

Tarea de la sesión 4

Osman Dennys Rafael Taipe

Análisis descriptivo

Inicialmente extraemos unos gráficos de nubes de palabras para explorar la frecuencia que tienen los tweets de Daniel Urresti, Mininter Perú y Minsa Perú. Claramente en los tweets de Daniel Urresti destacan las palabras: AFP, retiro, comisión, congreso, proyecto; debrido al proyecto de ley de retiro de AFP que se desarrollo por esa fecha. Para el caso de Minsa Perú destaca el nombre de Victor Zamora el ministro de Salud y palabras relacionados al contexto del Covid y finalmente para el Mininter Peru destaca el nombre del ministro y el contexto de la policía ante el covid.

Daniel Urresti



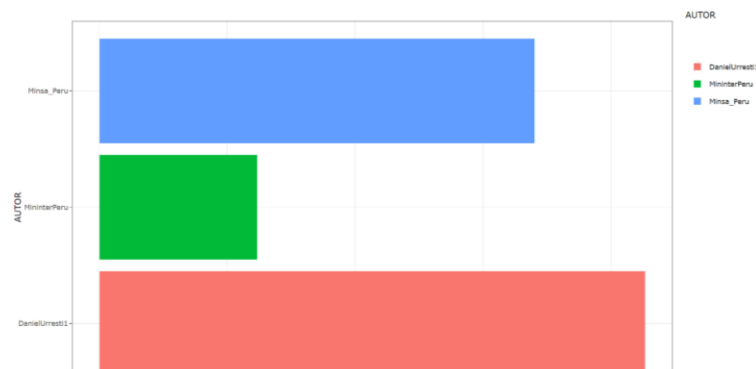
Minsa Peru

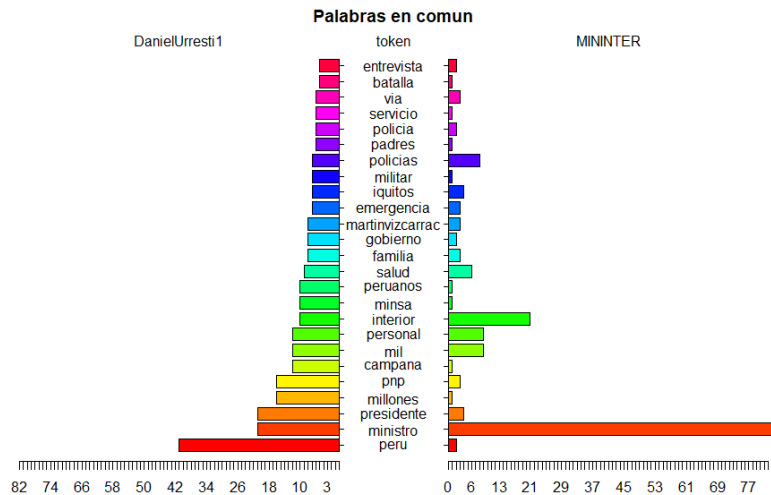


Mininter Peru



Para nuestro análisis de palabras comunes estamos eligiendo a las tweets de Daniel Urresti y de Mininter.





Modelos de Clasificación

En esta sección realizaremos un comparativo de los modelos de clasificación aplicados a los tweets de **Daniel Urresti y Mininter**, aplicando 3 modelos: Naive Bayes sin normalización, SVM con corpus limpio y itdf, SVM normalizado y Naive Bayes Normalizado.

Indicador	Naive Bayes Normalizado	Naive Bayes Normalizado	Naive Bayes Normalizado
Accuracy	83%	96%	94%
95% CI	(0.774, 0.8824)	(0.9028, 0.9719)	(0.9028, 0.9719)
No Information Rate	0.7374	-0.2475	0.7525
P-Value [Acc > NIR]	0.0009221	-1.00E+00	7.26E-13
Kappa	0.5559	-0.1442	0.8558
Mcnemar's Test P-Value	0.4862343	-0.7722	0.2278
Sensitivity	0.9041	0.9041	0.9463
Specificity	0.6346	0.6346	0.9388
Pos Pred Value	0.8742	0.8424	0.9792

Claramente el modelo que tiene un mejor performance es el de Naive Bayes Normalizado con un accuracy de 96% seguido por Naive Bayes Normalizado con un 94%. Pero si solo analizamos la clasificación adecuada de la clasificación positiva que es la Sensitividad el modelo a elegir sería el de Naive Bayes Normalizado