



# PRIMER PORTAFOLIO

Katerine Lisbeth Rafael Bourdierd 2022-0088

## PRIMER ACTIVIDAD

### Preguntas sobre NLP

#### 1.- ¿Qué es el procesamiento del lenguaje natural?

**Respuesta:** Es el campo de conocimiento de la Inteligencia Artificial que se ocupa de la investigar la manera de comunicar las máquinas con las personas mediante el uso de lenguas naturales, como el español, el inglés o el chino.

#### 2.- ¿Cuáles son las aplicaciones del NLP?

**Respuesta:**

**Reconocimiento del habla:** es la habilidad de una máquina para interpretar frases o palabras de un lenguaje. Esta aplicación de NLP se puede encontrar en teléfonos móviles y hasta en casas inteligentes. Simplemente con decir "Llamar a John", el dispositivo reconoce lo que esta frase significa y automáticamente empieza a llamar a John. Aquí se puede mencionar el caso de Alexa o Siri que son bastante populares.

**Análisis de sentimientos:** es la determinación del tono emocional implícito en una frase. Por ejemplo, si un texto anuncia una caída de las acciones en la bolsa, el programa va a predecir que posiblemente es un texto negativo. Por otra parte, si un texto se refiere a una fiesta y todos están invitados a asistir, se puede decir que su significado es positivo.

**Sistemas de Preguntas y Respuestas:** consiste en responder de forma automática a preguntas por medio de un programa. Esta aplicación se puede encontrar fácilmente en chats de redes sociales, llamadas o herramientas como Siri o IBM Watson.

**Generación automática de resúmenes:** dado un texto, el programa va a obtener las ideas principales del mismo y producir un resumen coherente. Por ejemplo, si una persona tiene que leer Don Quijote, un programa puede extraer las ideas principales del libro.

**Chatbots:** son programas que entablan conversaciones con humanos. Por ejemplo: cuando una persona necesita comprar algo en una tienda virtual y tiene ciertas preguntas sobre un producto, es muy probable que esta reciba respuestas automáticas generadas por una máquina.

**Inteligencia de mercado:** con base a lo que una persona ha buscado en internet, automáticamente se empieza a buscar anuncios relacionados. Un buen ejemplo sería cuando se busca un producto específico y automáticamente en redes sociales nos aparecen anuncios relacionados con dicho producto.

**Clasificación automática de textos:** se define como la asignación de una etiqueta a un texto según su contenido y semántica. Por ejemplo, cuando se recibe un correo se puede decir si es spam o no basado en su contenido.

**Revisión automática de gramática:** el computador reconoce los diferentes errores gramaticales o de ortografía de un texto según el contexto.

### 3.- ¿Cómo converge el NLP con el Machine Learning?

**Respuesta:** Mediante el término genérico de Inteligencia Artificial se hace referencia al proceso mediante el cual, la inteligencia humana se intenta replicar en las máquinas. De este modo, se proporciona a las mismas la habilidad de poder responder como los humanos, permitiendo llevar a cabo un proceso similar de aprendizaje y de resolución de problemas. Por lo tanto, se trabaja para que las máquinas puedan llegar a tomar decisiones inteligentes, basándose en su propia experiencia y el aprendizaje que deriva de la misma.

Como se ha comentado, dentro del campo de la Inteligencia Artificial, encontramos el Machine Learning, que ha revolucionado el concepto de aprendizaje automático. Mediante esta disciplina, se trabaja para construir algoritmos que sean capaces de aprender de los datos disponibles y explorar su capacidad para llevar a cabo todo tipo de predicciones. Por último, y con el objetivo de humanizar todo este proceso, nace el NLP (Natural Language Processing o Procesamiento del Lenguaje Natural), que intenta comprender los comandos o instrucciones que se le dan a una máquina sin necesidad de utilizar complejos códigos o programación, sino utilizando el lenguaje normal o natural.

Esta tecnología resulta especialmente disruptiva si se tiene en cuenta que antes éramos los humanos los que debíamos entender y adaptarnos al lenguaje de la máquina, mientras que ahora es la máquina la que debe entendernos a nosotros y nuestras maneras de interactuar con ella.

## SEGUNDA ACTIVIDAD

Preguntas sobre metodologías de preprocesamiento de texto

### 1.- ¿Qué es una palabra?

**Respuesta:** Es una unidad léxica formada por un sonido o un conjunto de sonidos articulados, que se asocia a uno o varios sentidos, y que posee una categoría gramatical determinada.

### 2.- ¿Qué es una oración?

**Respuesta:** Es un grupo de palabras que forman una unidad sintáctica con sentido completo, un conjunto lingüístico que representa la unidad mínima de expresión.

### 3.- ¿Qué es un token en NLP?

**Respuesta:** Los tokens son cosas diferentes dependiendo del tipo de tokenizador que se utilice. Un token puede ser una palabra, un carácter o una subpalabra (por ejemplo, en la palabra inglesa "higher", hay dos subpalabras: "high" y "er"). Los signos de puntuación como "!", ".", y ";", también pueden ser tokens.

### 4.- ¿Qué es tokenización? ¿Cuáles son los diferentes tipos de tokenización?

**Respuesta:** Separar palabras del texto en entidades llamadas tokens, con las que trabajaremos luego. Debemos pensar si utilizaremos los signos de puntuación como token, si daremos importancia o no a las mayúsculas y si unificamos palabras similares en un mismo token.

### Tokenización de plataforma:

La tokenización de las infraestructuras de blockchain que permiten el desarrollo de aplicaciones descentralizadas se conoce como tokenización de plataforma. DAI, que puede ayudar con las transacciones de contratos inteligentes, es uno de los ejemplos más conocidos de tokenización de plataforma. La tokenización de la plataforma utiliza la red blockchain que sirve como base para una mayor seguridad y soporte transaccional.

### Tokenización en blockchain de utilidad:

La tokenización de utilidad es el proceso de generar tokens de utilidad en un protocolo para usar los servicios del protocolo. Es crucial resaltar que la tokenización de servicios públicos no implica la creación de tokens de inversión. Los tokens de utilidad proporcionan la actividad de plataforma requerida para que crezca la economía de la plataforma, mientras que la plataforma garantiza la seguridad de los tokens.

### Tokenización de la gobernanza:

El auge de los protocolos descentralizados ha requerido la adición de otro tipo de tokenización importante para blockchain. Los sistemas de votación basados en blockchain son el foco de la tokenización de la gobernanza porque pueden mejorar el proceso de toma de decisiones en torno a los protocolos descentralizados. La utilidad de la tokenización de la gobernanza se puede ver en el valor de la gobernanza en cadena, que permite a todas las partes interesadas colaborar, debatir y votar sobre la administración de un sistema.

## 5.- ¿Qué es stemming o derivación?

**Respuesta:** Stemming es el proceso de producir variantes morfológicas de una raíz/palabra base. Los programas de derivación se conocen comúnmente como algoritmos de derivación o lematizadores. Un algoritmo de derivación reduce las

palabras «chocolates», «chocolatey», «choco» a la raíz de la palabra, «chocolate» y «recuperación», «recuperado», «recuperado» se reducen a la raíz «recuperar».

## 6.- ¿Qué es lematización?

**Respuesta:** Los lematizadores son programas simples que intentan adivinar la raíz de una palabra, que puede ser sumamente útil para la agrupación en clúster. Por ejemplo, el lematizador de Porter, un algoritmo de lematización muy conocido, encontrará que "universidad" y "universidades" comparten la misma raíz y por lo tanto pertenecen a la misma clase de raíz. Todas las personalizaciones más avanzadas se basan en clases de raíz. Desafortunadamente, estos programas de lematización simples no son perfectos.

## 7.- ¿Cuál es la diferencia entre stemming y lematización?

**Respuesta:** La lematización es un término que se escucha con frecuencia al hablar del stemming y, aunque son métodos muy similares, existen diferencias importantes entre ambos.

Por un lado, el stemming elimina sufijos y prefijos de las palabras, dejando un stem como "cicl", en lugar de "ciclista", o "escri", en vez de "escribir".

Por otra parte, la lematización utiliza diccionarios para encontrar la base de la palabra, lo que también se conoce como lema de las formas flexionadas. Según lo dicho, el resultado sería que palabras como ciclista y ciclismo, se acorten a, por ejemplo, "ciclo" (su base sin conjugaciones o paradigmas) y, aunque este método es más complejo y amerita de mayores recursos informáticos, puede ofrecer mejores resultados, debido a que todas las variaciones de una palabra se agrupan en un término base.

## 8.- ¿Qué son unigramas, bigramas, trigramas y n-gramas?

**Respuesta:** N-grama. En Inteligencia Artificial, procesamiento de lenguaje natural, Bioinformática, recuperación de información se llama n-grama a una subsecuencia

de  $n$  elementos consecutivos en una secuencia dada. Si  $n=2$  se denominan bigramas;  $n=3$ , trigramas; para  $n \geq 4$  entonces se llaman genéricamente  $n$ -gramas o Modelos de Markov de orden  $(n-1)$ .

Los  $n$ -gramas son de gran utilidad en el tratamiento de textos, la determinación del lenguaje de documentos, el procesamiento estadístico de los mismos o el descubrimiento o predicción de genes en tanto estos son subsecuencias dadas dentro del material genético.

### **9.- ¿En qué consiste la vectorización de texto y cuáles métodos de vectorización de texto existen?**

**Respuesta:** La vectorización consiste en convertir imágenes que están formadas por píxeles en imágenes formadas por vectores, esto incluye texto, símbolos y otros detalles.

### **10.- ¿Qué es el vectorizador por frecuencia binaria?**

Respuesta:

### **11.- ¿Qué es el vectorizador por TF-IDF?**

**Respuesta:**

Básicamente, tf-idf (Term Frequency - Inverse Document Frequency) es un dato estadístico que muestra el grado de frecuencia de una palabra en un corpus especializado, no genérico, de documentos.

### **12.- ¿Qué es web scrapping y para qué sirve?**

**Respuesta:**

El web scraping es una técnica que se utiliza para la extracción y almacenamiento de información de cualquier página web a través de un programa de software, que suele ser un crawler. Estos bots no siempre son de Google, quien rastrea un sitio con el fin de indexar las URL en sus SERP.

### **13.- ¿Cuáles son las librerías más importantes para scrapping web de Python?**

**Respuesta:**

#### **Pyspider**

Empecemos por Pyspider, una herramienta para web-crawler con una interface de usuario basada en la web que facilita el seguimiento de múltiples operaciones. Es una opción extensible, con varias bases de datos de back-end y colas de mensajes compatibles, y varias funciones útiles integradas, desde la priorización hasta la capacidad de reintentar páginas fallidas, el rastreo de páginas por antigüedad y otros. Pyspider trabaja con Python 2 y Python 3, y para realizar un crawling más rápido, puede usarlo en un formato distribuido con múltiples crawlers funcionando al mismo tiempo.

#### **MechanicalSoup**

MechanicalSoup es una librería para crawling desarrollada en torno a la librería de análisis de HTML BeautifulSoup. Si las necesidades de crawling son simples, pero es necesario que ingrese información en algunas casillas y no desea construir su propio rastreador para esta tarea, es una buena opción a ser considerada.

#### **Scrapy**

Scrapy es un framework para scraping soportado por una comunidad muy activa con la cual puede construir su propia herramienta. Además de las herramientas de scraping, se puede exportar la data recopilada en varios formatos, como JSON o CSV, y almacenar los datos en un backend de su elección. También tiene una serie de extensiones integradas para tareas como el manejo de cookies, suplantación de user-agent, restricción de la profundidad del rastreo, entre otras, así como un API para ampliar fácilmente las funcionalidades.



## TERCERA ACTIVIDAD

### Ejercicios de vectorización de texto

Dada la siguiente tabla aplique tokenización y vectorización de texto con fines de entrenamiento posterior para un modelo de NLP.

Comentario	Puntuación	Sentimiento
La mejor pizza que he comido en años. Literalmente sin desperdicios, excelente sabor y textura. 100% recomendado.	5	Positivo
Siento que la salsa de tomate estaba muy dulce. Terrible combinación con queso de cabra, habría preferido queso mozzarella normal.	2	Negativo
Menos es más. Fue tan terrible que solo me comí la punta de un pedazo. Demasiados ingredientes, es como si fuera para engordar animales.	1	Negativo
La idea está bien, pero la ejecución no fue la mejor. Lo que hace que esta pizza no termine de ser terrible son los embutidos. La alta calidad del pepperoni para mí la salvó, lo demás, a la basura.	3	Negativo
Me encantan las pizzas como estas. Esta incorporación de sabores exóticos fue increíble. De verdad es una experiencia que todos deben degustar.	5	Positivo
Excelente pizza, nunca había probado dátiles en una pizza y me sorprendió que combinara tan bien con el queso de cabra. Será una de las mejores pizzas en mi top.	4	Positivo
La compré con miedo porque había leído muchas reviews malas, pero al probarla quedé enamorado. ¡Qué delicia de pizza, cuánto sabor!	4	Positivo
Horrible el queso de cabra y no pega para nada con los dátiles. Siento que quisieron hacerla muy compleja y al final terminó siendo un desastre.	2	Negativo
La pizza definitivamente es deliciosa, una de las mejores que he probado. La textura de la arena fue delicada y crujiente, lo único es que siento que son muy caras para el producto que ofrecen.	4	Positivo
Me gustó mucho el sabor. Ya quiero que sea viernes otra vez para volver.	5	Positivo
¡Qué deliciosa pizza! Si no estuviera a dieta me verían todos los días aquí.	5	Positivo
Esta pizza es una basura, un desastre total. No la recomiendo, mejor vayan a otro sitio. Para colmo carísimas.	1	Negativo

### + PARTE I

#### Tokenización

Convierta cada una de estas oraciones en tokens y aplique los siguientes pasos:

1. Use el método de "Separación por Espacios en Blanco" para la obtención de los tokens crudos.

Tokens crudos !:

La mejor pizza que he comido en años . Literalmente sin desperdicios , excelente sabor y textura . 100 % recomendado .

Tokens crudos 2:

Siento que la salsa de tomate estaba muy dulce . Terrible combinación con queso de cabra , habría preferido queso mozzarella normal .

Tokens crudos 3:

Menos es más . Fue tan terrible que solo me comí la punta de un pedazo . Demasiados ingredientes , es como si fuera para engordar animales .

Tokens crudos 4:

La idea está bien , pero la ejecución no fue la mejor . Lo que hace que esta pizza no termine de ser terrible son los embutidos . La alta de calidad del pepperoni para mí la salvó , lo demás , a la basura .

Tokens crudos 5:

Me encantan las pizzas como estas . Esta incorporación de sabores exóticos fue increíble . De verdad es una experiencia que todos deben degustar .

Tokens crudos 6:

Excelente pizza , nunca había probado dátiles en una pizza y me sorprendió que combinara tan bien con el queso de cabra . Será una de las mejores pizzas en mi top .

Tokens crudos 7:

La compré con miedo porque había leído muchas reviews malas, pero al probarla quedé enamorado. ¡Qué delicia de pizza, cuánto sabor!

Tokens crudos 8:

Horrible el queso de cabra y no pega para nada con los dátiles. Siento que quisieron hacerla muy compleja y al final terminó siendo un desastre.

Tokens crudos 9:

La pizza definitivamente es deliciosa, una de las mejores que he probado. La textura de la arena fue delicada y crujiente, lo único es que siento que son muy caras para el producto que ofrecen.

Tokens crudos 10:

Me gustó mucho el sabor. Ya quiero que sea viernes otra vez para volver

Tokens crudos 11:

¡Qué deliciosa pizza! Si no estuviera a dieta me verían todos los días aquí.

Tokens crudos 12:

Esta pizza es una basura, un desastre total. No la recomiendo, mejor vayan a otro sitio. Para colmo carísimas.

## 2. Elimine los tokens que sean números, signos de puntuación y stopwords.

Tokens crudos 1:

mejor pizza que comido años Literalmente desperdicios excelente sabor y textura recomendado

Tokens crudos 2:

Siento que salsa de tomate dulce Terrible combinación queso de  
cabra habría preferido queso mozzarella normal

Tokens crudos 3:

Menos más tan terrible que me comí punta de pedazo Demasiados  
ingredientes fuera engordar animales

Tokens crudos 4:

idea ejecución no mejor que que pizza no termine de terrible son  
embutidos alta de calidad del pepperoni mí salvó demás basura

Tokens crudos 5:

Me encantan pizzas estas incorporación de sabores exóticos increíble De  
verdad experiencia que todos deben degustar

Tokens crudos 6:

Excelente pizza nunca había probado dátiles pizza y me sorprendió que  
combinara tan bien queso de cabra Será de mejores pizzas mi top

Tokens crudos 7:

compré miedo había leído muchas reviews malas al probarla quedé  
enamorado Qué delicia de pizza cuánto sabor

Tokens crudos 8:

Horrible queso de cabra y no pega nada dátiles Siento que quisieron  
hacerla compleja y al final terminó desastre

Tokens crudos 9:

pizza definitivamente deliciosa de mejores que probado textura de arena  
delicada y crujiente único que siento que son caras producto que ofrecen

Tokens crudos 10:

Me gustó mucho sabor Ya quiero que sea viernes otra vez volver

Tokens crudos 11:

Qué deliciosa pizza no estuviera dieta me verían todos días

Tokens crudos 12:

pizza basura desastre total No recomiendo mejor vayan otro sitio colmo  
carísimas

### 3. Aplique lematización a todos los tokens.

Tokens crudos 1:

mejor pizza comer años Literalmente desperdicio excelente sabor  
textura recomendar

Tokens crudos 2:

Sentir salsa tomate dulce Terrible combinar queso cabra haber  
preferir queso mozzarella normal

Tokens crudos 3:

Menos más tan terrible comer punta pedazo Demasiado ingrediente  
engordar animal

Tokens crudos 4:

idea ejecutar no mejor pizza no terminar de terrible ser embutido alta  
de calidad del pepperoni salvó demás basura

Tokens crudos 5:

encantar pizza esta incorporación sabor exóticos increíble verdad  
experiencia todos deber degustar

Tokens crudos 6:

Excelente pizza nunca haber probar dátil pizza sorprender combinar tan  
bien queso cabra Ser mejor pizzas top

Tokens crudos 7:

compra miedo haber leer mucha review mala probar quedé enamorar  
Qué delicia pizza cuánto sabor

Tokens crudos 8:

Horrible queso cabra no pega nada dátil Sentir quieren hacer compleja  
final terminar desastre

Tokens crudos 9:

pizza definitiva delicia mejor probar textura arena delicada crujiente  
único sentir son caras producto ofrecer

Tokens crudos 10:

gustar mucho sabor Ya querer ser viernes otro vez volver

Tokens crudos 11:

Qué delicia pizza no estar dieta ver todos días

Tokens crudos 12:

pizza basura desastre total No recomendar mejor vayan otro sitio  
colmo caro

**4. Después de aplicar el paso anterior elimine todos los tokens duplicados. Este será su vocabulario.**

mejor pizza comer años Literalmente desperdicio excelente sabor  
textura recomendar Sentir salsa tomate dulce cabra haber  
preferir queso mozzarella normal Menos más tan terrible punta  
pedazo Demasiado ingrediente engordar animal idea ejecutar no  
terminar ser embutido alta calidad pepperoni salvó demás basura  
encantar estar incorporación exóticos increíble verdad experiencia  
todos deber degustar nunca probar sorprender combinar bien top  
compra miedo leer mucha review mala quedar enamorar delicia cuánto  
Horrible pega nada dátil quieren hacer compleja final desastre definitiva  
arena delicada crujiente único caras producto ofrecer gustar mucho  
querer viernes otra vez volver dieta ver días total vayan otro sitio  
colmo carísima

+ PARTE 2

Vectorización

Utilizando el vocabulario del paso anterior aplique una vectorización por los siguientes métodos a cada comentario tokenizado. Esto debe convertir su tabla en una matriz numérica.

### ■ Binary Term Frequency

	mejor	pizza	comer	años	Literal	Desperdi..	Excele..	sabor
D1	1	1	1	1	1	1	1	1
D2	0	0	0	0	0	0	0	0
D3	0	0	1	0	0	0	0	0
D4	1	1	0	0	0	0	0	0
D5	0	1	0	0	0	0	0	1
D6	0	1	0	0	0	0	1	0
D7	0	1	0	0	0	0	0	1
D8	0	0	0	0	0	0	0	0
D9	0	1	0	0	0	0	0	0
D10	0	0	0	0	0	0	0	1
D11	0	1	0	0	0	0	0	0
D12	1	1	0	0	0	0	0	0

	textura	recomendar	sentir	salsa	tomate	dulce	cabra	haber
D1	1	1	0	0	0	0	0	1
D2	0	0	1	1	1	1	1	1
D3	0	0	0	0	0	0	0	0
D4	0	0	0	0	0	0	0	0
D5	0	0	0	0	0	0	0	0
D6	0	0	0	0	0	0	1	1
D7	0	0	0	0	0	0	0	1
D8	0	0	1	0	0	0	1	0
D9	1	0	1	0	0	0	0	1
D10	0	0	0	0	0	0	0	0
D11	0	0	0	0	0	0	0	0
D12	0	1	0	0	0	0	0	0



	preferir	queso	mozzarella	normal	meno	mas	tan	terrible	punta
D1	0	0	0	0	0	0	0	0	0
D2	1	1	1	1	0	0	0	1	0
D3	0	0	0	0	1	1	1	1	1
D4	0	0	0	0	0	0	0	1	0
D5	0	0	0	0	0	0	0	0	0
D6	0	1	0	0	0	0	1	0	0
D7	0	0	0	0	0	0	0	0	0
D8	0	1	0	0	0	0	0	0	0
D9	0	0	0	0	0	0	0	0	0
D10	0	0	0	0	0	0	0	0	0
D11	0	0	0	0	0	0	0	0	0
D12	0	0	0	0	0	0	0	0	0

	pedazo	demasiado	ingrediente	engordar	animal	idea	no	ejecutar	ser
D1	0	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0	0	0
D3	1	1	1	1	1	0	0	0	1
D4	0	0	0	0	0	1	1	1	1
D5	0	0	0	0	0	0	0	0	1
D6	0	0	0	0	0	0	0	0	0
D7	0	0	0	0	0	0	0	0	0
D8	0	0	0	0	0	0	1	0	1
D9	0	0	0	0	0	0	0	0	1
D10	0	0	0	0	0	0	0	0	1
D11	0	0	0	0	0	0	1	0	0
D12	0	0	0	0	0	0	1	0	1

	terminar	embutido	alta	calidad	pepperoni	salvo	demás	basura
D1	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0	0
D3	1	0	0	0	0	0	0	0
D4	0	1	1	1	1	1	1	1
D5	0	0	0	0	0	0	0	0
D6	0	0	0	0	0	0	0	0
D7	0	0	0	0	0	0	0	0
D8	1	0	0	0	0	0	0	0
D9	0	0	0	0	0	0	0	0
D10	0	0	0	0	0	0	0	0
D11	0	0	0	0	0	0	0	0
D12	0	0	0	0	0	0	0	1

	encantar	estar	incorporación	exótico	increíble	verdad	experiencia	todos
D1	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0	0
D3	0	0	0	0	0	0	0	0
D4	0	0	0	0	0	0	0	0
D5	1	0	1	1	1	1	1	1
D6	0	0	0	0	0	0	0	0
D7	0	0	0	0	0	0	0	0
D8	0	0	0	0	0	0	0	0
D9	0	0	0	0	0	0	0	0
D10	0	0	0	0	0	0	0	0
D11	0	1	0	0	0	0	0	0
D12	0	0	0	0	0	0	0	0

	deber	degustar	nunca	probar	dátil	sorprender	combinar	bien	top
D1	0	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	1	0	0
D3	0	0	0	0	0	0	0	0	0
D4	0	0	0	0	0	0	0	1	0
D5	1	1	0	0	0	0	0	0	0

D6	0	0	1	1	1	1	1	1	1
D7	0	0	0	0	0	0	0	0	0
D8	0	0	0	0	1	0	0	0	0
D9	0	0	0	1	0	0	0	0	0
D10	0	0	0	0	0	0	0	0	0
D11	0	0	0	0	0	0	0	0	0
D12	0	0	0	0	0	0	0	0	0

	compra	miedo	enamorar	leer	mucha	review	mala	quedar	delicia
D1	0	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0	0	0
D3	0	0	0	0	0	0	0	0	0
D4	0	0	0	0	0	0	0	0	0
D5	0	0	0	0	0	0	0	0	0
D6	0	0	0	0	0	0	0	0	0
D7	1	1	1	1	1	1	1	1	1
D8	0	0	0	0	0	0	0	0	0
D9	0	0	0	0	0	0	0	0	1
D10	0	0	0	0	0	0	0	0	0
D11	0	0	0	0	0	0	0	0	1
D12	0	0	0	0	0	0	0	0	0

	cuanto	horrible	pegar	Querer	hacer	compleja	final	desastre	arena
D1	0	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0	0	0
D3	0	0	0	0	0	0	0	0	0
D4	0	0	0	0	0	0	0	0	0
D5	0	0	0	0	0	0	0	0	0
D6	0	0	0	0	0	0	0	0	0
D7	1	0	0	0	0	0	0	0	0
D8	0	1	1	1	1	1	1	1	0
D9	0	0	0	0	0	0	0	0	1
D10	0	0	0	1	0	0	0	0	0
D11	0	0	0	0	0	0	0	0	0

D12	0	0	0	0	0	0	0	1	0
-----	---	---	---	---	---	---	---	---	---

	definitiva	delicada	crujiente	único	Caro	producto	ofrecer
D1	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0
D3	0	0	0	0	0	0	0
D4	0	0	0	0	0	0	0
D5	0	0	0	0	0	0	0
D6	0	0	0	0	0	0	0
D7	0	0	0	0	0	0	0
D8	0	0	0	0	0	0	0
D9	1	1	1	1	1	1	1
D10	0	0	0	0	0	0	0
D11	0	0	0	0	0	0	0
D12	0	0	1	0	1	0	0

	gustar	mucho	viernes	otro	vez	volver	dieta
D1	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0
D3	0	0	0	0	0	0	0
D4	0	0	0	0	0	0	0
D5	0	0	0	0	0	0	0
D6	0	0	0	0	0	0	0
D7	0	1	0	0	0	0	0
D8	0	0	0	0	0	0	0
D9	0	0	0	0	0	0	0
D10	1	1	1	1	1	1	0
D11	0	0	0	0	0	0	1
D12	0	0	0	1	0	0	0

	ver	día	Total	vaya	sitio	colmo
D1	0	0	0	0	0	0
D2	0	0	0	0	0	0
D3	0	0	0	0	0	0

D4	0	0	0	0	0	0
D5	0	0	0	0	0	0
D6	0	0	0	0	0	0
D7	0	0	0	0	0	0
D8	0	0	0	0	0	0
D9	0	0	0	0	0	0
D10	0	0	0	0	0	0
D11	1	1	0	0	0	0
D12	0	0	1	1	1	1

### ■ Bag of Words

	mejor	pizza	comer	años	Literal	Desperdi..	Excele..	sabor
D1	1	1	1	1	1	1	1	1
D2	0	0	0	0	0	0	0	0
D3	0	0	1	0	0	0	0	0
D4	1	1	0	0	0	0	0	0
D5	0	1	0	0	0	0	0	1
D6	0	3	0	0	0	0	1	0
D7	0	1	0	0	0	0	0	1
D8	0	0	0	0	0	0	0	0
D9	0	1	0	0	0	0	0	0
D10	0	0	0	0	0	0	0	1
D11	0	1	0	0	0	0	0	0
D12	1	1	0	0	0	0	0	0

	<b>textura</b>	<b>recomendar</b>	<b>sentir</b>	<b>salsa</b>	<b>tomate</b>	<b>dulce</b>	<b>cabra</b>	<b>haber</b>
<b>D1</b>	1	1	0	0	0	0	0	1
<b>D2</b>	0	0	1	1	1	1	1	1
<b>D3</b>	0	0	0	0	0	0	0	0
<b>D4</b>	0	0	0	0	0	0	0	0
<b>D5</b>	0	0	0	0	0	0	0	0
<b>D6</b>	0	0	0	0	0	0	1	1
<b>D7</b>	0	0	0	0	0	0	0	1
<b>D8</b>	0	0	1	0	0	0	1	0
<b>D9</b>	1	0	1	0	0	0	0	1
<b>D10</b>	0	0	0	0	0	0	0	0
<b>D11</b>	0	0	0	0	0	0	0	0
<b>D12</b>	0	1	0	0	0	0	0	0

	<b>preferir</b>	<b>queso</b>	<b>mozzarella</b>	<b>normal</b>	<b>meno</b>	<b>mas</b>	<b>tan</b>	<b>terrible</b>	<b>punta</b>
<b>D1</b>	0	0	0	0	0	0	0	0	0
<b>D2</b>	1	1	1	1	0	0	0	1	0
<b>D3</b>	0	0	0	0	1	1	1	1	1
<b>D4</b>	0	0	0	0	0	0	0	1	0
<b>D5</b>	0	0	0	0	0	0	0	0	0
<b>D6</b>	0	1	0	0	0	0	1	0	0
<b>D7</b>	0	0	0	0	0	0	0	0	0
<b>D8</b>	0	1	0	0	0	0	0	0	0

D9	0	0	0	0	0	0	0	0	0
D10	0	0	0	0	0	0	0	0	0
D11	0	0	0	0	0	0	0	0	0
D12	0	0	0	0	0	0	0	0	0

	pedazo	demasiado	ingrediente	engordar	animal	idea	no	ejecutar	ser
D1	0	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0	0	0
D3	1	1	1	1	1	0	0	0	2
D4	0	0	0	0	0	1	2	1	2
D5	0	0	0	0	0	0	0	0	1
D6	0	0	0	0	0	0	0	0	0
D7	0	0	0	0	0	0	0	0	0
D8	0	0	0	0	0	0	1	0	1
D9	0	0	0	0	0	0	0	0	2
D10	0	0	0	0	0	0	0	0	1
D11	0	0	0	0	0	0	1	0	0
D12	0	0	0	0	0	0	1	0	1

	terminar	embutido	alta	calidad	pepperoni	salvo	demás	basura
D1	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0	0
D3	1	0	0	0	0	0	0	0
D4	0	1	1	1	1	1	1	1
D5	0	0	0	0	0	0	0	0
D6	0	0	0	0	0	0	0	0
D7	0	0	0	0	0	0	0	0
D8	1	0	0	0	0	0	0	0
D9	0	0	0	0	0	0	0	0
D10	0	0	0	0	0	0	0	0
D11	0	0	0	0	0	0	0	0
D12	0	0	0	0	0	0	0	1

	encantar	estar	incorporación	exótico	increíble	verdad	experiencia	todos
D1	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0	0
D3	0	0	0	0	0	0	0	0
D4	0	0	0	0	0	0	0	0
D5	1	0	1	1	1	1	1	1
D6	0	0	0	0	0	0	0	0
D7	0	0	0	0	0	0	0	0
D8	0	0	0	0	0	0	0	0



D9	0	0	0	0	0	0	0	0
D10	0	0	0	0	0	0	0	0
D11	0	1	0	0	0	0	0	0
D12	0	0	0	0	0	0	0	0

	deber	degustar	nunca	probar	dátil	sorprender	combinar	bien	top
D1	0	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	1	0	0
D3	0	0	0	0	0	0	0	0	0
D4	0	0	0	0	0	0	0	1	0
D5	1	1	0	0	0	0	0	0	0
D6	0	0	1	1	1	1	1	1	1
D7	0	0	0	0	0	0	0	0	0
D8	0	0	0	0	1	0	0	0	0
D9	0	0	0	1	0	0	0	0	0
D10	0	0	0	0	0	0	0	0	0
D11	0	0	0	0	0	0	0	0	0
D12	0	0	0	0	0	0	0	0	0

	compra	miedo	enamorar	leer	mucha	review	mala	quedar	delicia
D1	0	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0	0	0
D3	0	0	0	0	0	0	0	0	0

D4	0	0	0	0	0	0	0	0	0
D5	0	0	0	0	0	0	0	0	0
D6	0	0	0	0	0	0	0	0	0
D7	1	1	1	1	1	1	1	1	1
D8	0	0	0	0	0	0	0	0	0
D9	0	0	0	0	0	0	0	0	1
D10	0	0	0	0	0	0	0	0	0
D11	0	0	0	0	0	0	0	0	1
D12	0	0	0	0	0	0	0	0	0

	cuanto	horrible	pegar	hacer	compleja	final	desastre	arena
D1	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0	0
D3	0	0	0	0	0	0	0	0
D4	0	0	0	0	0	0	0	0
D5	0	0	0	0	0	0	0	0
D6	0	0	0	0	0	0	0	0
D7	1	0	0	0	0	0	0	0
D8	0	1	1	1	1	1	1	0
D9	0	0	0	0	0	0	0	1
D10	0	0	0	0	0	0	0	0
D11	0	0	0	0	0	0	0	0
D12	0	0	0	0	0	0	1	0

	<b>definitiva</b>	<b>delicada</b>	<b>crujiente</b>	<b>único</b>	<b>Caro</b>	<b>producto</b>	<b>ofrecer</b>
D1	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0
D3	0	0	0	0	0	0	0
D4	0	0	0	0	0	0	0
D5	0	0	0	0	0	0	0
D6	0	0	0	0	0	0	0
D7	0	0	0	0	0	0	0
D8	0	0	0	0	0	0	0
D9	1	1	1	1	1	1	1
D10	0	0	0	0	0	0	0
D11	0	0	0	0	0	0	0
D12	0	0	1	0	1	0	0

	<b>gustar</b>	<b>mucho</b>	<b>querer</b>	<b>viernes</b>	<b>otro</b>	<b>vez</b>	<b>volver</b>	<b>dieta</b>
D1	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0	0
D3	0	0	0	0	0	0	0	0
D4	0	0	0	0	0	0	0	0
D5	0	0	0	0	0	0	0	0
D6	0	0	0	0	0	0	0	0
D7	0	1	0	0	0	0	0	0
D8	0	0	1	0	0	0	0	0

D9	0	0	0	0	0	0	0	0
D10	1	1	1	1	1	1	1	0
D11	0	0	0	0	0	0	0	1
D12	0	0	0	0	1	0	0	0

	ver	día	Total	vaya	sitio	colmo
D1	0	0	0	0	0	0
D2	0	0	0	0	0	0
D3	0	0	0	0	0	0
D4	0	0	0	0	0	0
D5	0	0	0	0	0	0
D6	0	0	0	0	0	0
D7	0	0	0	0	0	0
D8	0	0	0	0	0	0
D9	0	0	0	0	0	0
D10	0	0	0	0	0	0
D11	1	1	0	0	0	0
D12	0	0	1	1	1	1

### ■ WordNet Embeddings

	mejor	pizza	que	comer	años	Literal	Desperdi..	Excele..	sabor
D1	1	1	1	1	1	1	1	1	1
D2	0	0	1	0	0	0	0	0	0
D3	0	0	1	1	0	0	0	0	0
D4	1	1	1	0	0	0	0	0	0
D5	0	1	1	0	0	0	0	0	1
D6	0	1	0	0	0	0	0	1	0
D7	0	1	1	0	0	0	0	0	1
D8	0	0	1	0	0	0	0	0	0
D9	0	1	1	0	0	0	0	0	0
D10	0	0	1	0	0	0	0	0	1
D11	0	1	1	0	0	0	0	0	0
D12	1	1	0	0	0	0	0	0	0

	y	textura	recomendar	sentir	salsa	de	tomate	dulce	cabra	haber
D1	1	1	1	0	0	0	0	0	0	1
D2	0	0	0	1	1	1	1	1	1	1
D3	0	0	0	0	0	1	0	0	0	0
D4	0	0	0	0	0	1	0	0	0	0
D5	0	0	0	0	0	1	0	0	0	0
D6	1	0	0	0	0	0	0	0	1	1
D7	0	0	0	0	0	0	0	0	0	1
D8	1	0	0	1	0	0	0	0	1	0
D9	1	1	0	1	0	1	0	0	0	1
D10	0	0	0	0	0	0	0	0	0	0
D11	0	0	0	0	0	0	0	0	0	0
D12	0	0	1	0	0	0	0	0	0	0

	preferir	queso	mozzarella	normal	meno	mas	tan	terrible	punta
D1	0	0	0	0	0	0	0	0	0
D2	1	1	1	1	0	0	0	1	0
D3	0	0	0	0	1	1	1	1	1
D4	0	0	0	0	0	0	0	1	0
D5	0	0	0	0	0	0	0	0	0

D6	0	1	0	0	0	0	1	0	0
D7	0	0	0	0	0	0	0	0	0
D8	0	1	0	0	0	0	0	0	0
D9	0	0	0	0	0	0	0	0	0
D10	0	0	0	0	0	0	0	0	0
D11	0	0	0	0	0	0	0	0	0
D12	0	0	0	0	0	0	0	0	0

	pedazo	demasiado	ingrediente	engordar	animal	idea	no	ejecutar	ser
D1	0	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0	0	0
D3	1	1	1	1	1	0	0	0	1
D4	0	0	0	0	0	1	1	1	1
D5	0	0	0	0	0	0	0	0	1
D6	0	0	0	0	0	0	0	0	0
D7	0	0	0	0	0	0	0	0	0
D8	0	0	0	0	0	0	1	0	1
D9	0	0	0	0	0	0	0	0	1
D10	0	0	0	0	0	0	0	0	1
D11	0	0	0	0	0	0	1	0	0
D12	0	0	0	0	0	0	1	0	1

	terminar	embutido	alta	calidad	del	pepperoni	mi	salvo	demás	basura
D1	0	0	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0	0	0	0
D3	1	0	0	0	0	0	0	0	0	0
D4	0	1	1	1	1	1	1	1	1	1
D5	0	0	0	0	0	0	0	0	0	0
D6	0	0	0	0	0	0	1	0	0	0
D7	0	0	0	0	0	0	0	0	0	0
D8	1	0	0	0	0	0	0	0	0	0
D9	0	0	0	0	0	0	0	0	0	0
D10	0	0	0	0	0	0	0	0	0	0
D11	0	0	0	0	0	0	0	0	0	0

D12	0	0	0	0	0	0	0	0	0	1
-----	---	---	---	---	---	---	---	---	---	---

	encantar	estar	incorporación	exótico	increíble	verdad	experiencia	todos
D1	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0	0
D3	0	0	0	0	0	0	0	0
D4	0	0	0	0	0	0	0	0
D5	1	0	1	1	1	1	1	1
D6	0	0	0	0	0	0	0	0
D7	0	0	0	0	0	0	0	0
D8	0	0	0	0	0	0	0	0
D9	0	0	0	0	0	0	0	0
D10	0	0	0	0	0	0	0	0
D11	0	1	0	0	0	0	0	0
D12	0	0	0	0	0	0	0	0

	deber	degustar	nunca	probar	dátil	sorprender	combinar	bien	top
D1	0	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	1	0	0
D3	0	0	0	0	0	0	0	0	0
D4	0	0	0	0	0	0	0	1	0
D5	1	1	0	0	0	0	0	0	0
D6	0	0	1	1	1	1	1	1	1
D7	0	0	0	0	0	0	0	0	0
D8	0	0	0	0	1	0	0	0	0
D9	0	0	0	1	0	0	0	0	0
D10	0	0	0	0	0	0	0	0	0
D11	0	0	0	0	0	0	0	0	0
D12	0	0	0	0	0	0	0	0	0

	compra	miedo	enamorar	leer	mucha	review	mala	al	quedar	delicia
D1	0	0	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0	0	0	0
D3	0	0	0	0	0	0	0	0	0	0

D4	0	0	0	0	0	0	0	0	0	0
D5	0	0	0	0	0	0	0	0	0	0
D6	0	0	0	0	0	0	0	0	0	0
D7	1	1	1	1	1	1	1	1	1	1
D8	0	0	0	0	0	0	0	1	0	0
D9	0	0	0	0	0	0	0	0	0	1
D10	0	0	0	0	0	0	0	0	0	0
D11	0	0	0	0	0	0	0	0	0	1
D12	0	0	0	0	0	0	0	0	0	0

	cuanto	horrible	pega	Querer	hacer	compleja	final	desastre	arena
D1	0	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0	0	0
D3	0	0	0	0	0	0	0	0	0
D4	0	0	0	0	0	0	0	0	0
D5	0	0	0	0	0	0	0	0	0
D6	0	0	0	0	0	0	0	0	0
D7	1	0	0	0	0	0	0	0	0
D8	0	1	1	1	1	1	1	1	0
D9	0	0	0	0	0	0	0	0	1
D10	0	0	0	1	0	0	0	0	0
D11	0	0	0	0	0	0	0	0	0
D12	0	0	0	0	0	0	0	1	0

	definitiva	delicada	crujiente	único	Caro	producto	ofrecer	me
D1	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0	0
D3	0	0	0	0	0	0	0	1
D4	0	0	0	0	0	0	0	0
D5	0	0	0	0	0	0	0	0
D6	0	0	0	0	0	0	0	1
D7	0	0	0	0	0	0	0	0
D8	0	0	0	0	0	0	0	0
D9	1	1	1	1	1	1	1	0



D10	0	0	0	0	0	0	0	0
D11	0	0	0	0	0	0	0	1
D12	0	0	1	0	1	0	0	0

	gustar	mucho	ya	querer	viernes	otro	vez	volver	dieta
D1	0	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0	0	0
D3	0	0	0	0	0	0	0	0	0
D4	0	0	0	0	0	0	0	0	0
D5	0	0	0	0	0	0	0	0	0
D6	0	0	0	0	0	0	0	0	0
D7	0	1	0	0	0	0	0	0	0
D8	0	0	0	1	0	0	0	0	0
D9	0	0	0	0	0	0	0	0	0
D10	1	1	1	1	1	1	1	1	0
D11	0	0	0	0	0	0	0	0	1
D12	0	0	0	0	0	1	0	0	0

	ver	día	Total	vaya	sitio	colmo
D1	0	0	0	0	0	0
D2	0	0	0	0	0	0
D3	0	0	0	0	0	0
D4	0	0	0	0	0	0
D5	0	0	0	0	0	0
D6	0	0	0	0	0	0
D7	0	0	0	0	0	0
D8	0	0	0	0	0	0
D9	0	0	0	0	0	0
D10	0	0	0	0	0	0
D11	1	1	0	0	0	0
D12	0	0	1	1	1	1

## ULTIMA ACTIVIDAD

Dadas las siguientes 9 frases y el siguiente vocabulario admitido, construya un modelo de procesamiento del lenguaje natural que permita clasificar el sentimiento de una frase de entrada empleando el algoritmo de Naive Bayes. Debe mostrar el proceso paso a paso y hacerlo de forma detallada. Para el cálculo de label de las estimaciones debe desarrollar cómo lo calcularía Naive Bayes también.

NOTA: Considerar tokens similares como mismo token (lematización).

Estimar qué devolvería su modelo de resultado para:

- Si necesitas dinero rápido y fácil, haz click en este enlace, no te arrepentirás.
- Hola, no he recibido el dinero que acordaron pagarme, favor recordar convertir el mismo.

Frases:

Frase	Sentimiento
Gana dinero fácil conmigo.	Spam
Hola, necesito dinero mami.	No Spam
Hola, puedes ganar dinero fácil y rápido haciendo click aquí.	Spam
Haz click aquí para concursar en línea por 1000 dólares.	Spam
Hola Laura, te acabo de enviar el dinero que me pediste.	No Spam
¿Quieres conseguir dinero rápido? Haz click aquí para ver cómo.	Spam
Hola Marco, yo hice click en el botón que me dijiste, pero no se transfirió el dinero.	No Spam
No pude convertir el dinero de dólares a pesos dominicanos.	No Spam
Respóndeme rápido, que necesito saber cuánto dinero usarás.	No Spam

Vocabulario:

Vocabulario
aquí
click
convertir
cuánto
dinero
dólares
enviar
fácil

gana
haciendo
hola
necesito
pesos
dominicanos
puedes
rápido

**Respuesta:**

Vocabulario	SPAM			No spam		
	NVT	nw	$p(w ctv)$	NVT	nw	$p(w ctv)$
aqui	10	2	0.125	19	1	0.060606061
click	10	2	0.125	19	2	0.090909091
convertir	10	0	0.041666667	19	1	0.060606061
cuanto	10	0	0.041666667	19	1	0.060606061
dinero	10	2	0.125	19	2	0.090909091
dolares	10	1	0.083333333	19	1	0.060606061
enviar	10	0	0.041666667	19	1	0.060606061
facil	10	1	0.083333333	19	1	0.060606061
gana	10	1	0.083333333	19	1	0.060606061
haciendo	10	0	0.041666667	19	0	0.03030303
hola	10	0	0.041666667	19	3	0.121212121
necesito	10	0	0.041666667	19	1	0.060606061
pesos dominicanas	10	0	0.041666667	19	1	0.060606061
puedes	10	1	0.083333333	19	1	0.060606061
rapido	10	0	0.041666667	19	1	0.060606061

Frases	Probabilidad Spam	Probabilidad No Spam	resultado
• Si necesitas dinero rápido y fácil, haz click en este enlace, no te arrepentirás.	38%	30%	Spam
• Hola, no he recibido el dinero que acoraron pagarme, favor recordar convertir el mismo.	21%	27%	No Spam