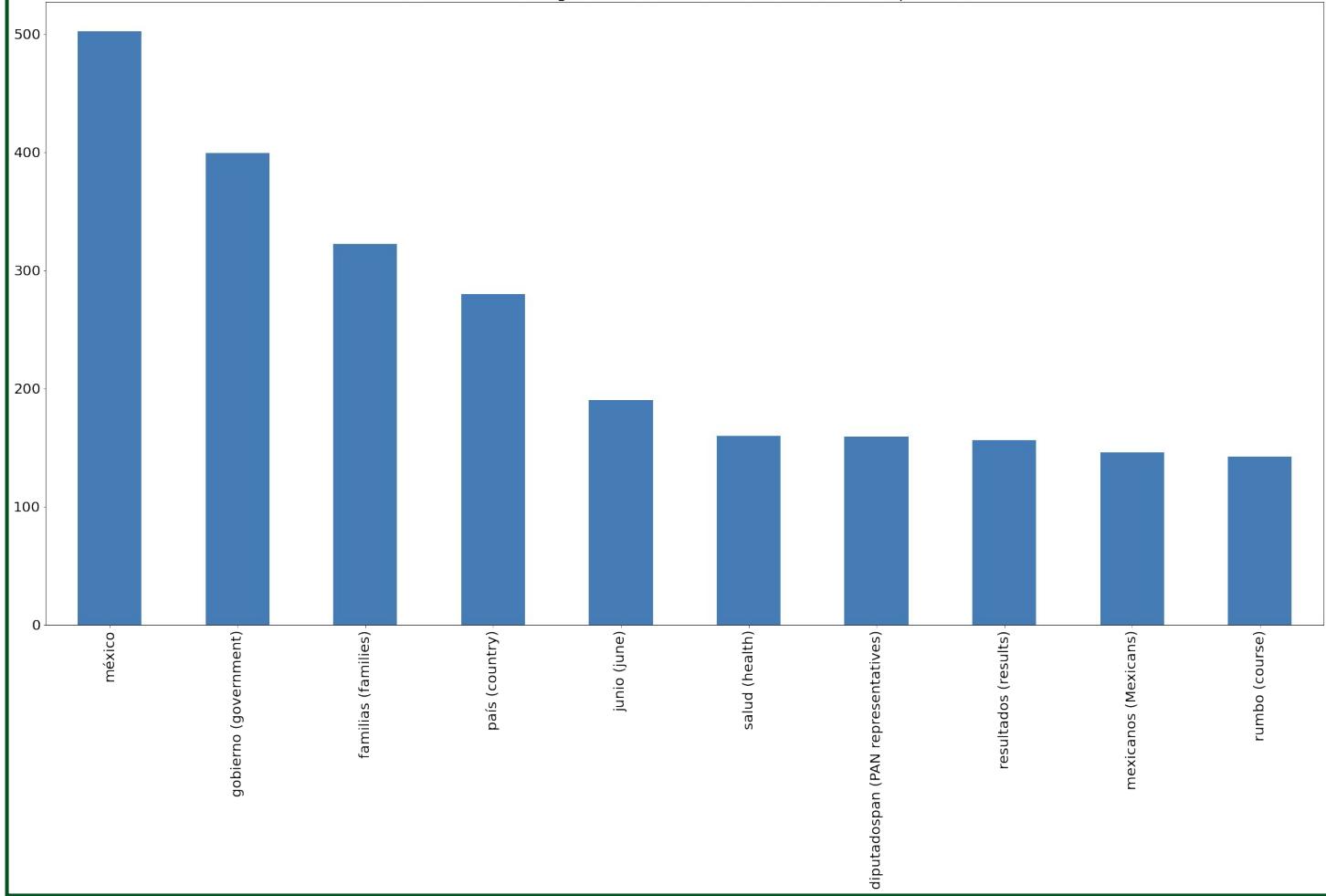
Stop Words List 2

780 words

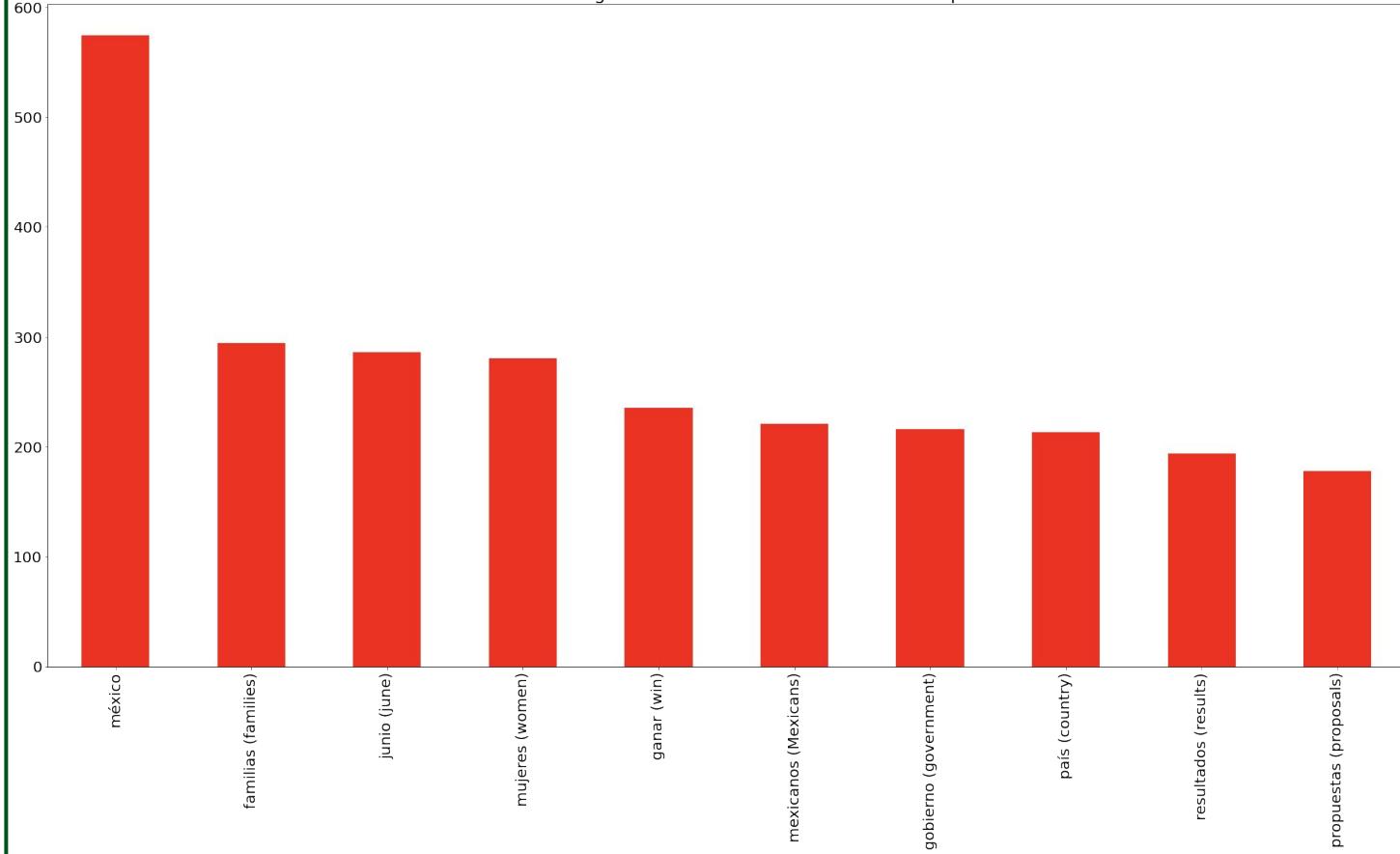
*Stop Words List 1 + party names +
party acronyms + politician
names in first top 10 +
current President's initials (amlo)*



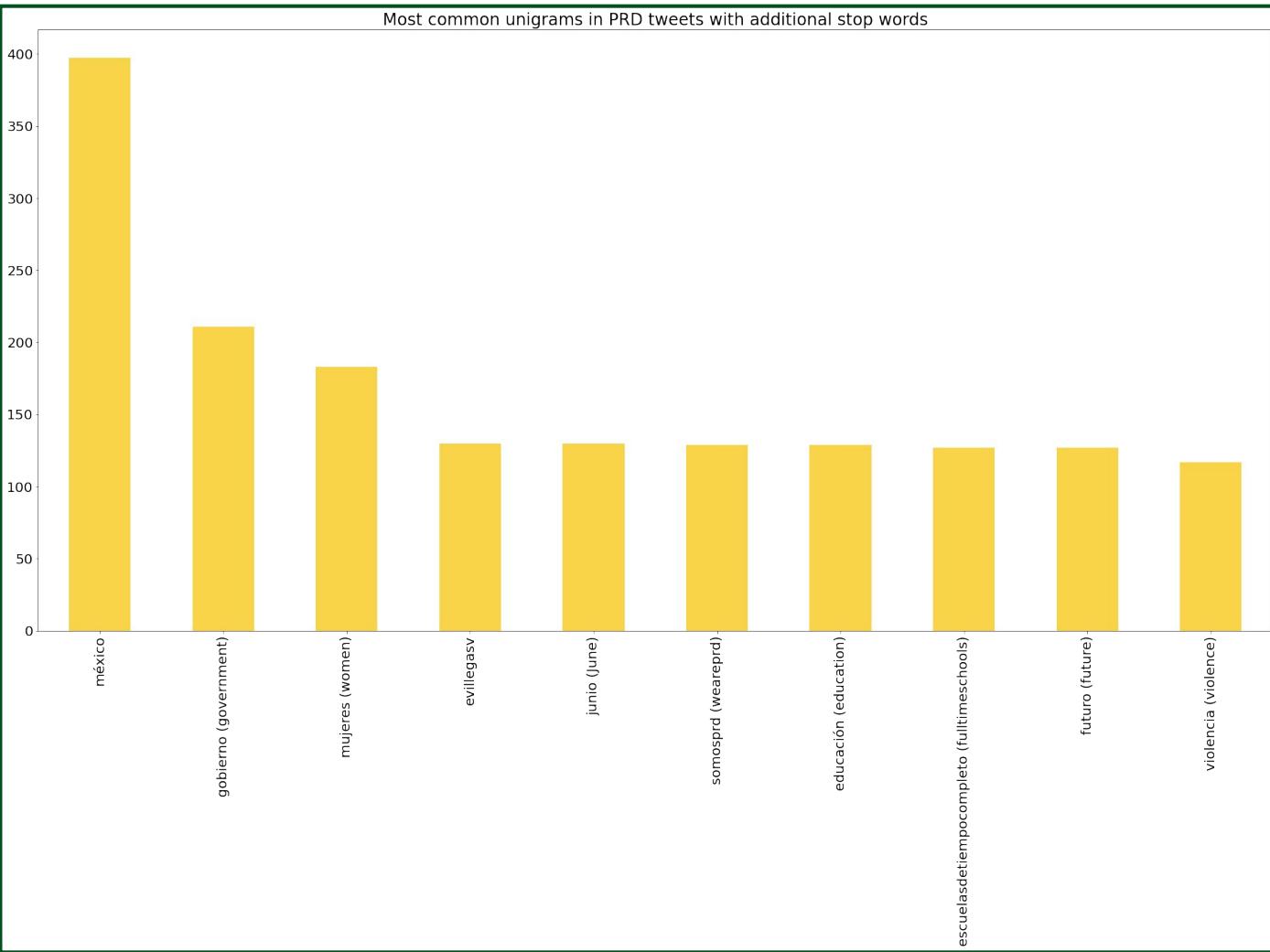
Most common unigrams in PAN tweets with additional stop words



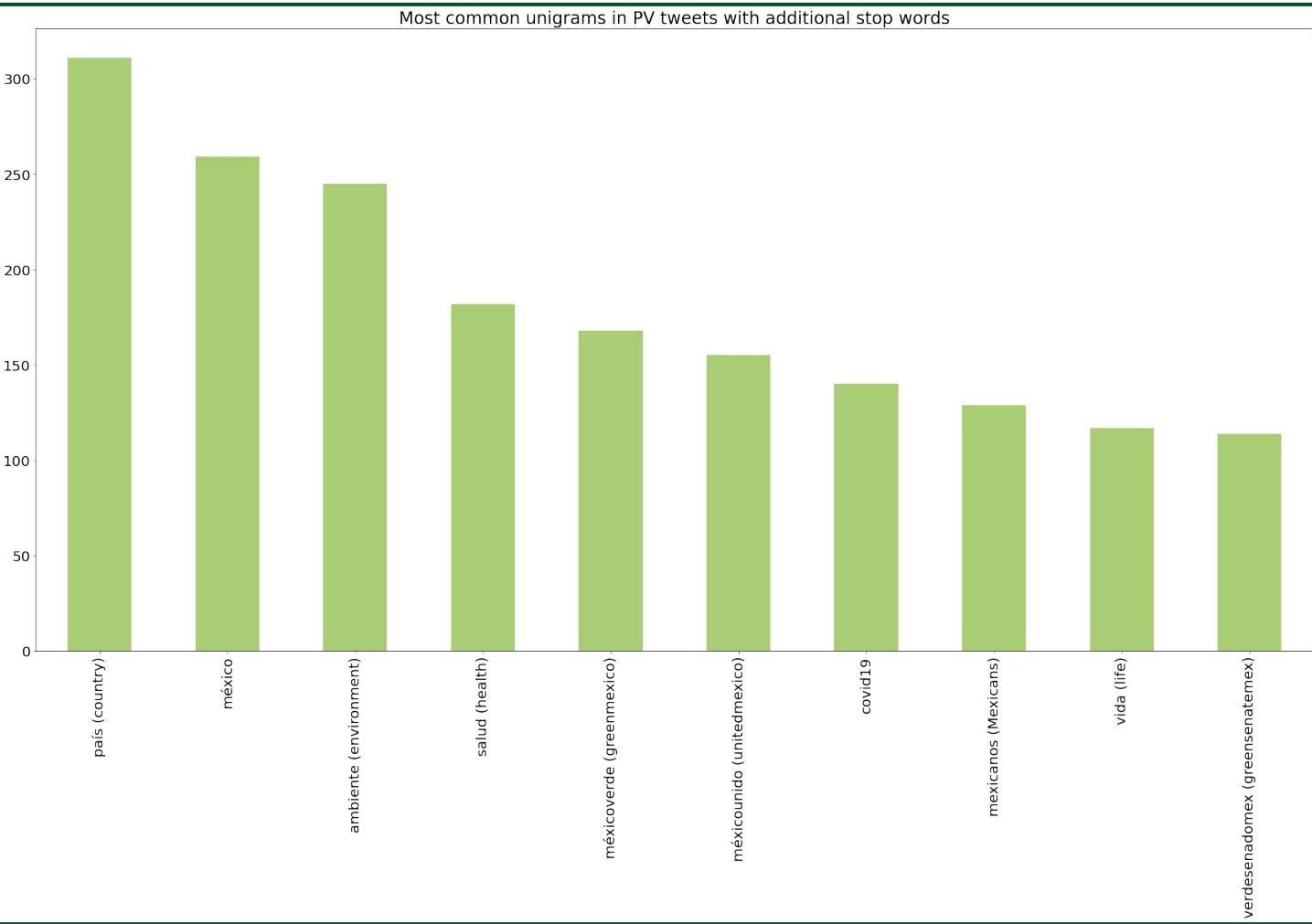
Most common unigrams in PRI tweets with additional stop words



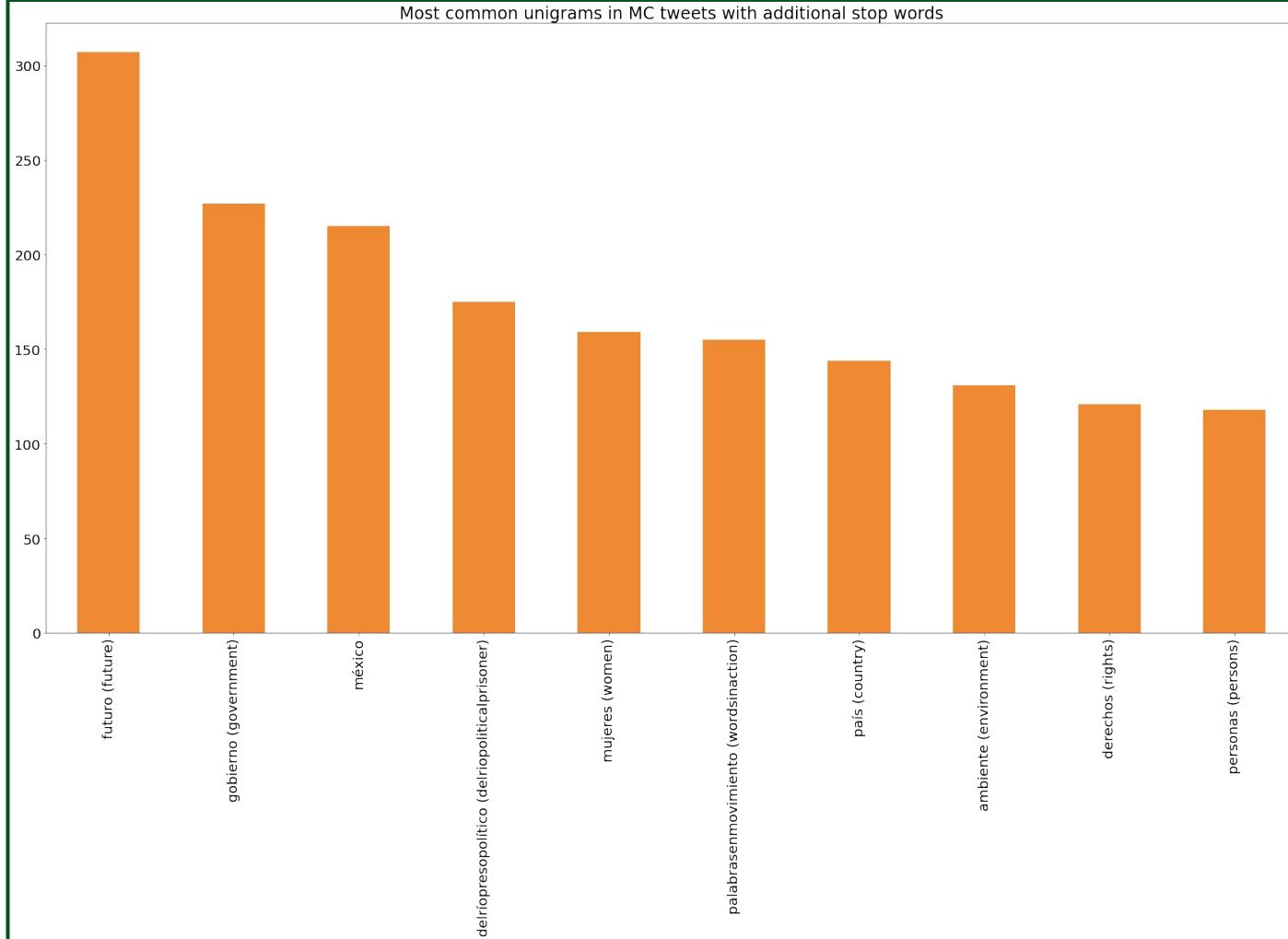
Most common unigrams in PRD tweets with additional stop words



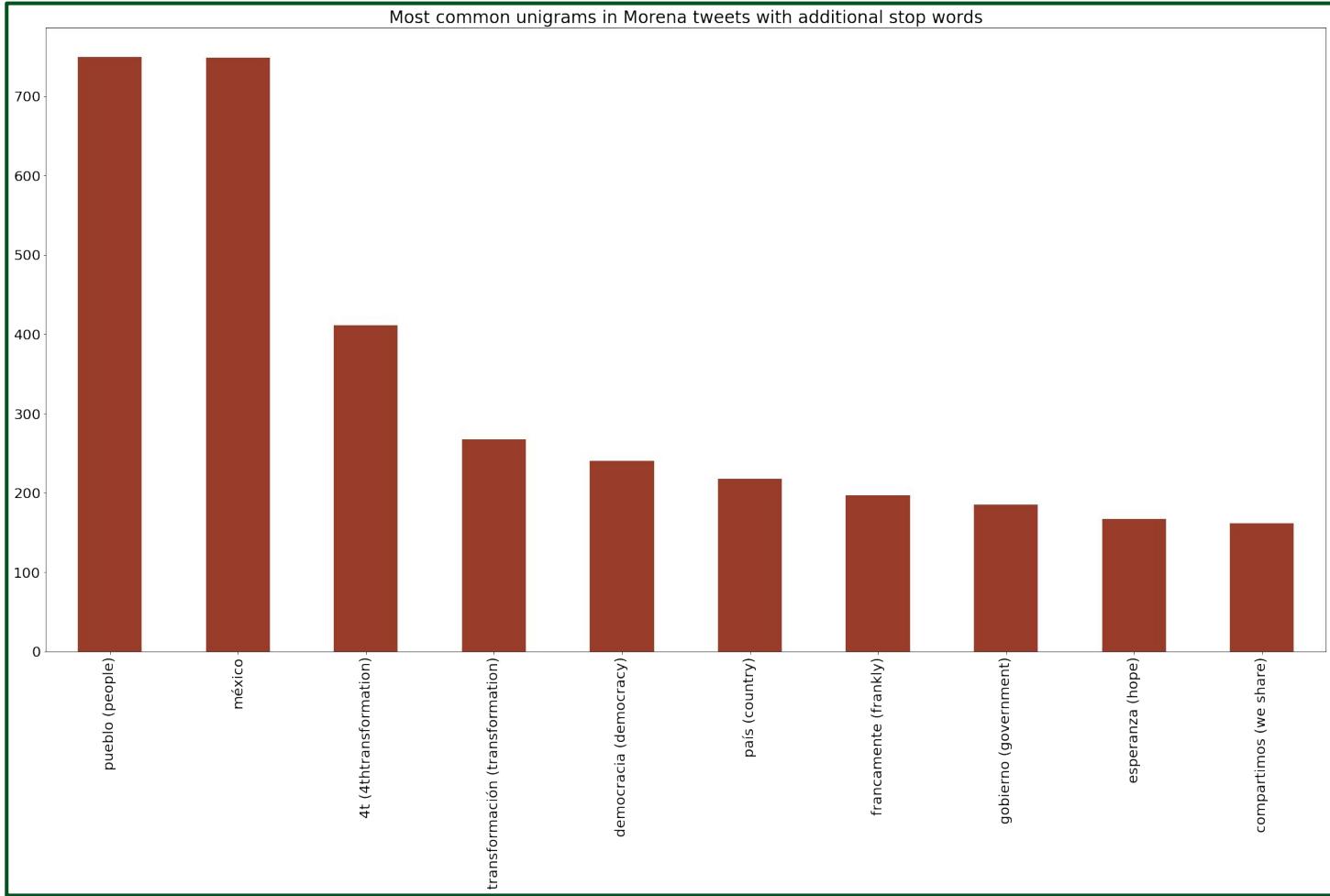
Most common unigrams in PV tweets with additional stop words



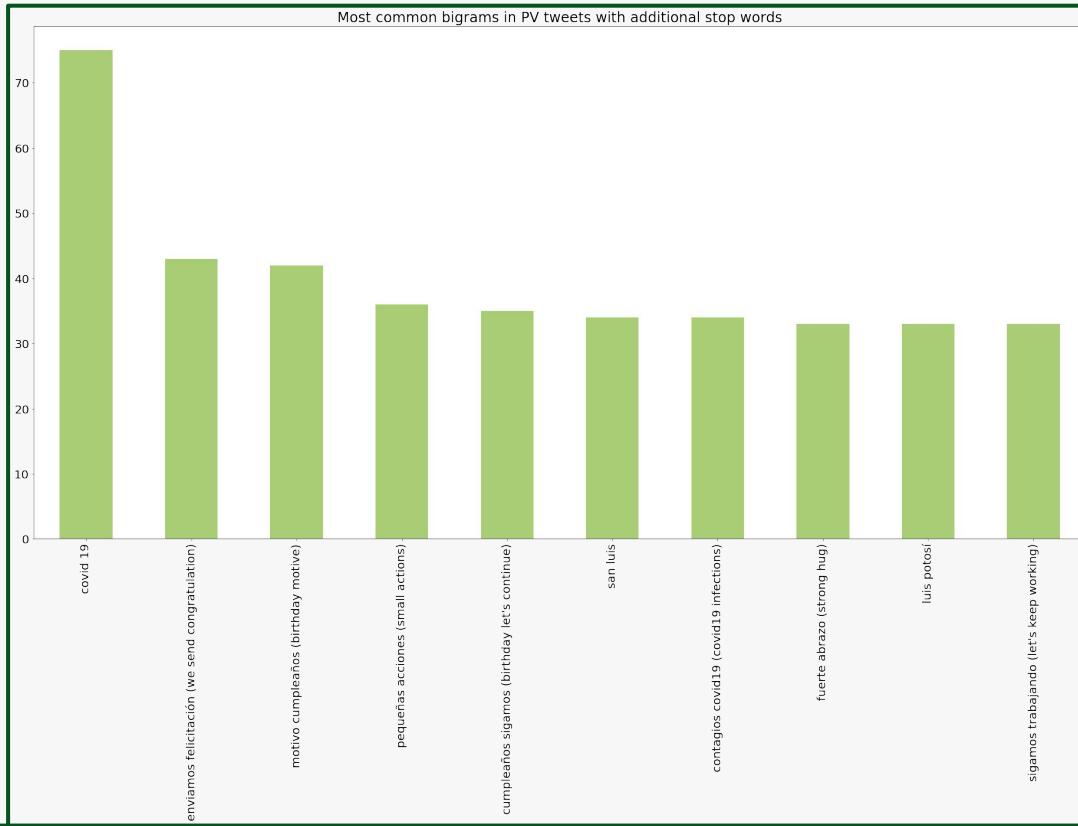
Most common unigrams in MC tweets with additional stop words



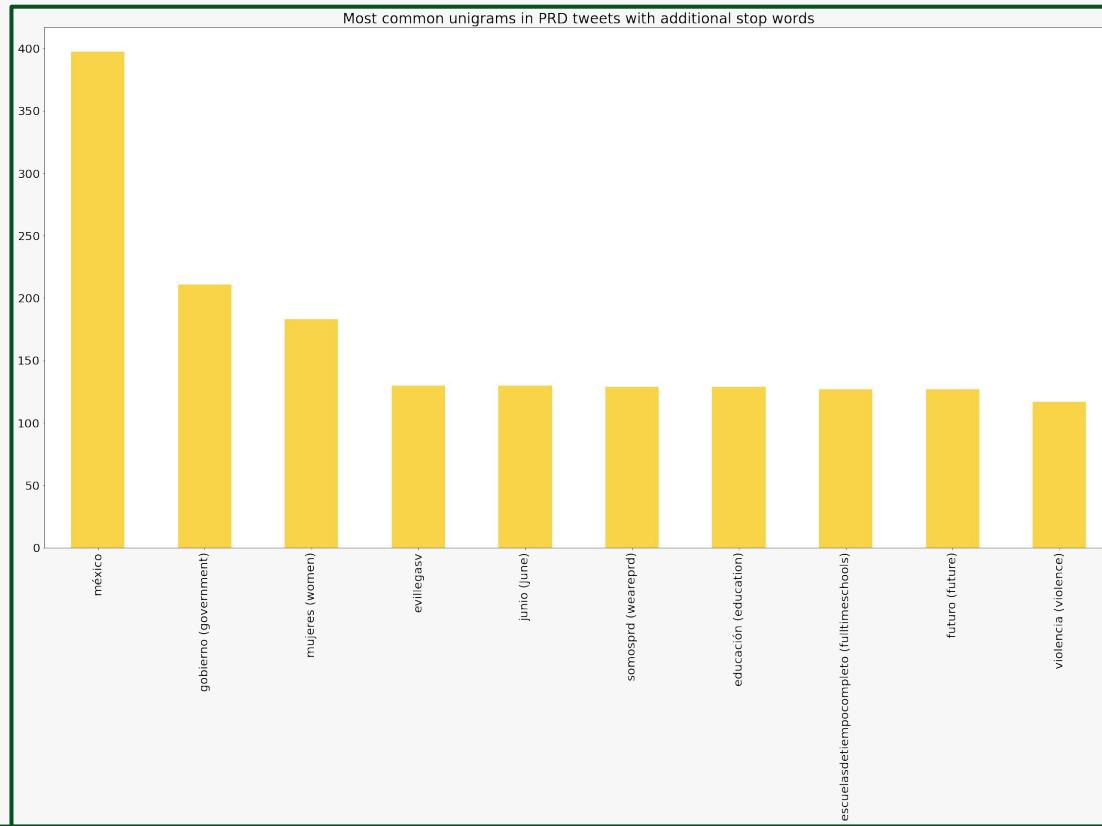
Most common unigrams in Morena tweets with additional stop words

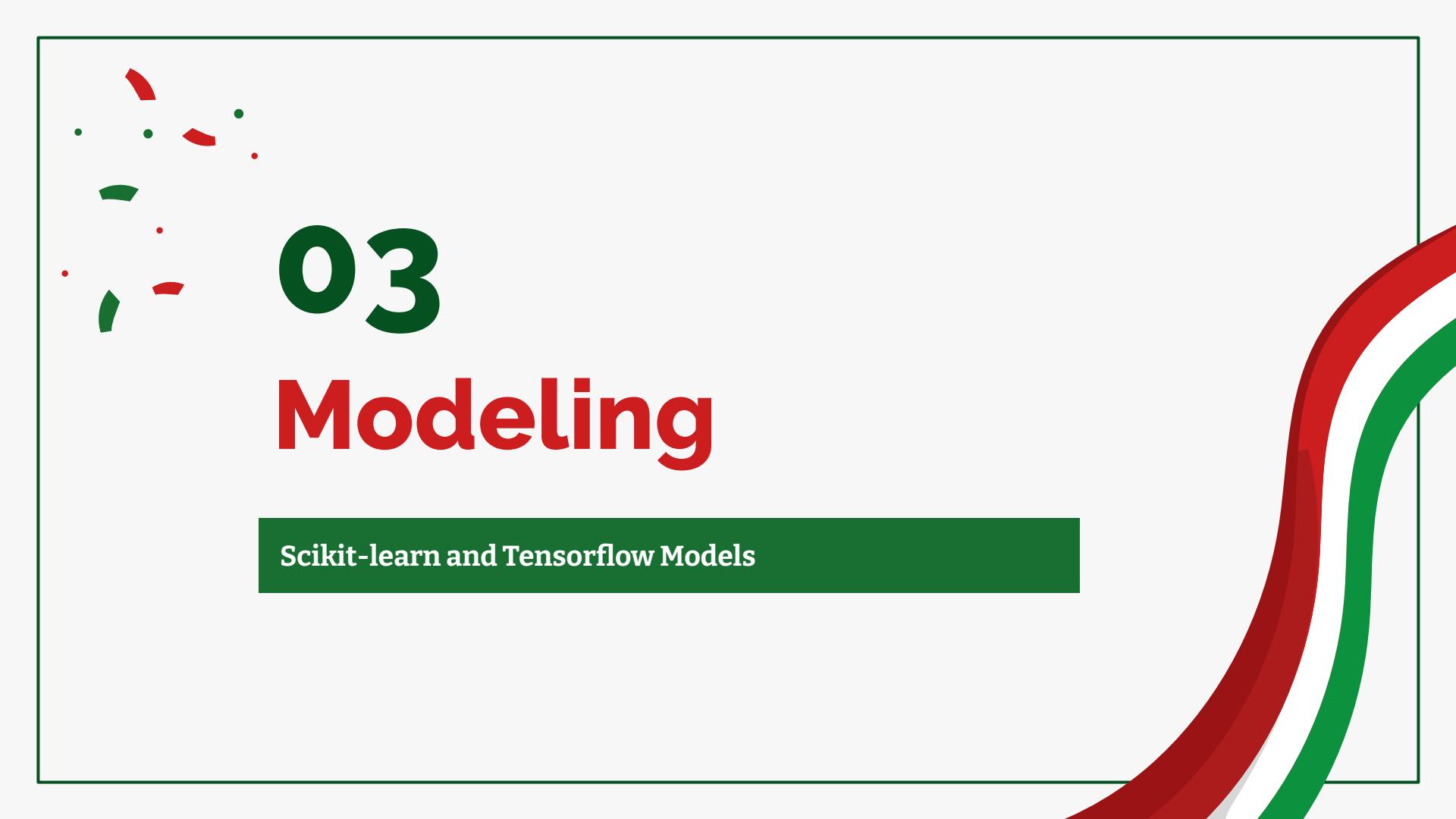


Clean energies?



Addressing violence?





03

Modeling

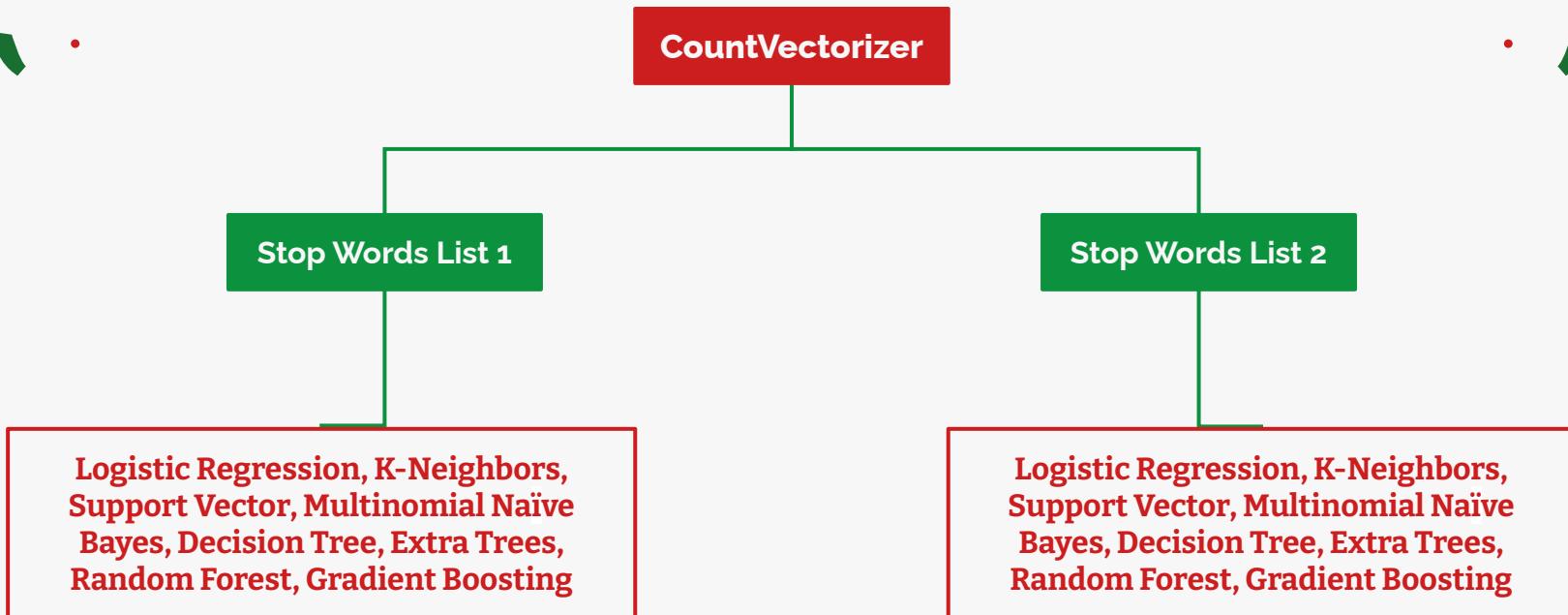
Scikit-learn and Tensorflow Models



scikit
learn

The background features abstract, wavy shapes in red, green, white, and blue, partially visible behind the central orange circle.

Model Diagram



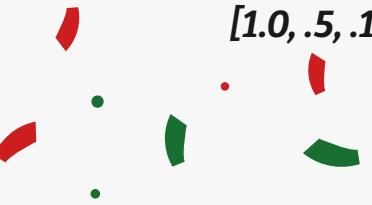
Hyperparameter Tuning



RandomizedSearchCV

Performed randomized search over a wide range of hyperparameters

- **'multinomialnb_alpha':**
[1.0, .5, .1]



GridSearchCV

Performed thorough search over a narrowed down range of hyperparameters:

- **'multinomialnb_alpha':**
[.4, .5, .6]



Best hyperparameters

Obtained the best hyperparameters using both searches:

- **'multinomialnb_alpha':**
.4



Best Scikit-Learn Models

Classifier	Stop Words List	Accuracy Score
Multinomial Naïve Bayes	1	.9040
Multinomial Naïve Bayes	2	.8773
Gradient Boosting	1	.8771
Logistic Regression	2	.8696

CountVectorizer + Multinomial Naive Bayes Classifier Hyperparameters

(2,2)

N-gram range

1

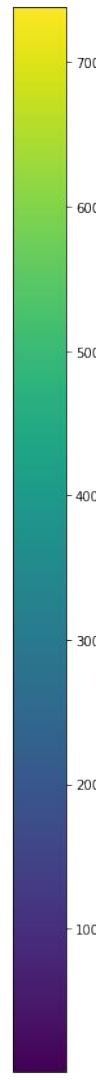
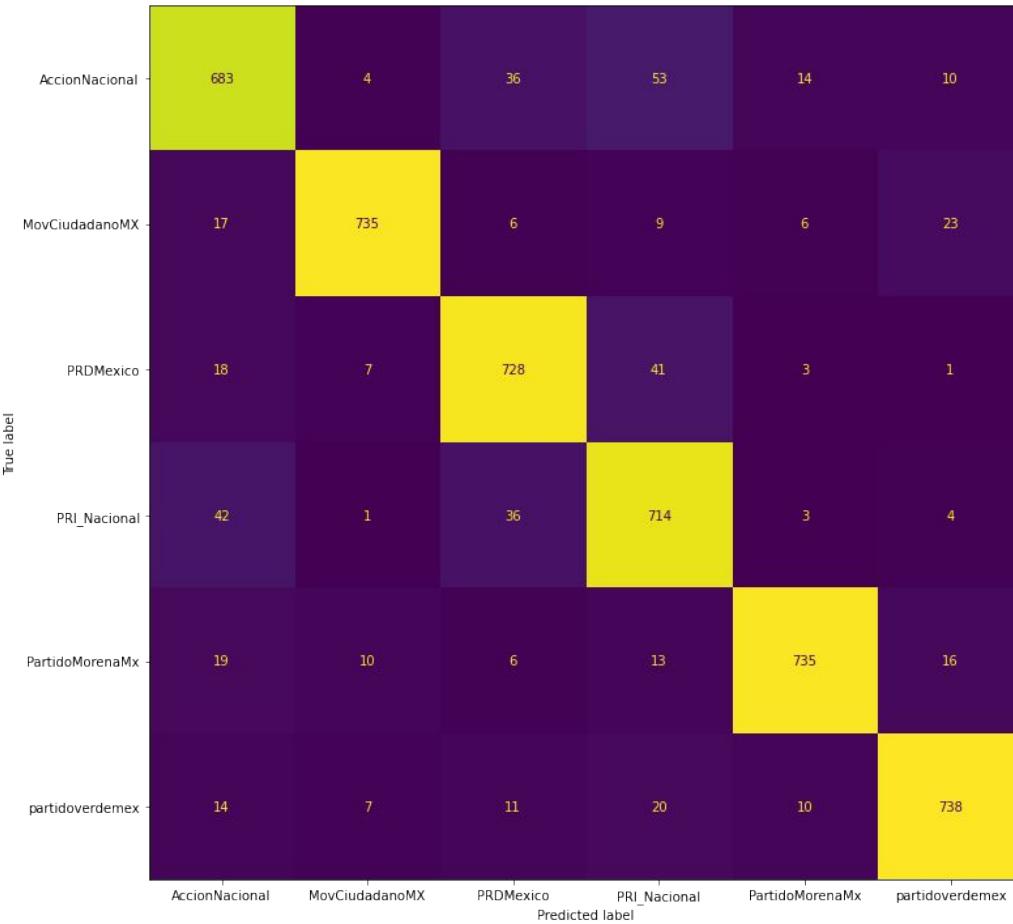
Stop Words List

Stemming

.4

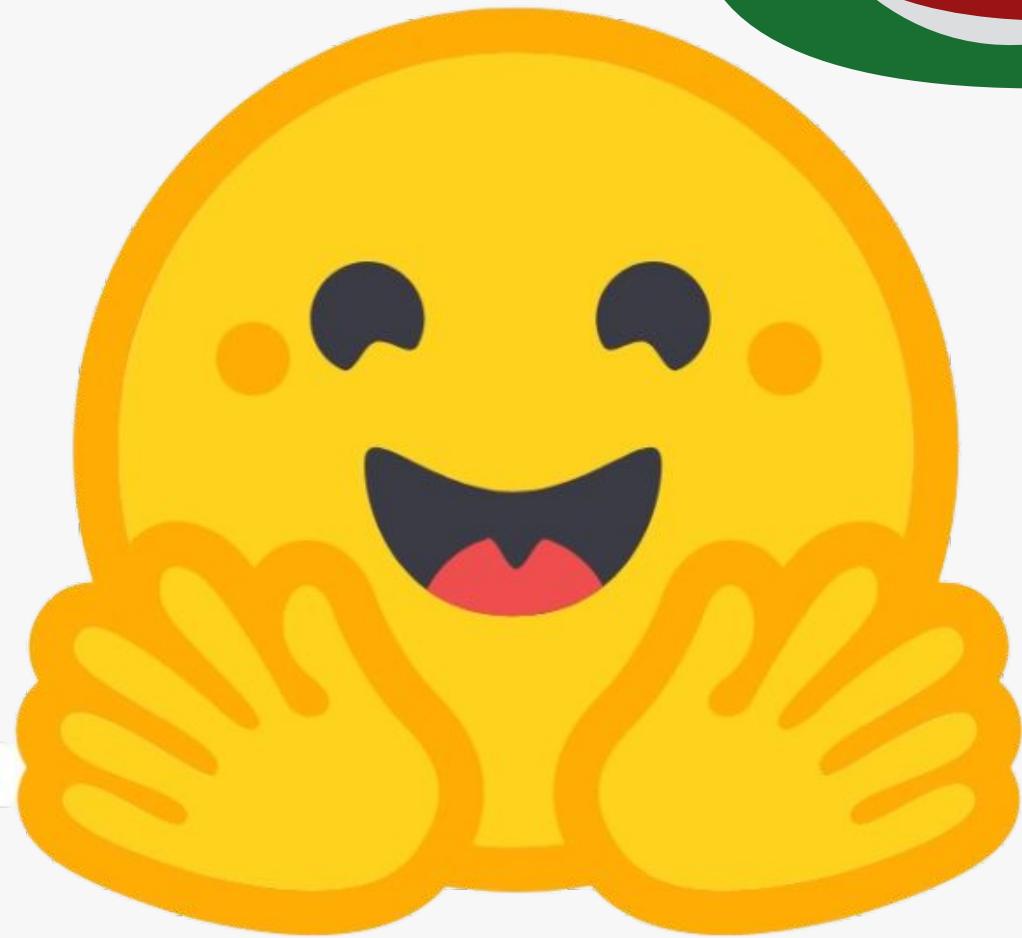
Tokenizer

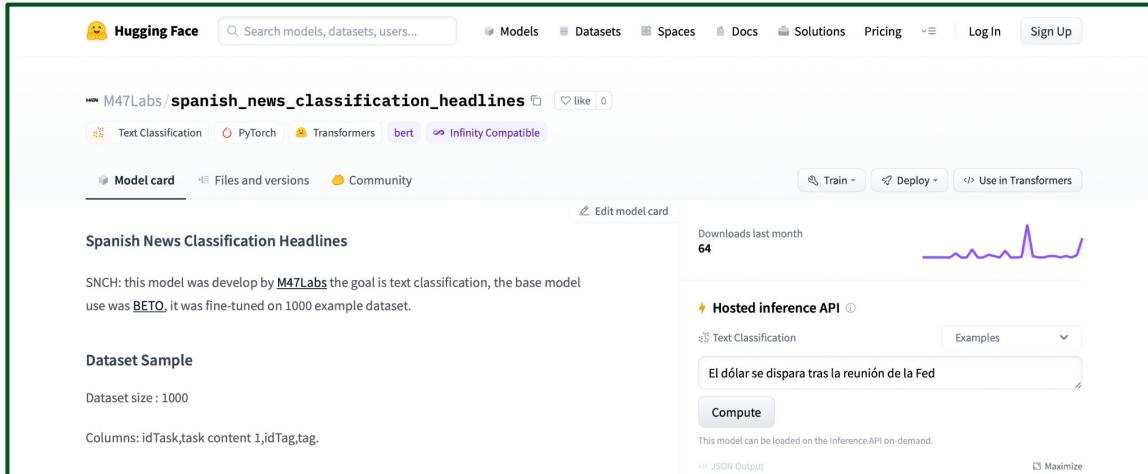
Alpha





TensorFlow





Input Layer

1. BERT Model
2. Pooling

Hidden Layer

1. Dense
2. ReLU activation

Output Layer

1. Dense
2. Softmax Activation



.9930

Validation accuracy score after 10 epochs

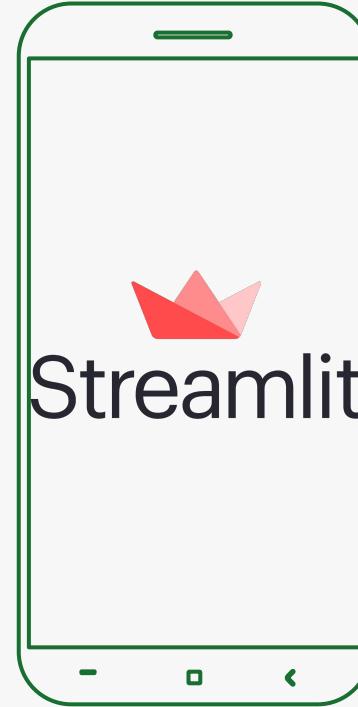


04

Evaluation

Let's try it out!

App: Mexican Political Party Tweet Classifier



Limitations

Hashtags...

make tokenizing challenging.

The stop words...

were NOT in their lemma or stem form.

The character limit...

was supposed to be 280?!?

@azcarmx @ernesto_mendez @WWF_LAC
@unep_espanol @prensaanimal4 @lajornadaonline
@ContraReplicaMX @heraldodemexico @ImagenTVMex
@laoctavadigital @Reforma Buen día, igual que a
ustedes, nos preocupan muchas cuestiones en torno a
esta obra, una de ellas es el uso de diésel. Por ello, ya
pedimos a @SCT_mx informe el impacto ambiental que
genera su uso, al igual que priorizar fuentes
sustentables que eviten daños al medio ambiente.

Everyone can reply



-151



Tweet

Future Directions

- **Collect** 14K tweets from each party
- **Convert** stop words lists to their stem/lemma form
- **Implement** 16 TFIDFVectorizer models (8 with Stop Words List 1 and 8 with Stop Words List 2)



05

Conclusion

Making sense of Mexico's political parties

CONCLUSIONES

7-Party System



Even with ideological overlap and alliances, these parties are distinct enough to be successfully classified.

Talking like a Politician



Political speak might fool humans, but models learn the true patterns in party messaging..

Change Needed



Mexico needs change *now*, and it needs **honest** politicians with *clear* plans for enacting change.



¿Questions?



¡Gracias!