



TIC



▶ TALENTO TECH

Lección 3: Metodologías para análisis de sentimientos



Tiempo de ejecución: 4 horas

PLANTEAMIENTO DE LA SESIÓN

MATERIALES

El análisis de sentimientos se centra en determinar la actitud o emoción expresada en un texto, ya sea positiva, negativa o neutral. Existen varias metodologías y enfoques para realizar análisis de sentimientos, que van desde técnicas basadas en reglas hasta enfoques basados en aprendizaje automático. A continuación, se presentan algunas de las metodologías más comunes:

Análisis léxico basado en diccionarios: Esta metodología implica el uso de diccionarios léxicos que contienen palabras etiquetadas con su polaridad (positiva, negativa o neutral). El análisis consiste en asignar una puntuación de sentimiento a cada palabra en el texto y luego calcular una puntuación general de sentimiento para el documento basada en estas puntuaciones de palabras. Ejemplos de diccionarios léxicos incluyen el Sentiment140, SentiWordNet y VADER (Valence Aware Dictionary and sEntiment Reasoner).

Conexión a Internet.



Tiempo de ejecución: 4 horas



TIC

PLANTEAMIENTO DE LA SESIÓN

MATERIALES

Aprendizaje automático supervisado: Este enfoque implica el entrenamiento de modelos de aprendizaje automático utilizando datos etiquetados con la polaridad del sentimiento. Los algoritmos comúnmente utilizados incluyen máquinas de vectores de soporte (SVM), árboles de decisión, redes neuronales y modelos de clasificación basados en palabras clave. Los modelos se entrenan con un conjunto de datos de entrenamiento y luego se evalúan con un conjunto de datos de prueba para predecir la polaridad del sentimiento en nuevos textos.

Análisis de sentimientos basado en aspectos: Esta metodología se centra en identificar y analizar opiniones sobre aspectos específicos de un producto, servicio o evento. Implica la identificación de aspectos relevantes (entidades) en el texto y luego el análisis del sentimiento asociado con cada aspecto.

Tiempo de ejecución: 4 horas



TIC

| PLANTEAMIENTO DE LA SESIÓN | MATERIALES |
|---|------------|
| <p>Esta técnica es útil para comprender opiniones detalladas sobre diferentes aspectos de un tema, como la calidad, el precio o el servicio al cliente.</p> <p>Análisis de sentimientos basado en redes neuronales: Con los avances en el procesamiento de lenguaje natural y el aprendizaje profundo, las redes neuronales han demostrado ser efectivas para el análisis de sentimientos. Los enfoques basados en redes neuronales, como las redes neuronales recurrentes (RNN), las redes neuronales convolucionales (CNN) y las redes neuronales de transformadores (como BERT), han logrado resultados significativos en tareas de análisis de sentimientos debido a su capacidad para capturar el contexto y las relaciones semánticas en el texto.</p> | |





El análisis léxico es una metodología para el análisis de sentimientos que se basa en el uso de diccionarios léxicos, los cuales contienen palabras etiquetadas con su polaridad de sentimiento (positiva, negativa o neutral). Esta metodología es ampliamente utilizada debido a su simplicidad y eficiencia en términos de procesamiento computacional.

Introducción al análisis léxico: El análisis léxico se centra en analizar el contenido de un texto a través de la evaluación de la polaridad de las palabras individuales que lo componen. Esto se logra mediante el uso de un diccionario léxico que asigna una puntuación de sentimiento a cada palabra.



TIC





TIC



Descripción de diccionarios léxicos: Los diccionarios léxicos son conjuntos de palabras que han sido previamente etiquetadas con su polaridad de sentimiento. Estas etiquetas pueden ser valores numéricos (por ejemplo, -1 para negativo, 0 para neutral, 1 para positivo) o categorías (por ejemplo, 'positivo', 'negativo', 'neutral'). Ejemplos de diccionarios léxicos incluyen VADER, SentiWordNet y AFINN.

Proceso de asignación de puntuaciones de sentimiento: En el análisis léxico, cada palabra en el texto se busca en el diccionario léxico y se le asigna una puntuación de sentimiento según su entrada correspondiente en el diccionario. Estas puntuaciones pueden ser ponderadas por diversos factores, como la frecuencia de aparición de la palabra en el texto o la intensidad de su sentimiento.





TIC



Cálculo de la puntuación general de sentimiento: Una vez que se han asignado puntuaciones de sentimiento a todas las palabras del texto, se puede calcular una puntuación general de sentimiento sumando o promediando las puntuaciones individuales de las palabras. Esto proporciona una medida agregada de la polaridad del sentimiento en el texto.

Cargar un diccionario léxico y asignar puntuaciones de sentimientos a palabras:





```
# Ejemplo de carga de diccionario léxico (AFINN)
def cargar_diccionario_lexico(ruta):
    diccionario = {}
    with open(ruta, 'r') as archivo:
        for linea in archivo:
            palabra, puntuacion = linea.strip().split('\t')
            diccionario[palabra] = int(puntuacion)
    return diccionario
```



TIC





```
# Cargar diccionario léxico
ruta_diccionario = 'AFINN-111.txt' # Ejemplo de ruta del
diccionario AFINN
diccionario_lexico = cargar_diccionario_lexico(ruta_diccionario)

# Ejemplo de asignación de puntuaciones de sentimiento a
palabras
texto = "Este es un ejemplo de texto para análisis de
sentimientos."
palabras = texto.lower().split()
puntuaciones = [diccionario_lexico.get(palabra, 0) for palabra
in palabras]
print(puntuaciones)
```



TIC



Análisis de sentimientos en un texto utilizando un diccionario léxico

```
# Ejemplo de análisis de sentimientos utilizando un diccionario
léxico (AFINN)

def analisis_sentimientos(texto, diccionario):
    palabras = texto.lower().split()
    puntuaciones = [diccionario.get(palabra, 0) for palabra in
palabras]
    sentimiento = sum(puntuaciones)
    return sentimiento
```



TIC





```
# Cargar diccionario léxico
ruta_diccionario = 'AFINN-111.txt' # Ejemplo de ruta del
diccionario AFINN
diccionario_lexico = cargar_diccionario_lexico(ruta_diccionario)

# Ejemplo de análisis de sentimientos en un texto
texto = "Este es un ejemplo de texto para análisis de
sentimientos."
puntuacion_sentimiento = analisis_sentimientos(texto,
diccionario_lexico)
print("Puntuación de sentimiento:", puntuacion_sentimiento)
```



TIC





TIC



En el primer script, se asignan puntuaciones de sentimiento a cada palabra en el texto utilizando un diccionario léxico. Si una palabra no está presente en el diccionario, se le asigna una puntuación de sentimiento de 0. En el segundo script, se suma el total de las puntuaciones de sentimiento de todas las palabras en el texto para obtener una puntuación de sentimiento agregada para todo el texto.

En este caso particular, las puntuaciones de sentimiento para cada palabra en el texto fueron todas 0. Esto significa que ninguna de las palabras en el texto estaba presente en el diccionario léxico utilizado para el análisis de sentimientos. Como resultado, la puntuación de sentimiento agregada para el texto también fue 0.



TIC



El aprendizaje automático supervisado es una metodología que utiliza datos etiquetados para entrenar modelos que pueden realizar predicciones o clasificaciones en nuevos datos. En análisis de sentimientos, se utiliza para entrenar modelos que puedan identificar la polaridad del sentimiento en un texto dado.

Introducción al aprendizaje automático supervisado para análisis de sentimientos: Esta metodología implica el uso de algoritmos de aprendizaje automático supervisado, que aprenden patrones a partir de datos etiquetados para hacer predicciones sobre nuevos datos no etiquetados. En el análisis de sentimientos, se utilizan datos de entrenamiento que contienen textos etiquetados con su polaridad de sentimiento (positiva, negativa o neutral).



Descripción de algoritmos comunes: Los algoritmos comunes para el análisis de sentimientos incluyen máquinas de vectores de soporte (SVM), árboles de decisión, modelos de regresión logística, entre otros. Estos algoritmos aprenden a partir de ejemplos de entrenamiento para realizar predicciones sobre nuevos textos.

Proceso de preparación de datos etiquetados: El proceso de preparación de datos implica la recopilación y limpieza de datos etiquetados para el entrenamiento del modelo. Esto puede incluir la tokenización, lematización o eliminación de palabras de parada, así como la representación de los datos en una forma que el modelo pueda entender, como vectores de características.



TIC





TIC



Entrenamiento y evaluación de modelos de aprendizaje automático: Una vez que los datos están preparados, se utilizan para entrenar modelos de aprendizaje automático utilizando algoritmos supervisados. Luego, se evalúan los modelos utilizando un conjunto de datos de prueba para medir su rendimiento en la clasificación de sentimientos.





Preprocesar y preparar datos etiquetados para análisis de sentimientos:

```
import pandas as pd
import matplotlib.pyplot as plt

from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder

# Datos de ejemplo: textos etiquetados con polaridad de
sentimiento

textos = ['Este es un buen producto', 'No me gustó el servicio',
'La película fue excelente']

etiquetas = ['positivo', 'negativo', 'positivo']
```



TIC





```
# Crear un DataFrame para visualizar los datos
data = pd.DataFrame({'Texto': textos, 'Etiqueta': etiquetas})

# Mostrar los datos en forma de tabla
print("Datos etiquetados:")
print(data)
print()
```



TIC





```
# Visualizar la distribución de las etiquetas en un gráfico de  
barras
```

```
plt.figure(figsize=(6, 4))
```

```
data['Etiqueta'].value_counts().plot(kind='bar',
```

```
color='skyblue')
```

```
plt.title('Distribución de etiquetas')
```

```
plt.xlabel('Etiqueta')
```

```
plt.ylabel('Cantidad')
```

```
plt.xticks(rotation=0)
```

```
plt.show()
```



TIC





```
# Codificar etiquetas
```

```
encoder = LabelEncoder()
```

```
y = encoder.fit_transform(etiquetas)
```

```
# Representación vectorial de los textos
```

```
vectorizer = CountVectorizer()
```

```
X = vectorizer.fit_transform(textos)
```

```
# Dividir datos en conjuntos de entrenamiento y prueba
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y,
```

```
test_size=0.2, random_state=42)
```



TIC





Output

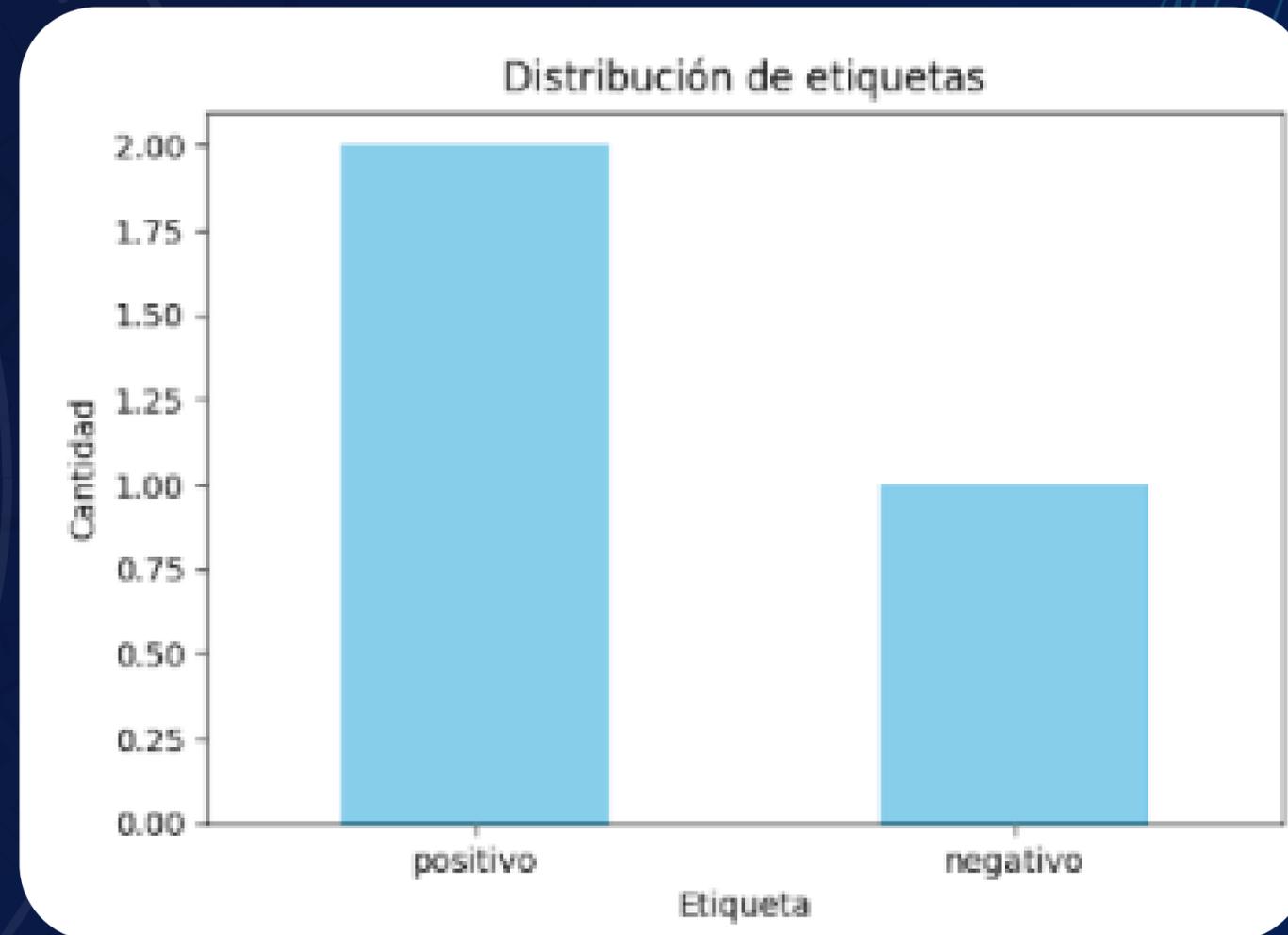
Datos etiquetados:

Texto Etiqueta

0 Este es un buen producto
positivo

1 No me gustó el servicio negativo

2 La película fue excelente positivo



TIC





Entrenar y evaluar un modelo de aprendizaje automático para análisis de sentimientos

```
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, classification_report,
confusion_matrix
import seaborn as sns

# Entrenar modelo SVM
modelo = SVC(kernel='linear')
modelo.fit(X_train, y_train)

# Realizar predicciones en el conjunto de prueba
y_pred = modelo.predict(X_test)
```



TIC





```
# Evaluar el rendimiento del modelo
accuracy = accuracy_score(y_test, y_pred)
print("Exactitud del modelo:", accuracy)
print()

# Mostrar informe de clasificación
print("Informe de clasificación:")
print(classification_report(y_test, y_pred, labels=[0, 1])) # Especificar
las etiquetas de las clases
print()
```



TIC





```
# Visualizar la matriz de confusión en un mapa de calor
conf_matrix = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(6, 4))
sns.heatmap(conf_matrix, annot=True, cmap='Blues', fmt='d',
            xticklabels=encoder.classes_, yticklabels=encoder.classes_)
plt.title('Matriz de Confusión')
plt.xlabel('Etiqueta Predicha')
plt.ylabel('Etiqueta Verdadera')
plt.show()
```



TIC





Output:

Exactitud del modelo: 1.0

Informe de clasificación:

precision recall f1-score support

0 0.00 0.00 0.00 0

1 1.00 1.00 1.00 1

micro avg 1.00 1.00 1.00 1

macro avg 0.50 0.50 0.50 1

weighted avg 1.00 1.00 1.00 1

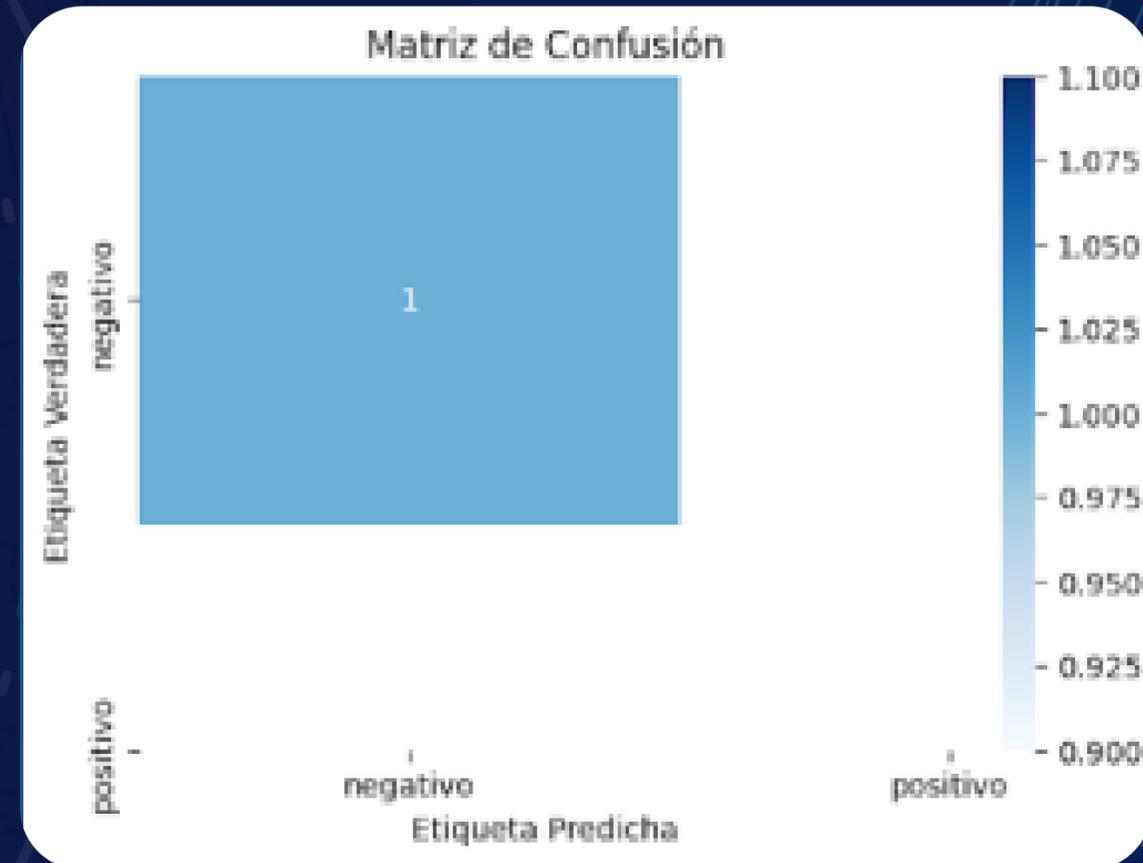


TIC





El primer script muestra cómo preparar los datos etiquetados y representar los textos como vectores de características. El segundo script entrena un modelo SVM utilizando los datos preparados y evalúa su rendimiento en un conjunto de datos de prueba.



TIC





TIC



El análisis de sentimientos basado en aspectos es una técnica avanzada que se centra en identificar aspectos específicos dentro de un texto y analizar el sentimiento asociado con cada uno de esos aspectos.

Introducción al análisis de sentimientos basado en aspectos: Esta técnica va más allá de simplemente clasificar un texto como positivo, negativo o neutral, y se enfoca en identificar aspectos específicos dentro del texto que generan sentimientos. Por ejemplo, en una revisión de un hotel, los aspectos podrían incluir la ubicación, la limpieza, el servicio, etc.





Proceso de identificación de aspectos relevantes en el texto: Para llevar a cabo este análisis, es necesario identificar los aspectos relevantes dentro del texto. Esto puede implicar técnicas de procesamiento de lenguaje natural (PLN) como el análisis sintáctico y la extracción de entidades nombradas para identificar sustantivos y frases relevantes que representen aspectos.

Métodos para analizar el sentimiento asociado con cada aspecto: Una vez que se han identificado los aspectos, se pueden aplicar métodos de análisis de sentimientos para determinar el sentimiento asociado con cada uno de ellos. Esto puede implicar el uso de diccionarios léxicos, aprendizaje automático supervisado o técnicas más avanzadas como el análisis de sentimientos basado en el contexto.



TIC





TIC



Aplicaciones y casos de uso en la industria: El análisis de sentimientos basado en aspectos tiene numerosas aplicaciones en la industria, como la gestión de la reputación en línea, la mejora de productos y servicios, la detección de problemas en tiempo real y la personalización de la experiencia del cliente.

Análisis de sentimientos basado en redes neuronales Actividad: Realice los scripts para procesar datos y construir una red neuronal para análisis de sentimientos. Además haga el script para entrenar y evaluar diferentes arquitecturas de redes neuronales para análisis de sentimientos.