



# Herramientas y Aplicaciones de la IA. IA en Amazon Web Services Amazon Comprehend

Jaume Jordán

# Amazon Comprehend

- Servicio de NLP que utiliza machine learning para extraer información de texto
- Reconocimiento de idioma, detección de entidades, análisis de frases clave, análisis de sentimientos, eventos, sintaxis...
- Gestionado y entrenado por AWS
- Tiempo real (síncrono) o batch (asíncrono)
- Integración en aplicaciones y con otros servicios
- <https://docs.aws.amazon.com/comprehend/latest/dg/what-is.html>
- Idiomas soportados:  
<https://docs.aws.amazon.com/comprehend/latest/dg/supported-languages.html>



# Casos de uso de Amazon Comprehend

- Análisis de documentos
- Análisis de clientes
- Determinar sentimiento de clientes
- Clasificación de tickets de soporte



# Análisis en Tiempo Real de Comprehend

- Análisis instantáneo de textos, ideal para aplicaciones en vivo y respuestas automáticas.
- Escenarios de uso: Servicios de atención al cliente, monitoreo de redes sociales, y sistemas de alerta inmediata.
- Permite reaccionar rápidamente a la información entrante y ajustar operaciones o comunicaciones de forma dinámica.

# Análisis en Tiempo Real de Comprehend

- Identificación del idioma dominante:

- +100 idiomas

- <https://docs.aws.amazon.com/comprehend/latest/dg/how-languages.html>

- Detección de entidades

- Entidad, categoría y confianza

- Únicamente en idiomas principales

- Ejemplos:

- Date

- Event

- Location

- Person

- Quantity

# Análisis en Tiempo Real de Comprehend

## ■ Detección de frases clave


- ☐ Sintagma nominal
- ☐ Solo en idiomas principales
- ☐ <https://docs.aws.amazon.com/comprehend/latest/dg/how-key-phrases.html>

## ■ Detección de sentimiento

- ☐ Opinión general del texto
- ☐ Neutral, positiva, negativa, mixta
- ☐ Solo en idiomas principales
- ☐ <https://docs.aws.amazon.com/comprehend/latest/dg/how-sentiment.html>

## ■ Sintaxis

- ☐ Desglose de elementos del texto
- ☐ Tokens: adjetivos, adverbios, nombres propios, verbos, preposiciones...
- ☐ Solo en idiomas principales
- ☐ <https://docs.aws.amazon.com/comprehend/latest/dg/how-syntax.html>



# Funcionamiento de Comprehend

- Tecnología subyacente: Uso de modelos de aprendizaje automático y técnicas de NLP.
- Personalización: Opción de entrenar modelos personalizados para necesidades específicas de análisis de texto.
  - Document processing, Custom classification, Custom entity recognition
- Seguridad y privacidad: Cumple con las políticas de privacidad y seguridad de AWS, garantizando la confidencialidad de los datos.

# Uso de Comprehend con la API desde AWS CLI

- <https://docs.aws.amazon.com/cli/latest/reference/comprehend/>
- `aws comprehend detect-dominant-language --text "It is a beautiful day in Seattle."`



# Detección del Idioma Principal en Python

```
import boto3

comprehend = boto3.client(service_name='comprehend')
text = "Escribe aquí tu texto"
response = comprehend.detect_dominant_language(Text=text)
print('Idiomas detectados:', response['Languages'])
```

# Detección de Entidades

- Reconoce nombres de personas, lugares, fechas, etc., dentro del texto.

```
import boto3

comprehend = boto3.client(service_name='comprehend')
text = "El BBVA lanza una opa hostil para comprar el Sabadell. El banco presidido por Carlos Torres presenta una oferta al mismo precio que la propuesta amistosa de la semana pasada. El Sabadell reitera su rechazo a la operación."
response = comprehend.detect_entities(Text=text, LanguageCode='es')
print('Entidades detectadas:')
for entity in response['Entities']:
    print(entity['Text'], entity['Type'], entity['Score'])
```

# Detección de entidades PII

## ■ Personally Identifiable Information (PII)

- ADDRESS, AGE, AWS\_ACCESS\_KEY, AWS\_SECRET\_KEY, CREDIT\_DEBIT\_CVV, CREDIT\_DEBIT\_EXPIRY, CREDIT\_DEBIT\_NUMBER, DATE\_TIME, DRIVER\_ID, EMAIL, INTERNATIONAL\_BANK\_ACCOUNT\_NUMBER, IP\_ADDRESS, LICENSE\_PLATE, MAC\_ADDRESS, NAME, PASSWORD, PHONE, PIN, SWIFT\_CODE, URL, USERNAME, VEHICLE\_IDENTIFICATION\_NUMBER

```
import boto3

comprehend = boto3.client(service_name='comprehend')
text = "El BBVA lanza una opa hostil para comprar el Sabadell. El banco presidido por Carlos Torres presenta una oferta al mismo precio que la propuesta amistosa de la semana pasada. El Sabadell reitera su rechazo a la operación."
response = comprehend.detect_pii_entities(Text=text, LanguageCode='es')
print('Entidades detectadas:')
for entity in response['Entities']:
    print(entity, text[entity['BeginOffset']:entity['EndOffset']])
```

# Frases Clave

- Extrae términos o frases que son importantes en el contexto del texto.

```
import boto3

comprehend = boto3.client(service_name='comprehend')
text = "El BBVA lanza una opa hostil para comprar el Sabadell. El banco presidido por Carlos Torres presenta una oferta al mismo precio que la propuesta amistosa de la semana pasada. El Sabadell reitera su rechazo a la operación."
response = comprehend.detect_key_phrases(Text=text, LanguageCode='es')
print('Frases clave detectadas:')
for phrase in response['KeyPhrases']:
    print(phrase)
```

# Análisis de Sentimientos

- Determina si el sentimiento del texto es positivo, negativo, neutral o mixto.

```
response = comprehend.detect_sentiment(Text=text,  
LanguageCode='es')
```


```
print('Sentimiento:', response['Sentiment'])
```

# Análisis de Sintaxis

- Desglosa el texto en tokens y proporciona información gramatical de cada palabra.

```
response = comprehend.detect_syntax(Text=text,  
LanguageCode='es')
```

```
print('Tokens:', response['SyntaxTokens'])
```



# Conexión y Combinación con Otros Servicios como AWS Transcribe y Translate

- Integración con AWS Transcribe: Transforma audio en texto y luego utiliza Comprehend para analizarlo.
- Integración con AWS Translate: Traduce textos a otros idiomas y luego aplica análisis de texto con Comprehend.
- Estas integraciones pueden ser usadas en escenarios reales como atención al cliente multilingüe o análisis de tendencias globales.

# Práctica propuesta. Tarea 4 AWS.

## Análisis de comentarios de clientes.

- Realizar una de las 4 tareas propuestas en la *Práctica AWS Rekognition y Comprehend* de PoliformaT.
- Utiliza un dataset de comentarios de clientes como los que puedes encontrar en <https://www.kaggle.com/datasets/vishweshsalodkar/customer-feedback-dataset/code>
- Detectar el idioma dominante de cada comentario.
- Dependiendo del idioma del comentario, detectar el sentimiento de dicho comentario.
- Obtener e imprimir estadísticas de comentarios, positivos, neutros y negativos.
- Detectar y mostrar las entidades que más se repiten en los comentarios, así como mostrar el número de entidades únicas en el dataset.





# Resumen servicios IA AWS

- Amazon SageMaker: Servicio integral que facilita a los científicos y desarrolladores de datos crear, entrenar y desplegar modelos de machine learning a escala.
- Amazon Rekognition: Servicio que permite añadir análisis de imagen y video a tus aplicaciones para identificar objetos, personas, texto, escenas y actividades.
- Amazon Comprehend: Servicio de procesamiento de lenguaje natural que utiliza machine learning para descubrir información y relaciones en texto.
- Amazon Forecast: Servicio de predicción basado en machine learning que genera predicciones de tiempo series con exactitud para planificación de recursos y más.
- Amazon Translate: Servicio de traducción automática que utiliza técnicas avanzadas de machine learning para traducir texto entre idiomas.
- Amazon Lex: Servicio para construir interfaces de conversación en cualquier aplicación usando voz y texto, impulsado por la misma tecnología que Alexa.
- Amazon Polly: Servicio que convierte el texto en habla realista, permitiendo añadir voces habladas a aplicaciones.
- AWS DeepLens: Cámara de video habilitada para deep learning que puede correr modelos de aprendizaje profundo localmente y en tiempo real.
- Amazon Personalize: Servicio de machine learning que facilita la creación de recomendaciones personalizadas para tus clientes.
- Amazon Kendra: Motor de búsqueda empresarial impulsado por machine learning que proporciona a los usuarios respuestas más precisas a preguntas en lenguaje natural.
- Amazon Textract: Servicio que extrae automáticamente texto y datos de documentos escaneados, incluyendo tablas y formularios.
- AWS DeepRacer: Plataforma autónoma de coche de carreras modelo a escala, diseñada para enseñar los fundamentos de la conducción autónoma y la competencia en la liga de carreras.
- Amazon Fraud Detector: Servicio que identifica actividades fraudulentas en línea en tiempo real, como pagos fraudulentos y cuentas falsas.

# Resumen servicios IA AWS (cont.)

- Amazon Lookout for Vision: Detecta defectos y anomalías en imágenes físicas utilizando machine learning para mejorar la calidad del producto.
- Amazon Lookout for Equipment: Analiza los datos de los sensores de equipos para detectar anomalías inusuales en el funcionamiento de las máquinas, lo que ayuda en el mantenimiento predictivo.
- Amazon Lookout for Metrics: Identifica y diagnostica anomalías en tus métricas empresariales automáticamente, facilitando la identificación rápida de problemas inesperados.
- Amazon HealthLake: Permite almacenar, transformar y analizar datos de salud en grandes escalas, facilitando la gestión y análisis de información médica con formatos compatibles como FHIR (Fast Healthcare Interoperability Resources).
- AWS Panorama: Permite a las empresas añadir capacidades de aprendizaje automático a sus cámaras de CCTV existentes para analizar imágenes de video y tomar decisiones automatizadas localmente.
- Amazon Augmented AI (A2I): Facilita la implementación de revisiones humanas para las predicciones de machine learning, permitiendo a los desarrolladores construir flujos de trabajo para la validación humana de las decisiones de IA.
- Amazon CodeGuru: Herramienta que revisa el código automáticamente y ofrece recomendaciones de optimización para mejorar la calidad del software y reducir costes.
- AWS DeepComposer: Servicio que proporciona una manera creativa y divertida de aprender sobre machine learning, permitiendo a los usuarios crear música con modelos generativos.
- Amazon Elastic Inference: Permite adjuntar capacidades de computación de inferencia acelerada a cualquier instancia de Amazon EC2 o Amazon SageMaker, reduciendo así los costos de inferencia de machine learning.
- AWS Glue Databrew: Herramienta de preparación de datos visuales que permite a los científicos de datos limpiar y normalizar datos sin escribir código.
- Amazon Monitron: Sistema de monitorización de condiciones de la máquina que usa sensores y machine learning para detectar anomalías en equipos industriales, ayudando a prevenir fallas no planificadas.
- Amazon Neptune ML: Servicio que facilita la creación y ejecución de predicciones de machine learning utilizando gráficos de bases de datos, ideal para conexiones complejas entre datos.

# Extensiones para TFM de IA en AWS

- Predicción y configuración de portfolio de inversión en AWS
- Sistemas de recomendación en AWS
- Chatbots en AWS
- Otras propuestas...
- Contactar con Juanmi Alberola ([jalberola@dsic.upv.es](mailto:jalberola@dsic.upv.es)) y Jaume Jordán ([jjordan@dsic.upv.es](mailto:jjordan@dsic.upv.es))