

Parámetros y pruebas en R

1

Clasificadores

2

Número de
palabras frecuentes

3

Stopwords

4

Bolsa de
palabras propias

Bolsa de palabras propias

nerviosa cumpleaños
feliz agradecida hija
bebé maravilloso

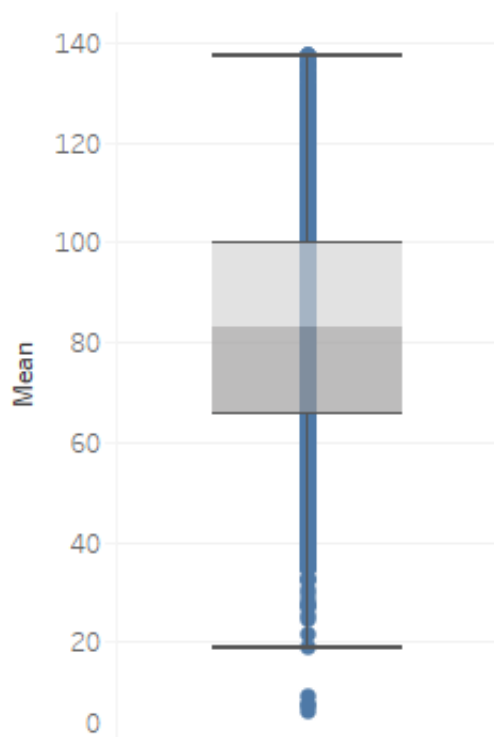
ganar batalla
libertad
perder enemigo

Modelo	Clasificador	Stopwords	Bag of words	Acc. Gender	Acc. Variety	Acc. Joint
RF n= 100, sw + bw	Random Forest	si	si	0,6571	0,4850	0,3107
RF s= 10, sw + bw	Random Forest	si	si	0,6571	0,2750	0,1800
SVM n= 10, sw + bw	SVM	si	si	0,6507	0,2642	0,1750
RF n= 10, sw	Random Forest	si	no	0,6514	0,2721	0,1714
SVM n=10, sw	SVM	si	no	0,6407	0,2492	0,1621
SVM n=10	SVM	no	no	0,5635	0,2157	0,1185
RF n= 10	Random Forest	no	no	0,5721	0,2064	0,1142

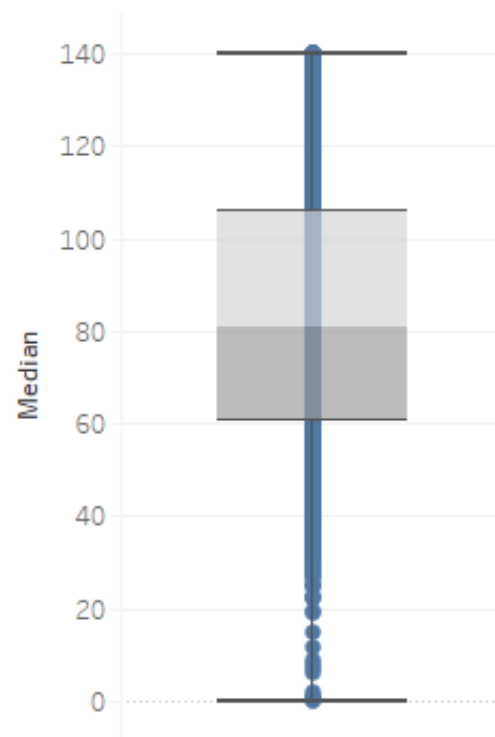
- El uso de las **stopwords** y la **bolsa de palabras** mejora considerablemente el accuracy.
- Usar una mayor cantidad de palabras mejora la precisión de la variedad debido a su gran margen de mejora.

Longitud del tweet

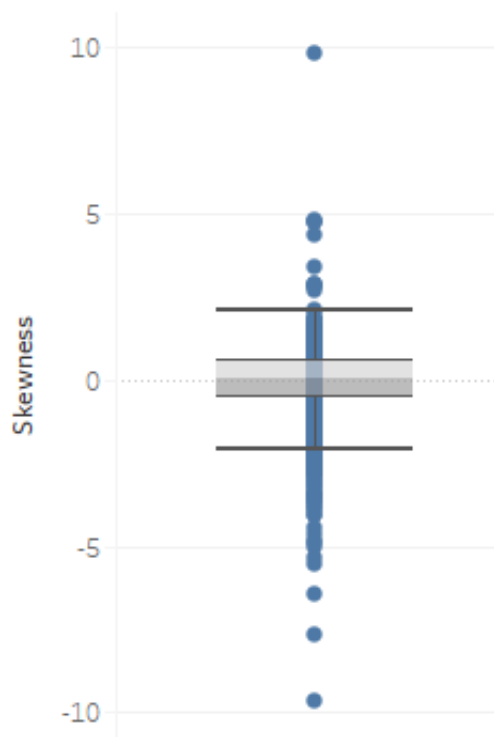
Mean



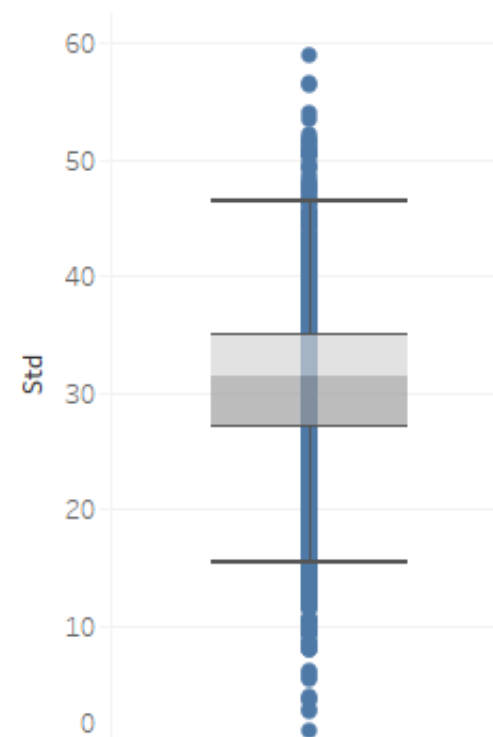
Median

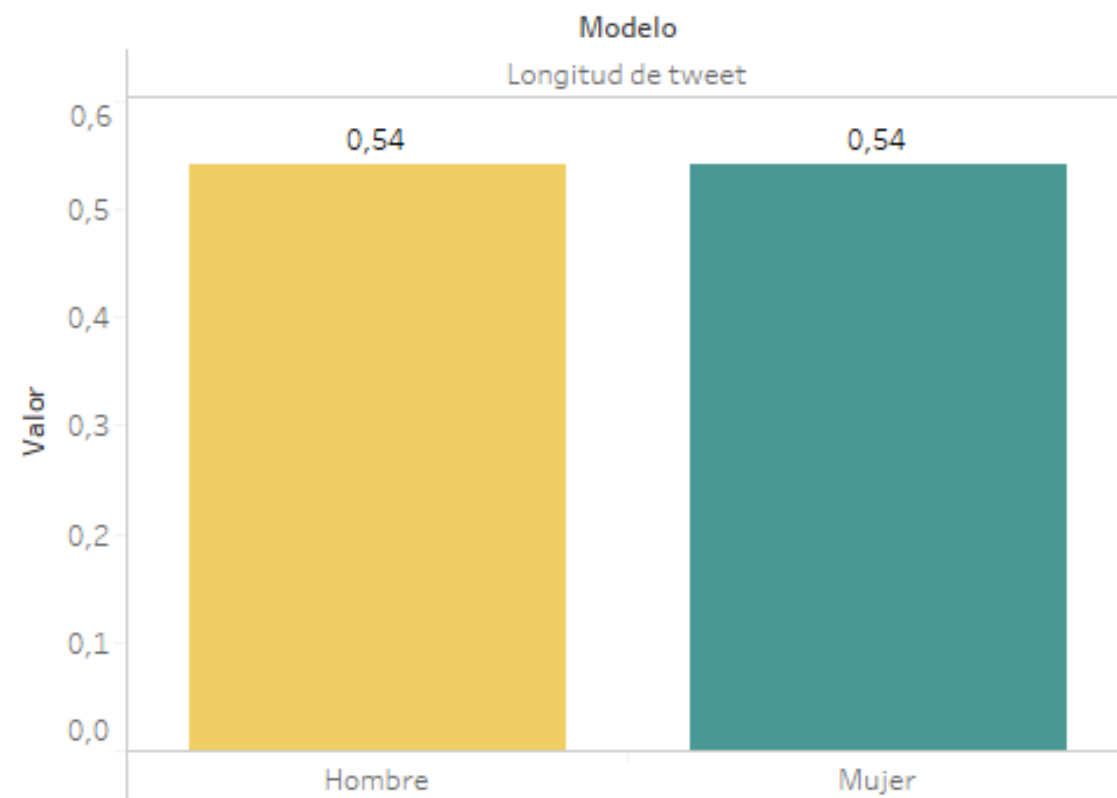
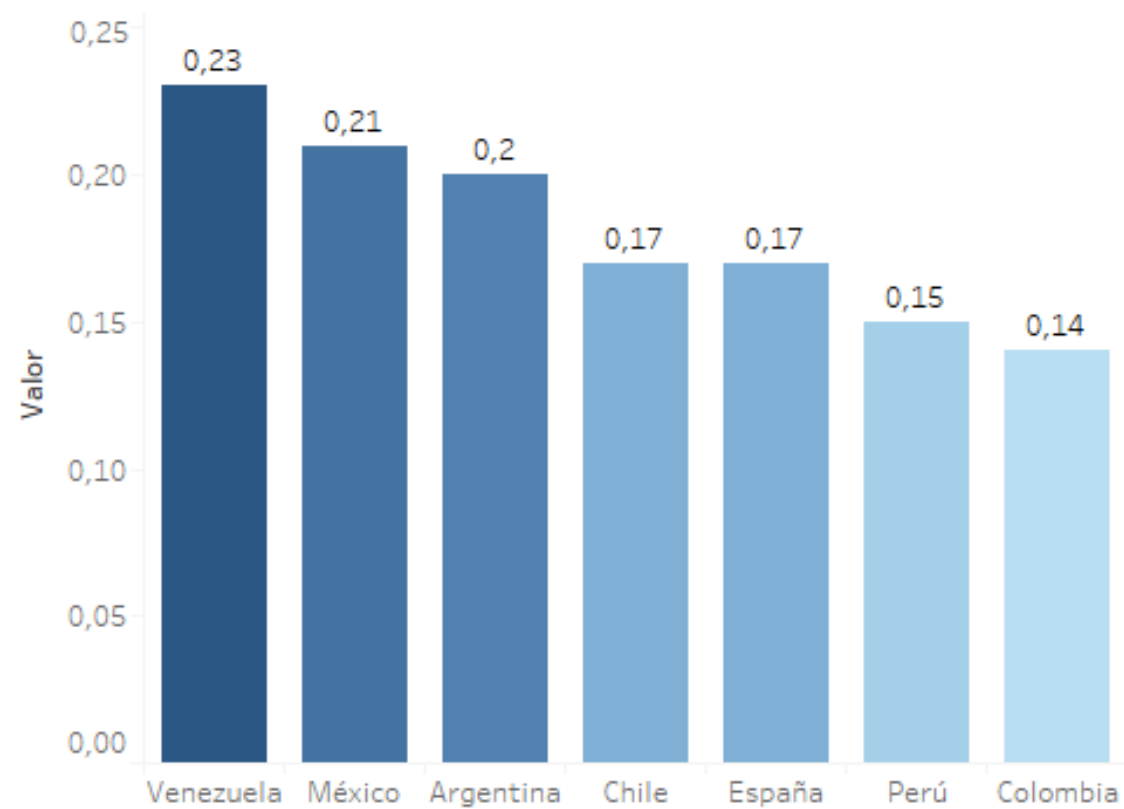


Skewness

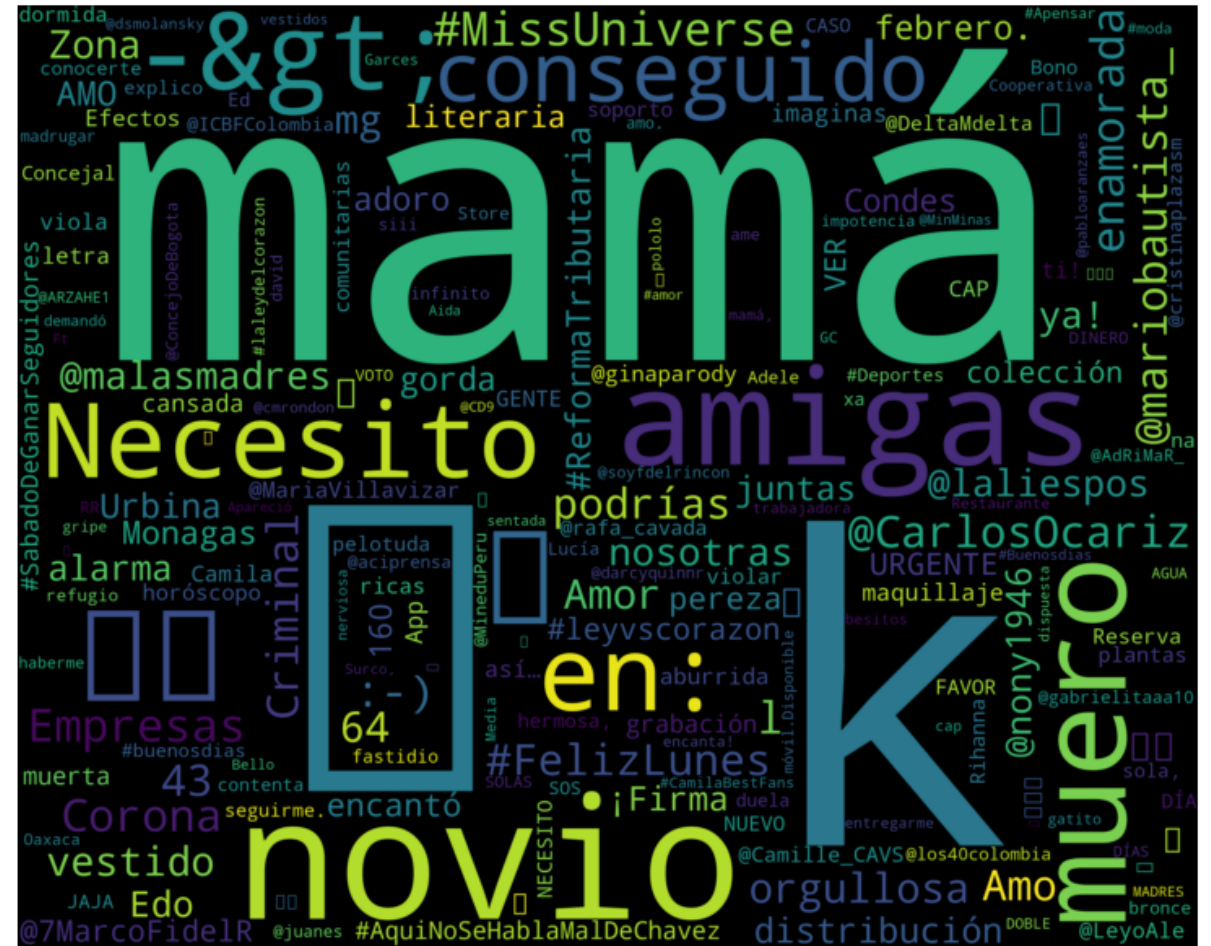


Std





Bag of Words

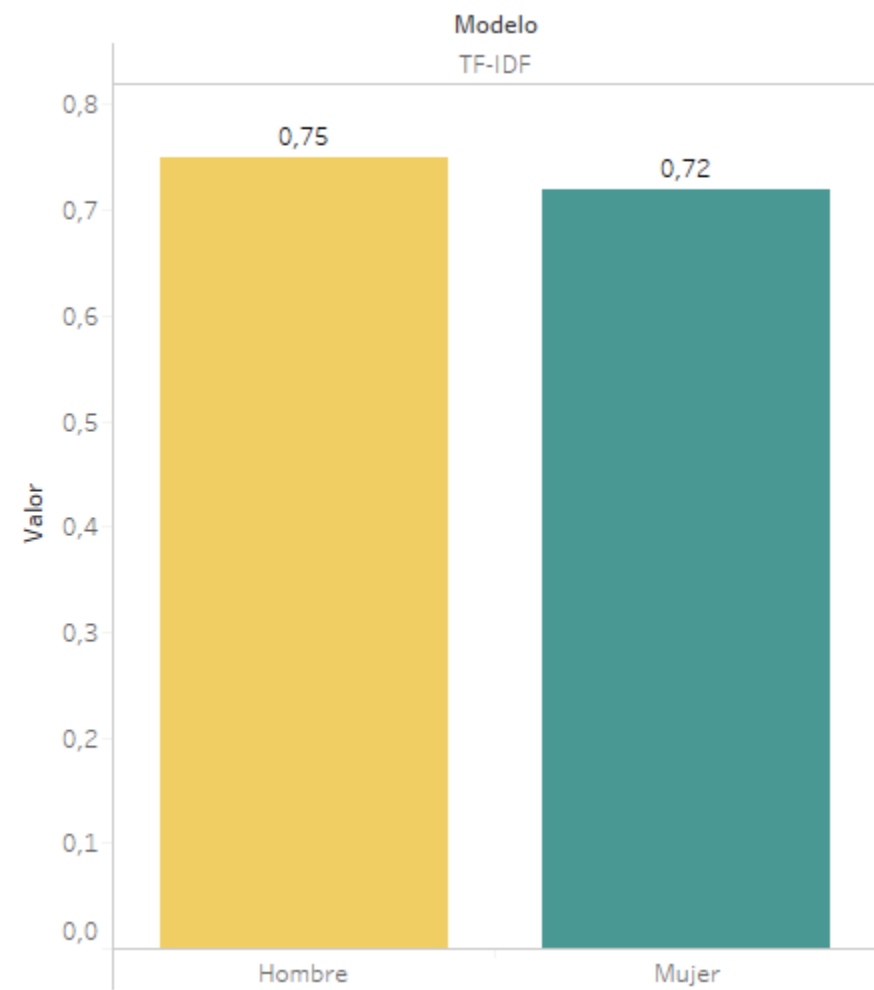
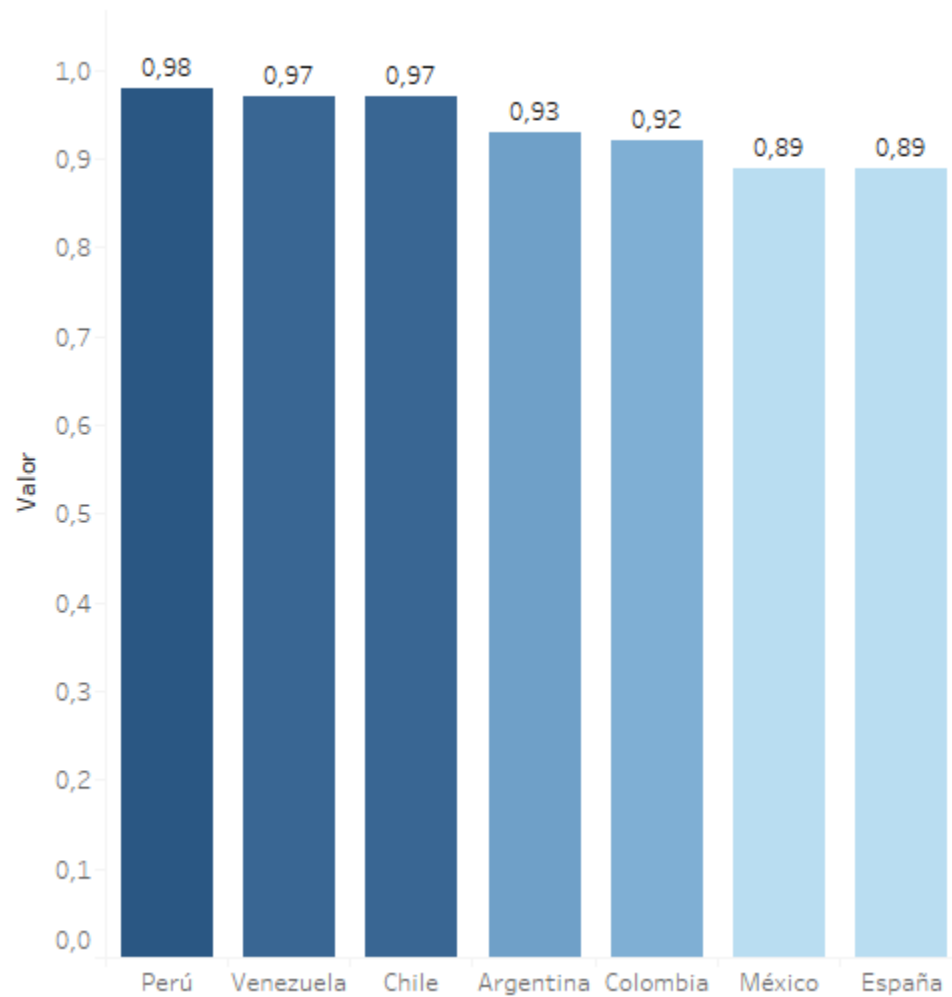


0	equipo	male
1	partido	male
2	fútbol	male
3	vs	male
4	acerca	male
5	gol	male
6	jugadores	male
7	juega	male
8	jugador	male
9	equipos	male
10	@MillosFCoficial	male
11	Colombia,	male
12	Copa	male
13	Diario	male
14	goles	male
15	latest	male

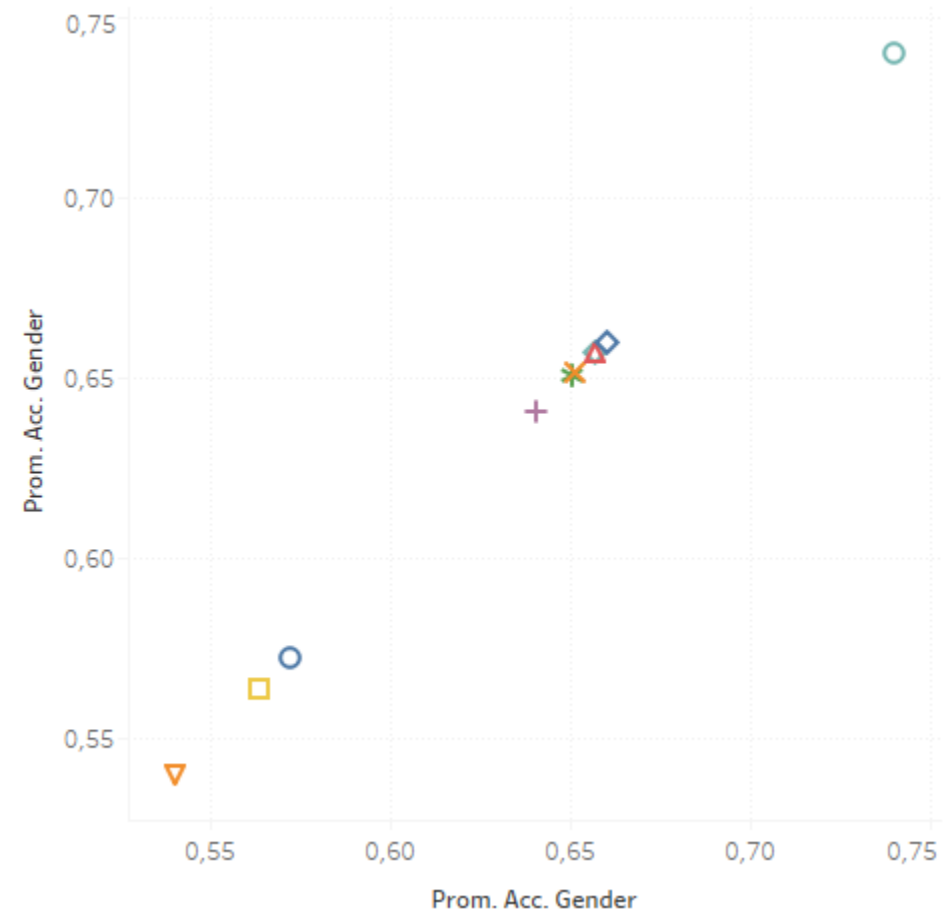
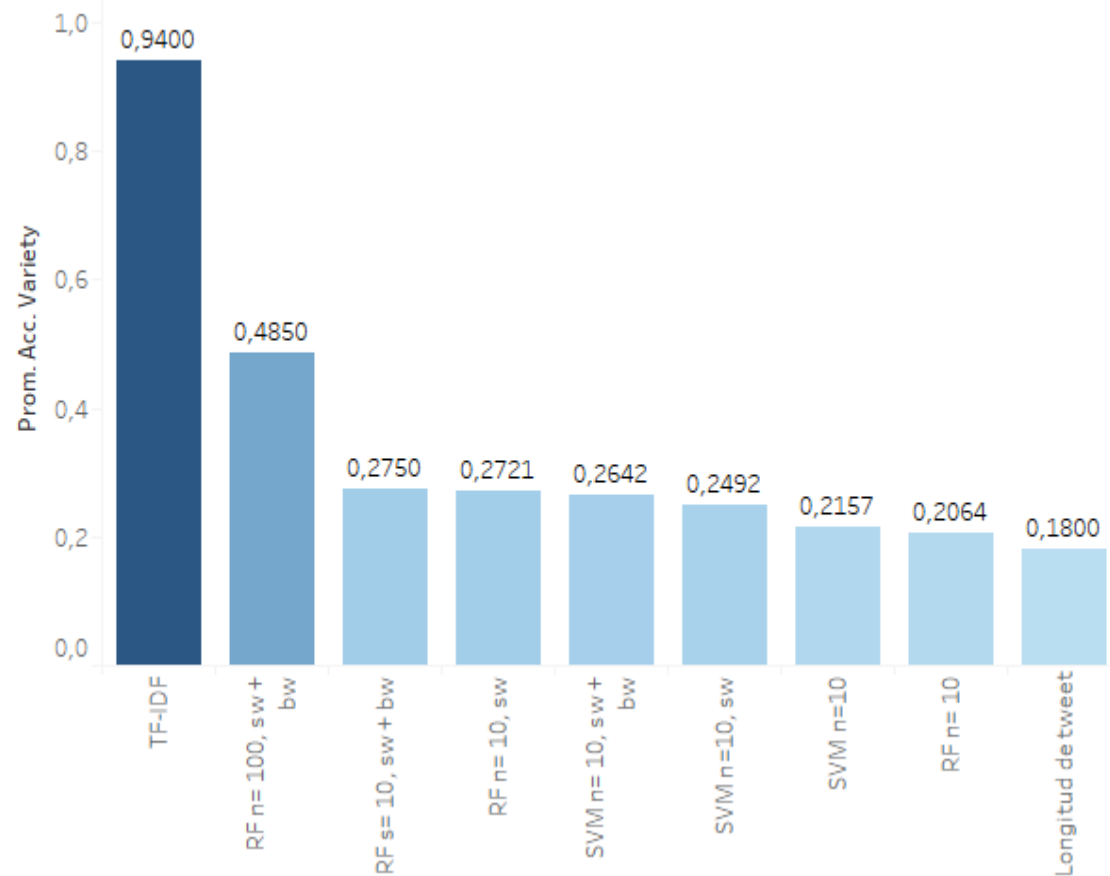
7515	❤️😄	female
7516	🌈	female
7517	💡	female
7518	🎵Me	female
7519	🎵No	female
7520	🐱	female
7521	👎	female
7522	❤️🇪🇸	female
7523	📖	female
7524	😂A	female
7525	😂Es	female
7526	😂Porque	female
7527	😂Qué	female
7528	😄😄	female
7529	😄Que	female
7530	😄.	female

- Las **mujeres** usan **emoticonos** en mayor frecuencia que los hombres
- Los **hombres** en cambio, suelen hablar mucho más de temas de **deportes**

TF-IDF



Resumen de modelos



Futuras mejoras

- Unir características

- TF-IDF
- Bolsa de palabras hombre/mujer
- Longitud de tweet
- Emoticonos
- N-gramas

- Añadir características

- Negación
- Pronombres
- Tiempos verbales
- Terminación de las palabras

