

Silvia Aroca

Natalia Cervelló

Elena Tejadillos



Publicidad

Marketing

Community Manager

Redes sociales

Twitter

Facebook

LinkedIn

Xing

Foursquare

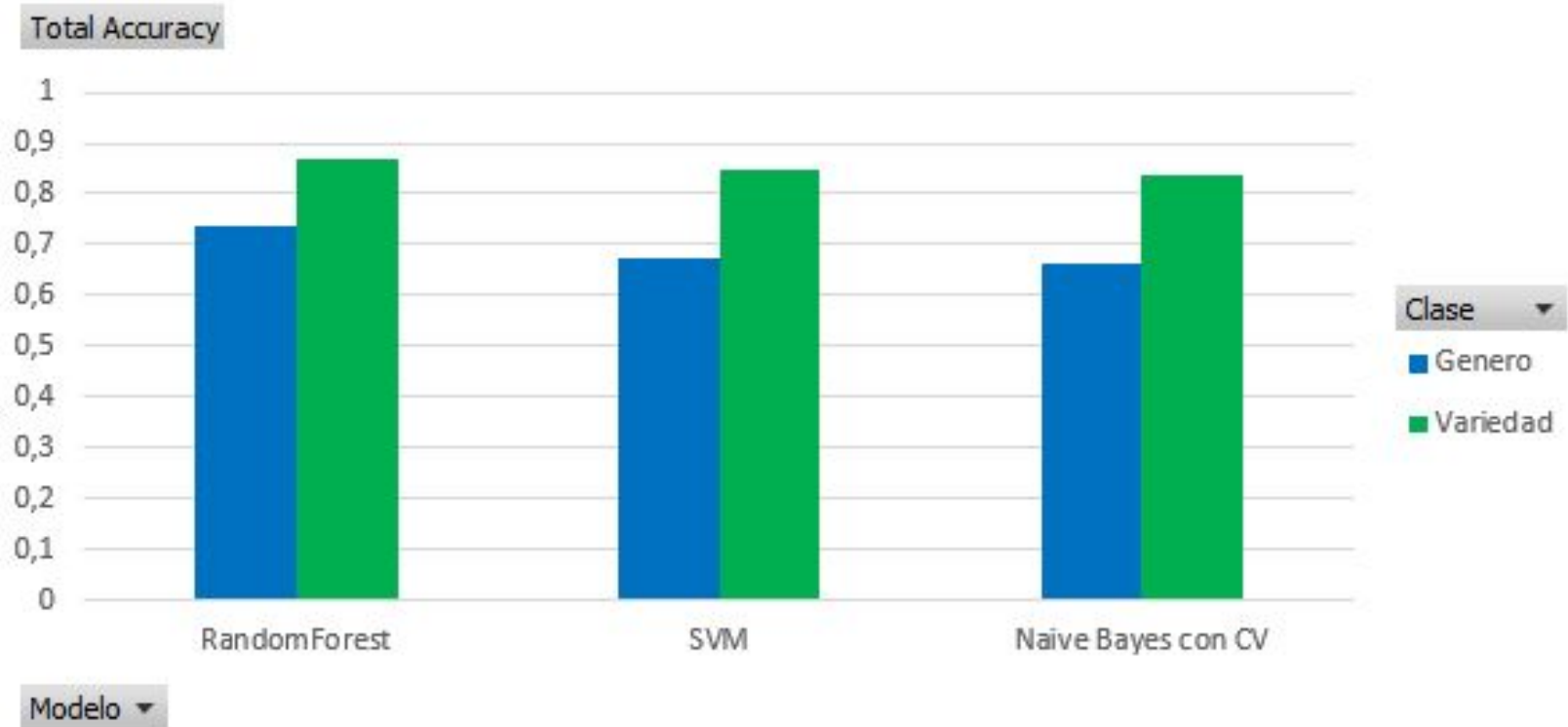
Investigación

- Nos enfocamos en la VARIEDAD
- De las 2000 palabras más frecuentes, nos centramos en las 100 primeras
- Descubrimos y eliminamos:
 - Abreviaturas a la hora de escribir (q,d,x)
 - Onomatopeyas (jajaja, jajajaja)
 - Nombres propios (Trump, Youtube)

Pasos Realizados

- Limpieza de datos eliminando:
 - Acentos
 - Palabras seleccionas en el studio (stopwords)
- Aumentamos de la bolsa de palabras a 2000
- Modelos aplicados
 - Support Vector Machine
 - Naive Bayes con Cross Validation
 - Radom Forest
 - Variedad: 86,79
 - Sexo: 73,43

Resultados ... por ahora



¡GRACIAS!

¡Y SUERTE A TODOS!

