Adjuntamos, este archivo en script “--------------”

Para ver que análisis será mas efectivo, dentro de nuestro análisis, ya sea aprendizaje supervisado como no supervisado, usaremos un conjunto de datos revisados previamente con su polaridad, en este caso usamos 367 de acuerdo a la formula que hemos expuesto anteriormente.

Partimos con estos datos del conjunto de la muestra que hemos obtenido

|  |  |
| --- | --- |
| No solicitados | Solicitados |
| 0 | 1 |
| 289 | 78 |

**Para SupportVectorMachine (SVM)**

Procedemos a realizar el analisis para poder ver la matriz de confusión:

1. Leeremos todas las palabras que están dentro del tweet, las usamos como vectores y los mandamos en un corpus.

Recorte de pantalla

1. Cambiamos a minúsculas y lo depuramos, con signos de puntuación, stopwords, links etc

Recorte de pantalla

Recorte de pantalla

1. Usamos Steamming para reducir las palabras a su simplicidad, ya que nos permite tratar muchas mas palabras de una manera mas homogénea, con el paquete de tm proveniente de R

Recorte de pantalla

1. Ahora lo que trataremos de hacer es sacar esas palabras, una por cada columna y asi poder observar cuales son las mejores

Recorte de pantalla

1. Le pegamos a esta matriz la polaridad antes resuelta en la variable ya escogida

Recorte de pantalla

Con estos pasos pues esta listo para el análisis de datos

Como necesitamos ver como se comportan nuestros datos dentro del modelo SVM (Support Vector Machine), Definimos un conjunto de entrenamiento o conjunto de evaluación, dividiendo de esta manera un 80% para entrenamiento, y 20% evaluación.

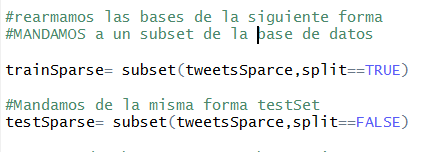
Cabe mencionar que con el entrenamiento vamos a enseñarle al modelo, vamos a enseñarle al algoritmo como clasificar, y con el 20% vamos a evaluar como lo hace el modelo, y con esto observar que poder de predicción tiene el modelo luego de haberlo creado.

Recorte de pantalla

Ahora procedemos a usar, una variable que me ayudara a ver que observaciones se van para el train(80%) y cuales para el test(20%), y diré que mi splitRatio sea 0.8 que quiere decir que el 80% de mis observaciones se vayan al conjunto de entrenamiento, asignando true o false según como convenga.

Recorte de pantalla

Rearmamos las bases tanto el **trainSparse** como el **testSparse**



Ya con estos datos procedemos a verificar con la matriz de confusión cual fue nuestro error, de predicción aplicamos el algoritmo SVM

Recorte de pantalla

Explicando que tomamos como un factor el sentimiento, y es representado por todas las columnas dentro del trainSparse con esto podemos observar que tenemos un total de 86% de exactitud

|  |  |
| --- | --- |
| No solicitado | Solicitado |
| 0 | 1 |
| 305 | 62 |

**Con esto podemos observar que tenemos 86% de exactitud.**

Recomendaciones

Tener en cuenta el enconding que usamos para ver los tweets en este caso es UTF-8,