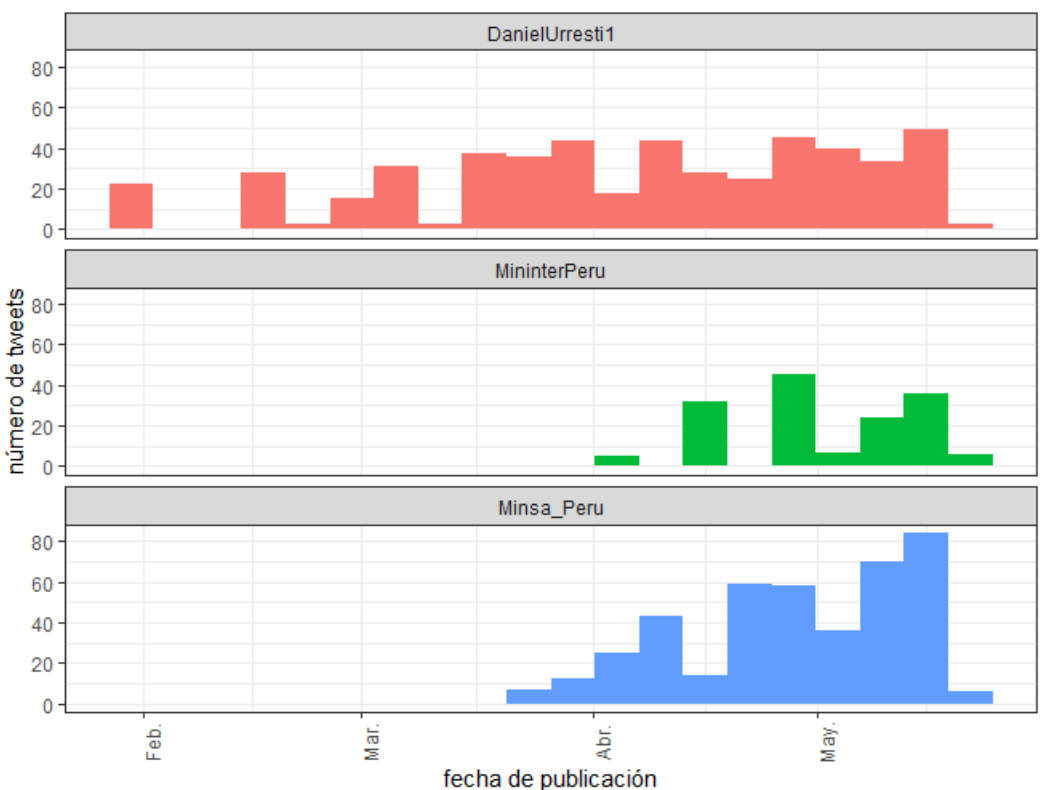
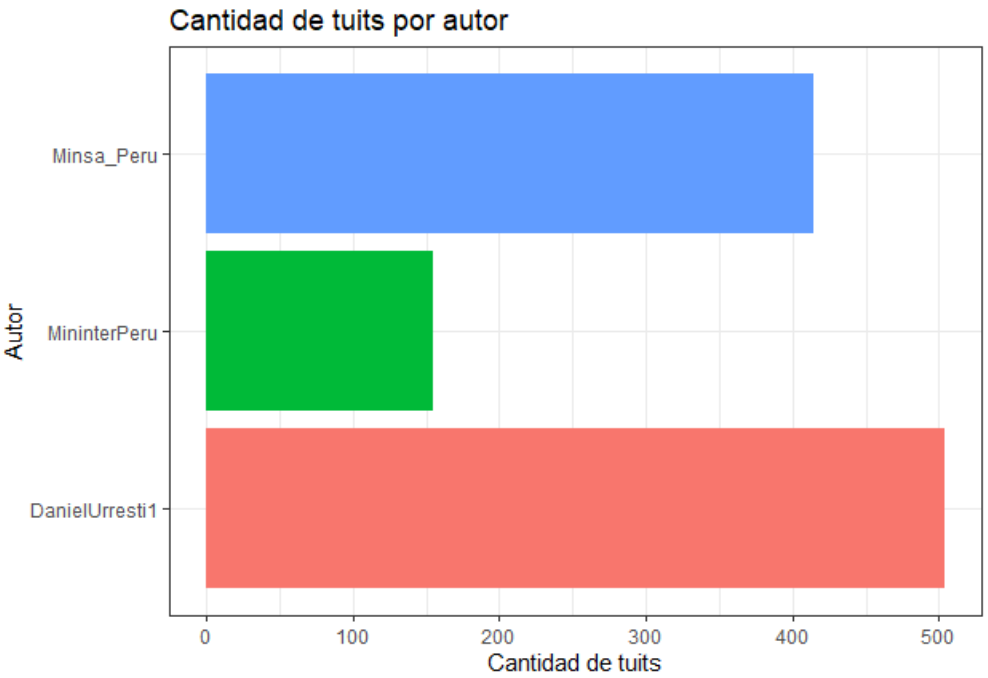
**Ejercicios Sesión 4**

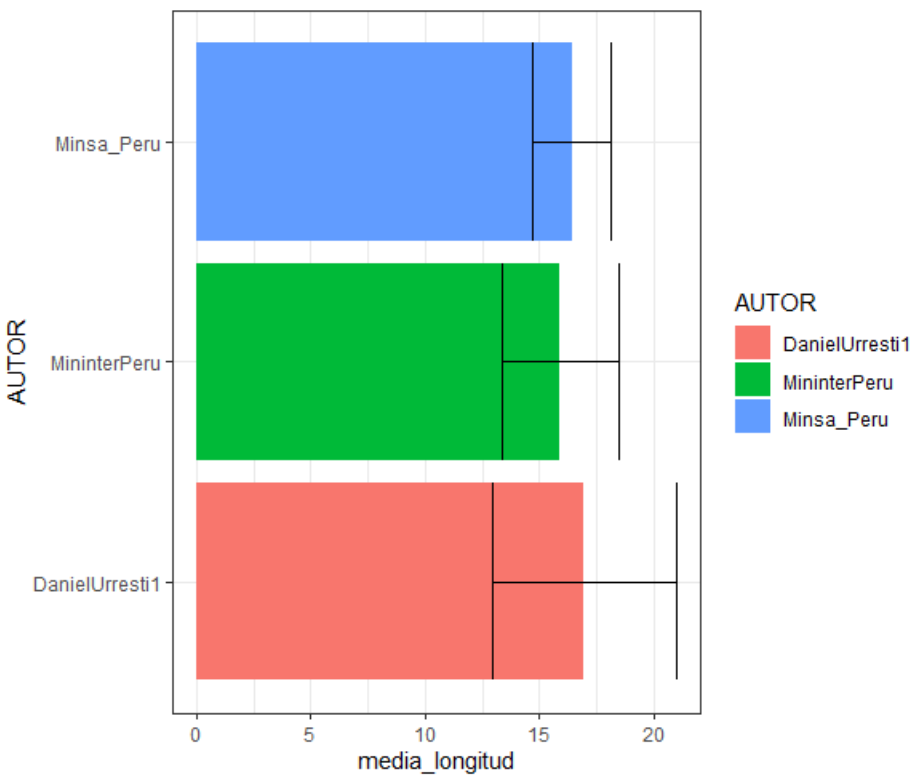
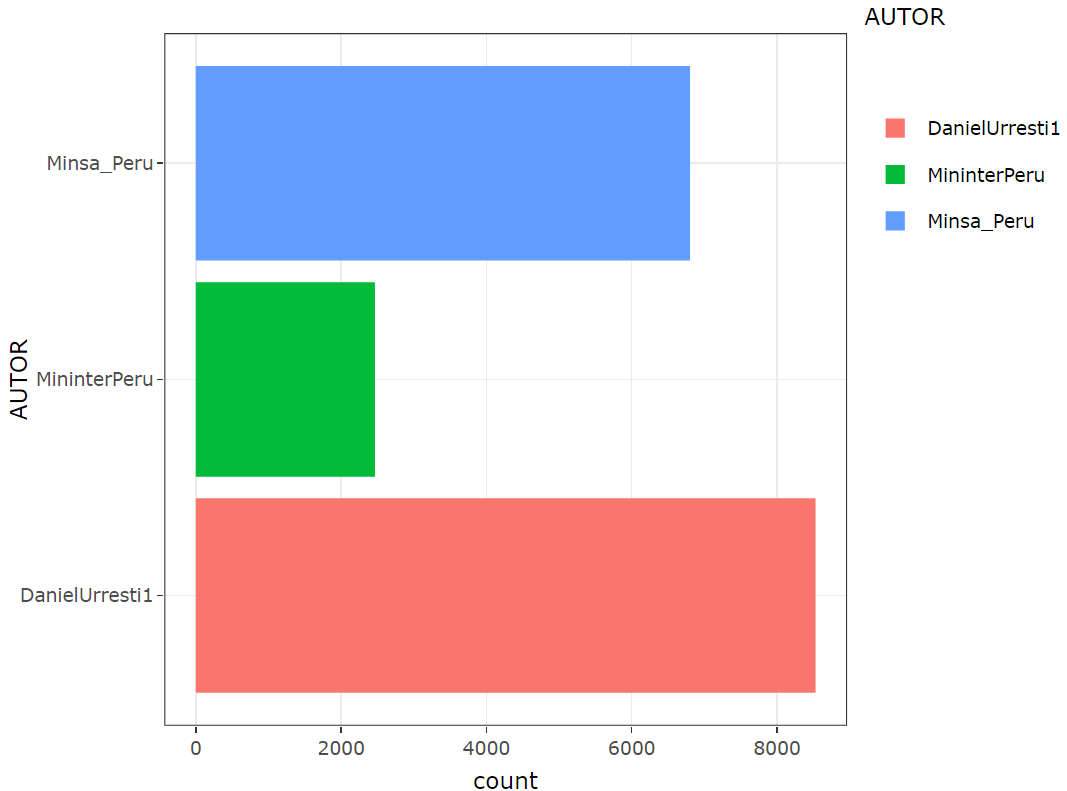
**Nombre: Alberto Armas Alvarado**

1. **Análisis descriptivo**



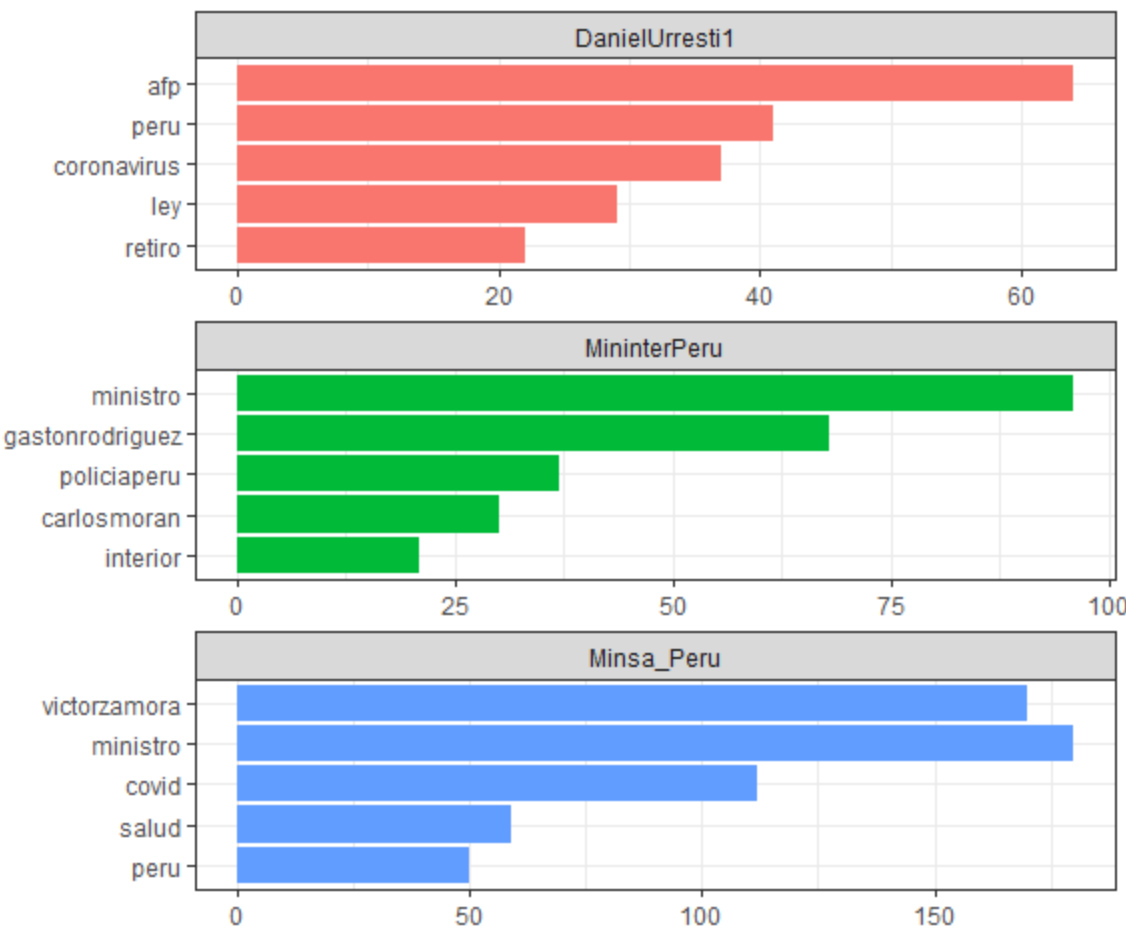
Los usuarios DanielUrresti1 y Minsa\_Peru tienen mayor cantidad de tuits, con 504 y 414 respectivamente.

Por fecha de publicación el usuario DanielUrresti1 ha sido más constante y tiene tuits más antiguos, mientras los usuarios MinterPeru y Minsa\_Peru tiene tuits recientes.



El usuario DanielUrresti1 usa más cantidad de palabras con 8532, seguido por Minsa\_Peru con 6804 palabras.

Como el usuario Daniel Urresti1 tiene mayor cantidad de tuits y mayor cantidad de palabras, tiene una mayor longitud media por tuits.

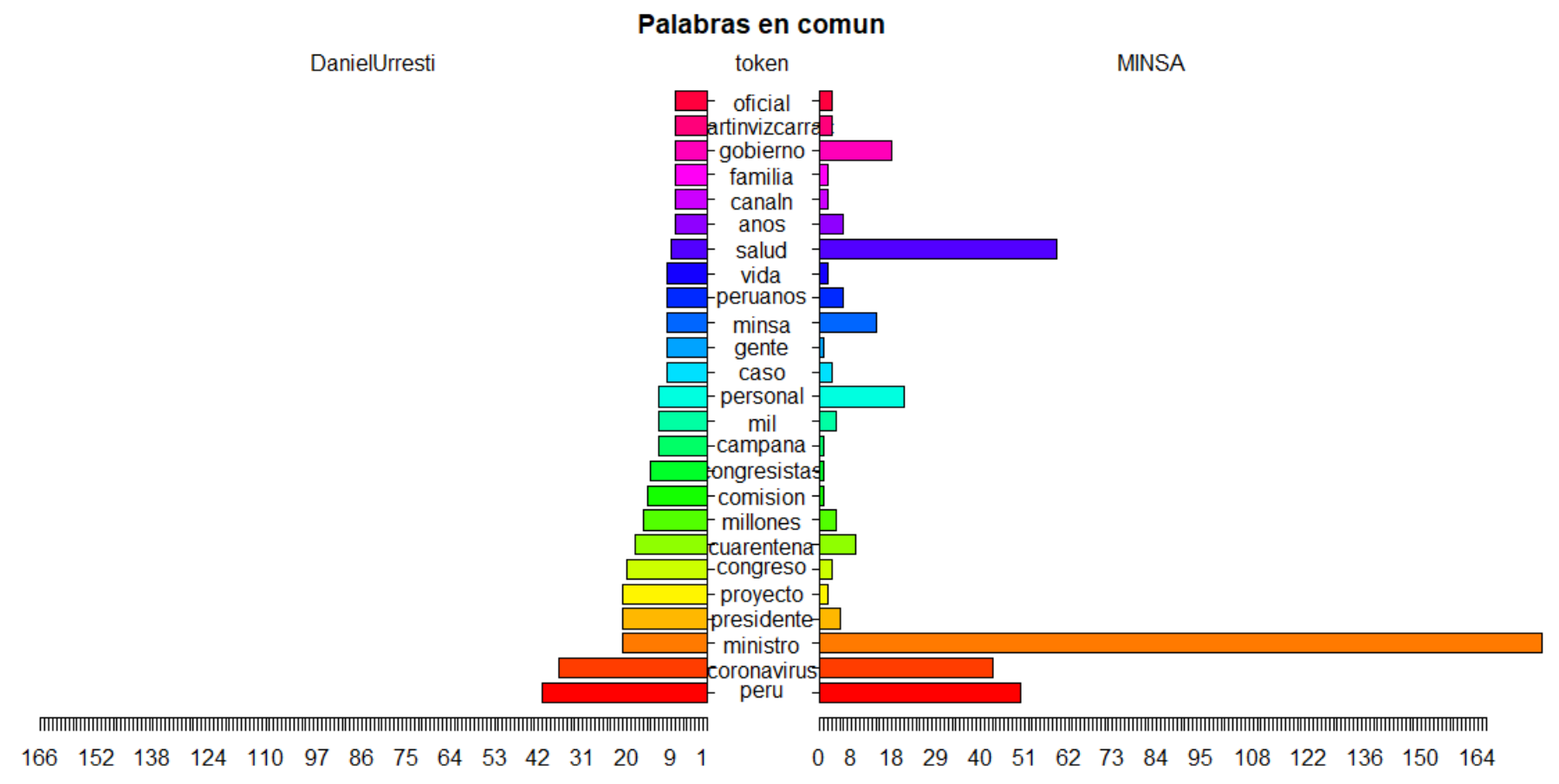


Las palabras más utilizadas por DanielUrresti1 son afp, peru, coronavirus, ley, refiriéndose al proyecto de les de retiro de la Afp. Las palabras más utilizadas por Minsa\_Peru son victorzamora, ministro, covid, salud, refiriéndose a las acciones tomadas para el manejo de la pandemia.

1. **Nube de palabras y palabras en común**

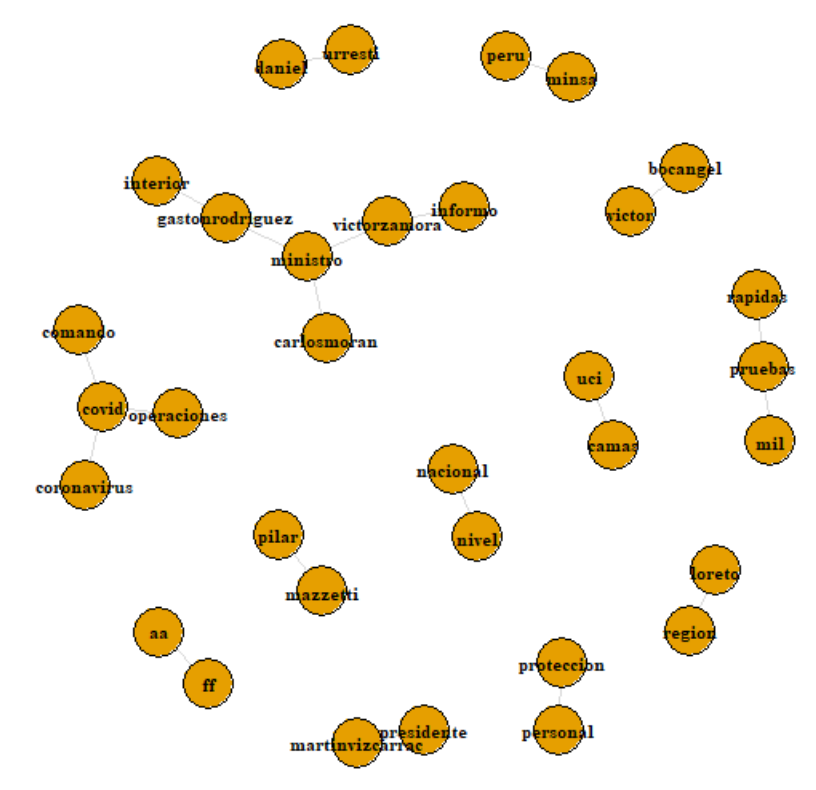


Las palabras con mayor frecuencia en el usuario DanielUrresti1 son afp, peru, ley, retiro, coronavirus, refiriéndose al proyecto de ley de retiro Afp, mientras que el usuario Minsa\_Peru usó con mayor frecuencia las palabras victorzamora, covid, ministro, refieriéndose a la gestión de la institución.



Las palabras en común entre los usuarios DanielUrresti1 y Minsa\_Peru son peru, coronavirus, ministro, cuarentena. Para el usuario Minsa\_Peru estas palabras refieren a la gestión de la institución durante el manejo de la pandemia, mientras que para el usuario DanielUrresti1 estas palabras pueden significar una observación o crítica a la gestión de Minsa\_Peru.

1. **Bigrama**



Se observan los Bigramas más utilizados en el documento.

1. **Modelos de Clasificación**

| **Modelo** | **Precision** | **Recall** | **F1 score** |
| --- | --- | --- | --- |
| Naive Bayes sin normalizar | 0.86440 | 0.68 | 0.76119 |
| SVM normalizado | 0.95 | 0.93827 | 0.94409 |
| Naive Bayes normalizado | 0.97419 | 0.93209 | 0.95268 |

De los modelos presentados el que presenta mejores indicadores en Precision y F1 score es el modelo Naive Bayes normalizado, mientras que el Recall del modelo SVM normalizado es ligeramente mayor al modelo Naive Bayes normalizado, debido a que no implica mucha diferencia en el Recall para este análisis se recomienda usar el modelo Naive Bayes normalizado.

El modelo que obtuvo los indicadores más bajos fue Naive Bayes sin normalizar, lo cual evidencia la importancia del preprocesamiento de datos, normalizar y remover stopwords antes de correr un modelo.