

Clasificación de texto con regresión logística

Práctica IV

Especificaciones

- Cargue el dataset corpus *Rest_Mex_2022_Sentiment_Analysis_Track_Train.xlsx*
- Aplique el proceso de tokenización y lematización con la herramienta Freeling a las columnas *Title* y *Opinion*
- Genere una nueva versión del dataset con el texto tokenizado y lematizado, agregando a este las columnas Polarity y Attraction que son las clases a predecir (target)
- Genere un conjunto de **entrenamiento con el 80%** de los datos y uno de **prueba con el 20%**. Debe aplicar Shuffle con el parámetro **random_state=0** (esto asegura que todos trabaje con las mismas instancias en cada conjunto)
- Con el conjunto de entrenamiento genere un conjunto de **validación de 10 pliegues**

Especificaciones

- Utilizando el conjunto de validación (80% de los datos en 10 pliegues) realice los siguientes experimentos
 - Genere la representación del texto en espacio vectorial binarizado
 - Para cada pliegue
 - Entrene un modelo de regresión logística y **prediga la polaridad de opinión (1-5)**
 - Aplique las métricas accuracy(exactitud), precision, recall y F-measure
 - Entrene un modelo de regresión logística y **prediga el tipo de lugar (Hotel, Restaurant, Attraction)**
 - Aplique las métricas accuracy(exactitud), precision, recall y F-measure
 - Obtenga el accuracy, precision, recall y F-measure promedio para la predicción de polaridad de opinión
 - Obtenga el accuracy, precision, recall y F-measure promedio para la predicción de lugar
 - Genere la representación del texto en espacio vectorial de frecuencia
 - Repita los experimentos anteriores

Especificaciones

- Utilizando el conjunto de prueba (20% de los datos)
 - Utilice el modelo con la representación del texto que mejor accuracy promedio obtuvo en la predicción de polaridad y prediga la polaridad
 - Aplique las métricas accuracy(exactitud), precision, recall y F-measure
 - Utilice el modelo con la representación del texto que mejor accuracy promedio obtuvo en la predicción del lugar y prediga el lugar
 - Aplique las métricas accuracy(exactitud), precision, recall y F-measure

Evidencias

- Código fuente
- Documento con la siguiente información
 - Tabla de los experimentos sobre el corpus de entrenamiento

Predicción de polaridad				
<i>Representación</i>	<i>Accuracy promedio</i>	<i>Precision promedio</i>	<i>Recall promedio</i>	<i>F-measure promedio</i>
Binarizado				
Frecuencia				

Predicción de lugar				
<i>Representación</i>	<i>Accuracy promedio</i>	<i>Precision promedio</i>	<i>Recall promedio</i>	<i>F-measure promedio</i>
Binarizado				
Frecuencia				

- Tabla de resultados sobre el conjunto de prueba

Predicción de polaridad				
<i>Representación</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>
<Mejor representación>				

Predicción de lugar				
<i>Representación</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>
<Mejor representación>				