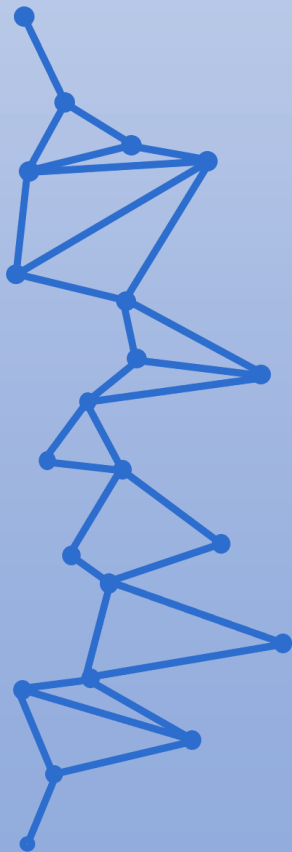




Curso de Especialización de Inteligencia Artificial y Big Data (IABD)



Programación de Inteligencia Artificial

UD02. Comparativa de dos API de
reconocimiento de imagen.
Tarea Online.

JUAN ANTONIO GARCÍA MUELAS

INDICE

	Pag
1. Apartado 1: Comparar dos herramientas de reconocimiento de imagen en la nube	2
2. Apartado 2: Comparar dos herramientas de análisis de lenguaje natural para análisis de sentimiento	12

Título de la tarea: Comparativa de dos API de reconocimiento de imagen

Unidad: 2

Curso de Especialización y módulo: Inteligencia Artificial y Big Data - Programación en Inteligencia Artificial.

Curso académico: 2022-2023

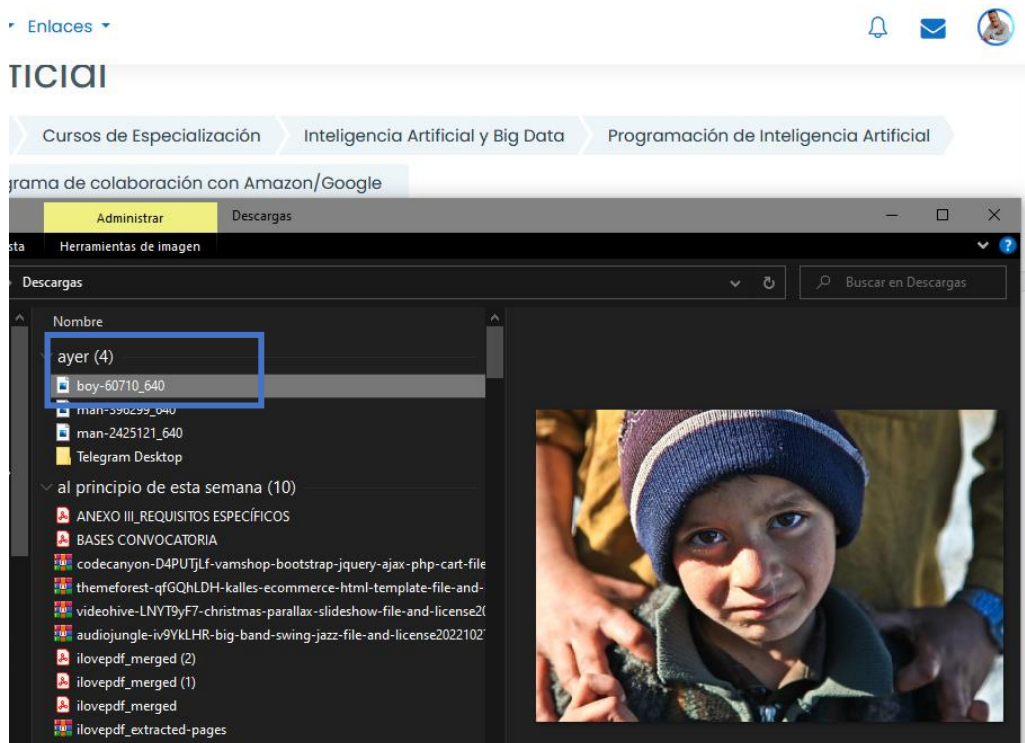
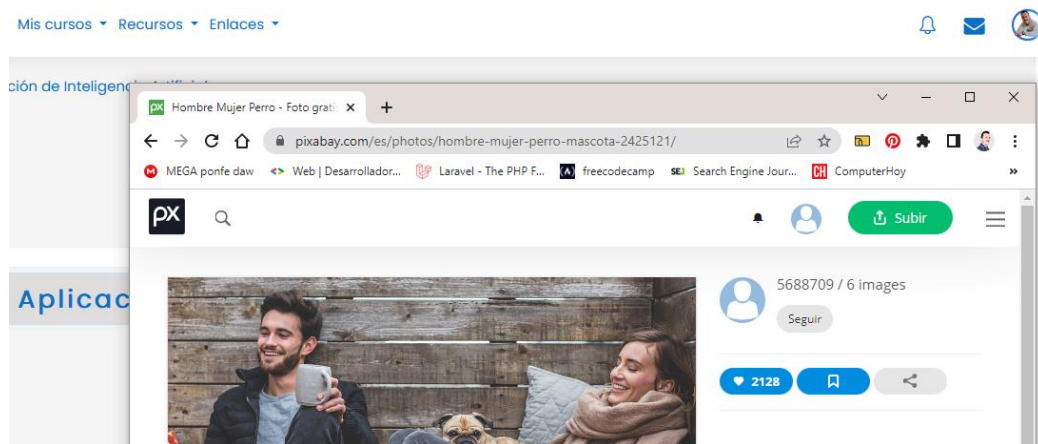
Tarea para PIA02

<https://cloud.google.com/vision/docs/drag-and-drop>

✓ Apartado 1: Comparar dos herramientas de reconocimiento de imagen en la nube

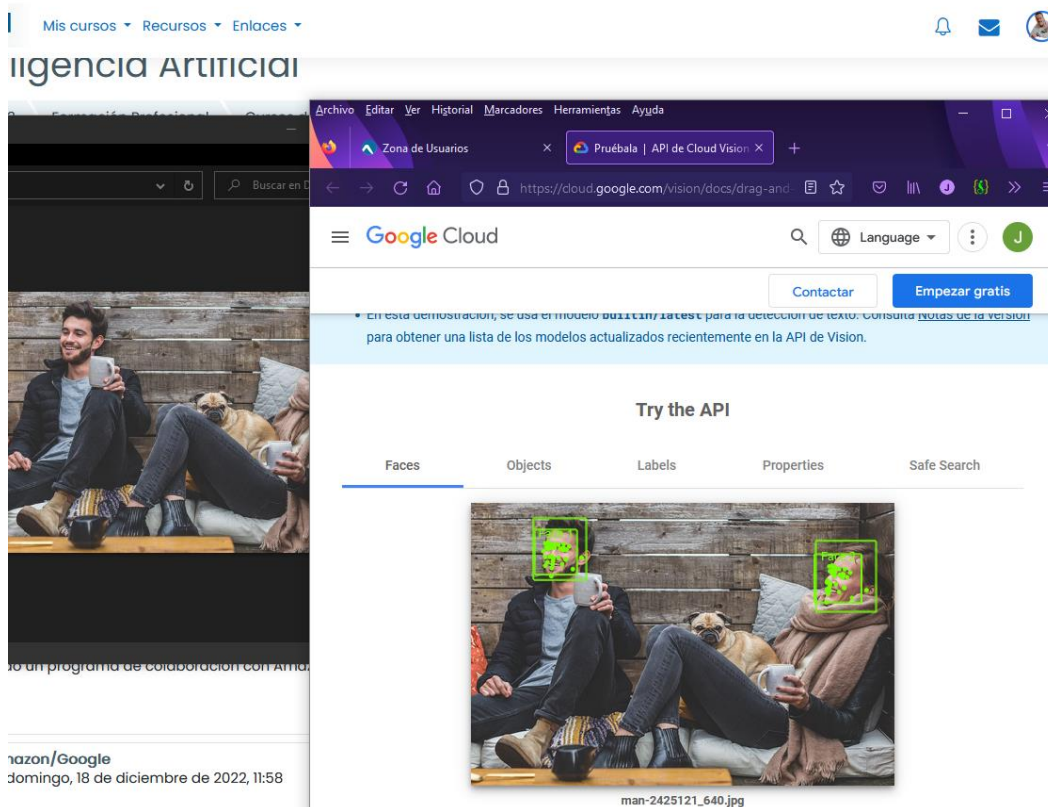
- Selecciona varias imágenes de rostros de personas expresando distintas emociones en un repositorio de imágenes gratuitas.

He accedido a través de mi cuenta en [Pixabay](#) y descargado tres imágenes en mi equipo.

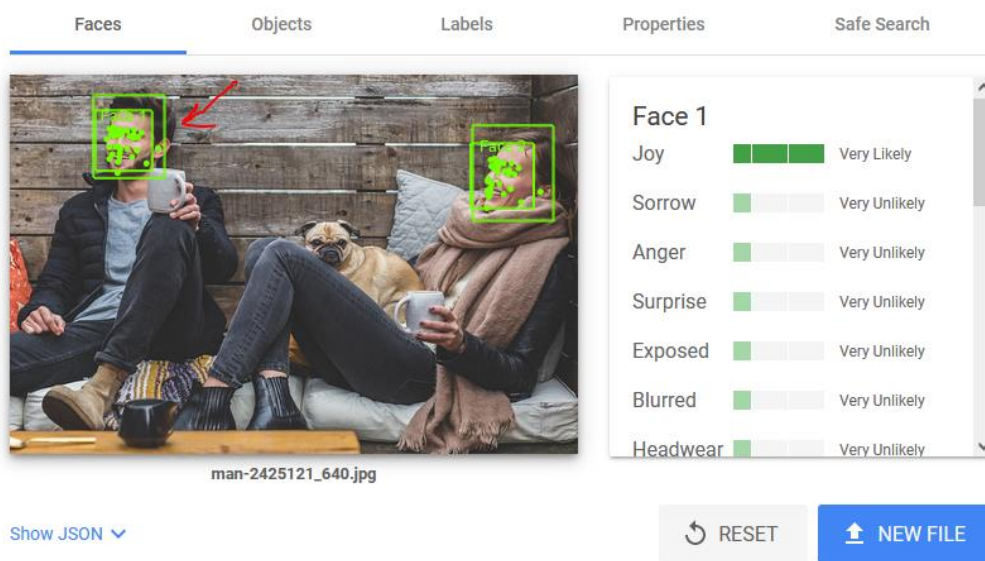


- Accede a la interfaz de prueba de Google Vision AI a través de este enlace. Sin pulsar en ningún botón o menú, haz scroll hacia abajo hasta llegar a la ventana para cargar imágenes. Sube una de las fotos y analiza los resultados que muestra la consola. También puedes observar la llamada y respuesta en formato JSON.

He comenzado este apartado con la imagen que me parecía poder ser la más compleja de análisis, dado el contexto global.



Tras analizar la imagen nos muestra los resultados:



Podemos observar cómo separa el reconocimiento de cada rostro de forma clara y determinando los puntos de análisis de cada uno.

The screenshot displays the Google Cloud Vision API interface with the 'Faces' tab selected. It shows two faces identified in a photo of two people sitting on a bench. The interface includes tabs for Faces, Objects, Labels, Properties, and Safe Search. The 'Properties' panel on the right shows various attributes for the selected face, including emotions, facial features, and orientation. The 'Confidence' score is 96%.

Property	Value
Anger	Very Unlikely
Surprise	Very Unlikely
Exposed	Very Unlikely
Blurred	Very Unlikely
Headwear	Very Unlikely
Roll	6°
Tilt	0°
Pan	-21°
Confidence	96%

The interface also shows a 'Safe Search' section with a 'Confidence' score of 96%.

The 'Faces' tab shows two faces identified in the image. The 'Properties' panel for 'Face 2' shows the following attributes:

Property	Value
Joy	Very Likely
Sorrow	Very Unlikely
Anger	Very Unlikely
Surprise	Very Unlikely
Exposed	Very Unlikely

The 'Safe Search' section shows a 'Confidence' score of 96%.

The interface also shows a 'Safe Search' section with a 'Confidence' score of 96%.

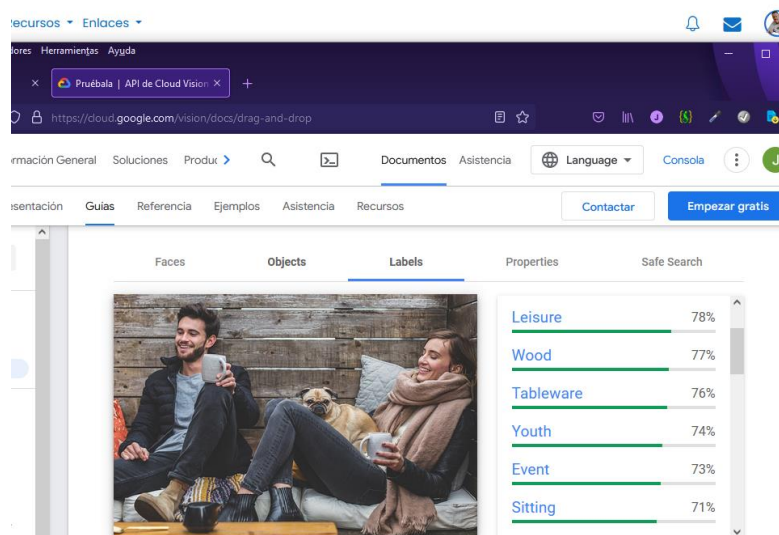
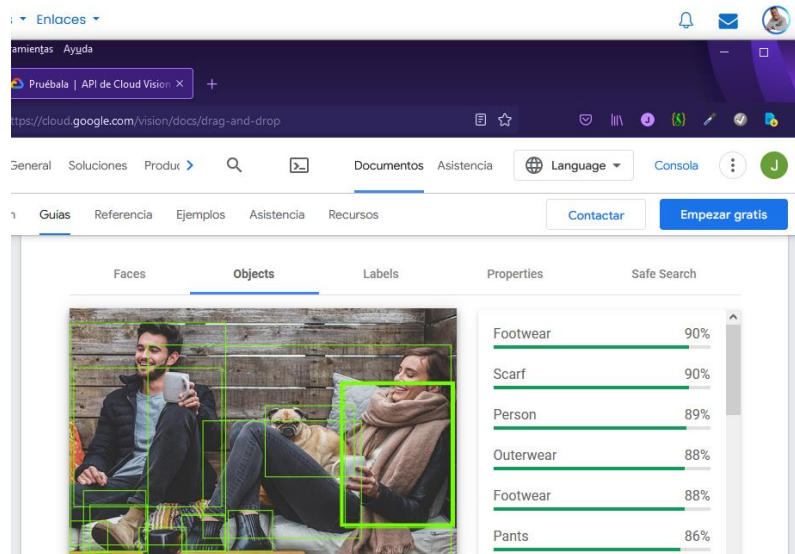
The 'Faces' tab shows two faces identified in the image. The 'Properties' panel for 'Face 2' shows the following attributes:

Property	Value
Anger	Very Unlikely
Surprise	Very Unlikely
Exposed	Very Unlikely
Blurred	Very Unlikely
Headwear	Very Unlikely
Roll	16°
Tilt	-2°
Pan	-45°
Confidence	73%

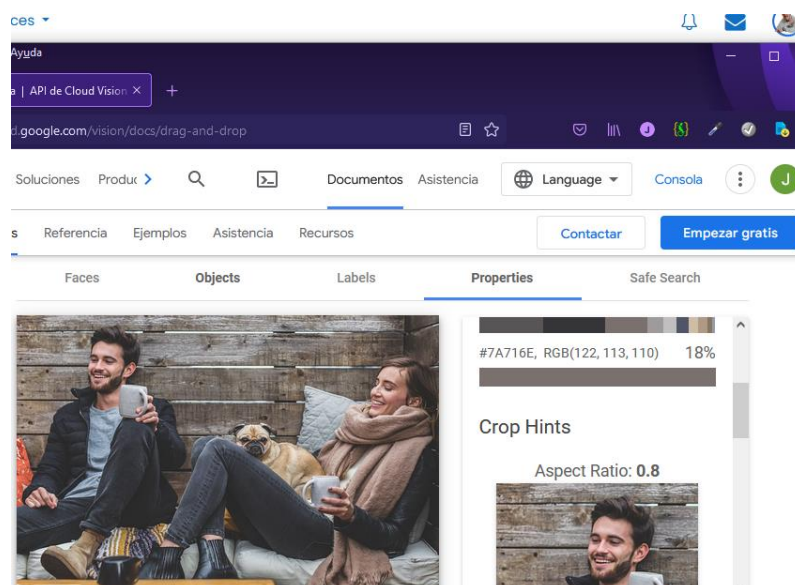
The 'Safe Search' section shows a 'Confidence' score of 96%.

En el caso del rostro femenino nos muestra una puntuación (porcentaje) algo menor respecto a la predicción anterior, pero sigue acertando de forma clara en el resto.

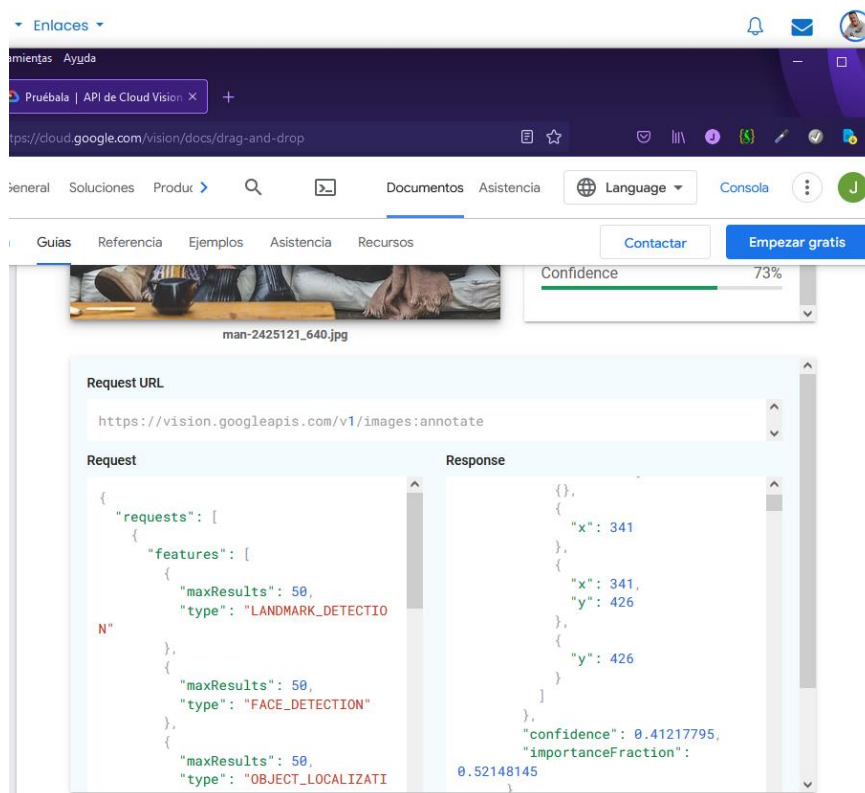
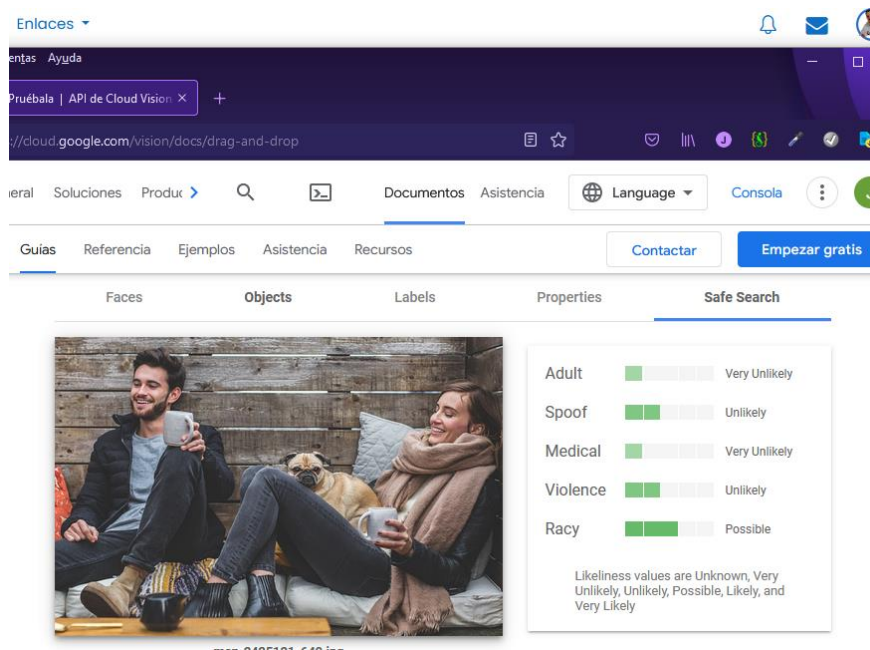
Podemos seguir avanzando por las distintas pestañas y ver como realiza el reconocimiento de los objetos presentes, las etiquetas y propiedades...



Me ha parecido muy interesante el análisis cromático.



Y por supuesto, los resultados para el uso seguro de la imagen.

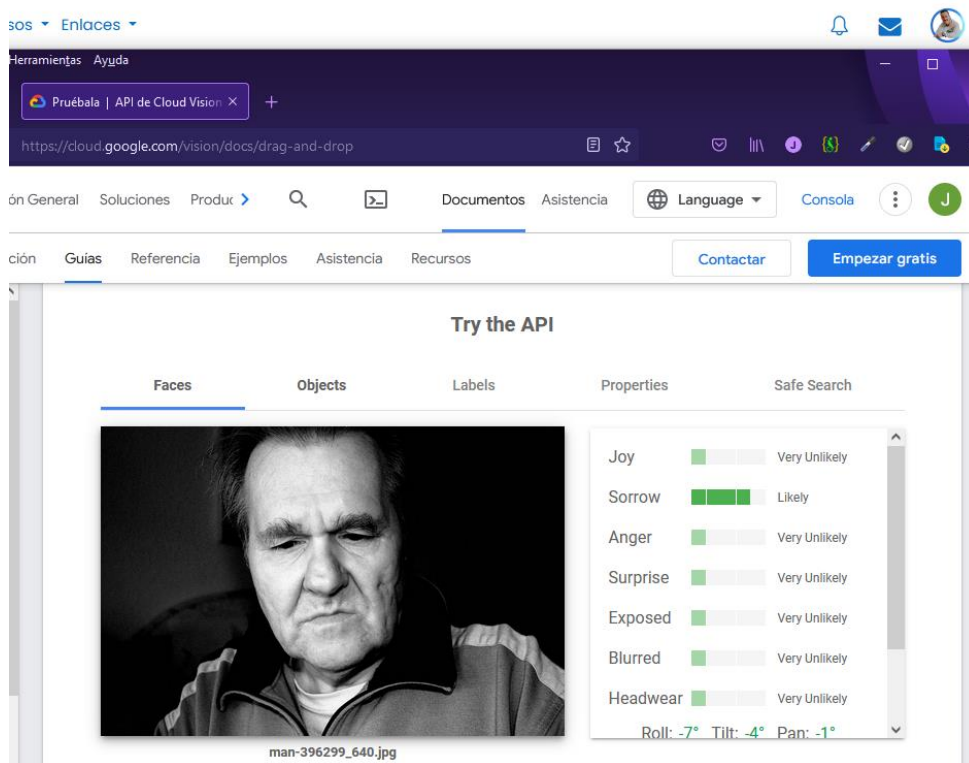


Los resultados en formato JSON tanto de la petición como de la respuesta son exportables.

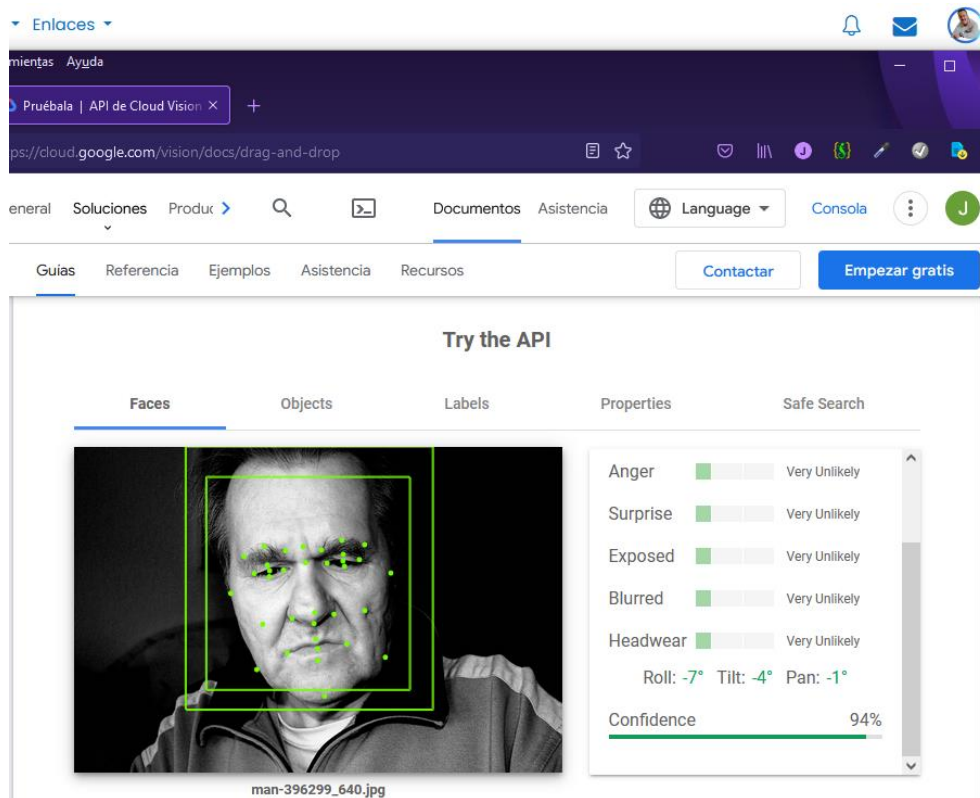
- Repite el proceso con todas las fotos, analizando en qué grado se captan correctamente las emociones.

He intentado buscar imágenes que pudieran mostrar sentimientos de varias formas. Que no fueran excesivamente claros, para observar cómo trabaja esta herramienta y resuelve con los parámetros que utiliza para esta versión de prueba.

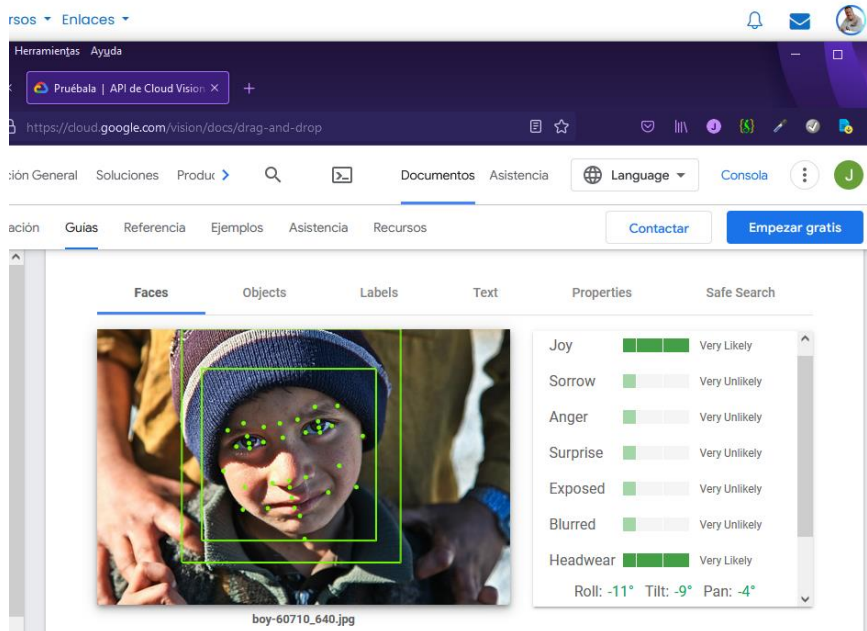
Vemos las imágenes para un adulto y un niño:



En esta imagen creo que acierta plenamente dentro de estas limitaciones comentadas para la versión gratuita (Podría ser triste, preocupado, pensativo..., por lo que encaja en la búsqueda de emociones).



Donde creo que no hace un análisis correcto es en esta última, donde los rasgos labiales pueden denotar una leve sonrisa, pero donde el resto del rostro muestra tristeza o dolor.



- Accede a la interfaz de prueba de Rekognition a través de [este enlace](#). Si es posible, pon delante de la cámara un dispositivo con una de las imágenes descargadas, y si no, Pon tú mismo un gesto de emoción para que lo capte la cámara. Pulsa el botón "Detect Labels" y analiza los resultados que arroja la herramienta.

Para este apartado, al no disponer de cámara en el dispositivo, he utilizado la cámara virtual de OBS Studio, para que Rekognition recogiera las imágenes utilizadas en el punto anterior.

Comienzo las pruebas para esta parte con el mismo orden de imágenes que con Google.

Al pulsar sobre "Detect Labels", compruebo que está recogiendo datos de toda la pantalla (ya que quería demostrar la autoría de la captura con la plataforma detrás), lo que dificulta o imposibilita que detecte alguna emoción en la imagen, aunque consigue mostrar las etiquetas Person, Woman, Adult y Female, relativas a las personas detectadas.

amazon Rekognition

Classification successful!

Label	Confidence
File	99.32327270507812
Person	98.18283081054688
Woman	98.18283081054688
Adult	98.18283081054688
Female	98.18283081054688
Monitor	88.98051452636719
Screen	88.98051452636719
Computer Hardware	88.98051452636719
Electronics	88.98051452636719
Hardware	88.98051452636719
Webpage	85.92578125
Game	56.343509674072266
Pc	56.33686447143555
Computer	56.33686447143555
Page	55.37959671020508
Text	55.37959671020508

Results:

Classification successful!

Label	Confidence
File	99.32327270507812
Person	98.18283081054688
Woman	98.18283081054688
Adult	98.18283081054688
Female	98.18283081054688
Monitor	88.98051452636719
Screen	88.98051452636719
Computer Hardware	88.98051452636719
Electronics	88.98051452636719
Hardware	88.98051452636719
Webpage	85.92578125
Game	56.343509674072266
Pc	56.33686447143555
Computer	56.33686447143555
Page	55.37959671020508
Text	55.37959671020508

Debemos por tanto, concentrar el ámbito de búsqueda ampliando la captación de imagen.

Step 1: Take picture

Detect Labels

Results: Classification successful!

Label	Confidence
Photography	99.5389994785156
Page	99.4169921875
Text	99.4169921875
Cup	99.19815626416016
Person	98.98507690429688
Woman	98.98507690429688
Adult	98.98507690429688
Female	98.98507690429688
Portrait	94.5710464477539
Face	94.5710464477539
Head	94.5710464477539
Screen	89.59030151367188
Electronics	89.59030151367188
Couch	82.10773468017578
Furniture	82.10773468017578
Monitor	81.61433410644531
Computer Hardware	81.61433410644531
Hardware	81.61433410644531
Pc	74.76030731201172
Computer	74.76030731201172
Parts	73.78761291503906
Clothing	73.78761291503906
Canine	62.878013610839844
Mammal	62.878013610839844
Animal	62.878013610839844
Blanket	57.46904246363758
Pet	57.46206329345703
Sleeping	56.70808780114258
Scarf	56.0502319359375
Indoors	56.03846828782286
Human Design	55.37210327523096

Sin embargo, no se consiguen resultados reales hasta dejar sólo la imagen en pantalla, momento en el que ya refleja etiquetas relativas al escaneo de los rostros y sus sentimientos.

Step 1: Take picture

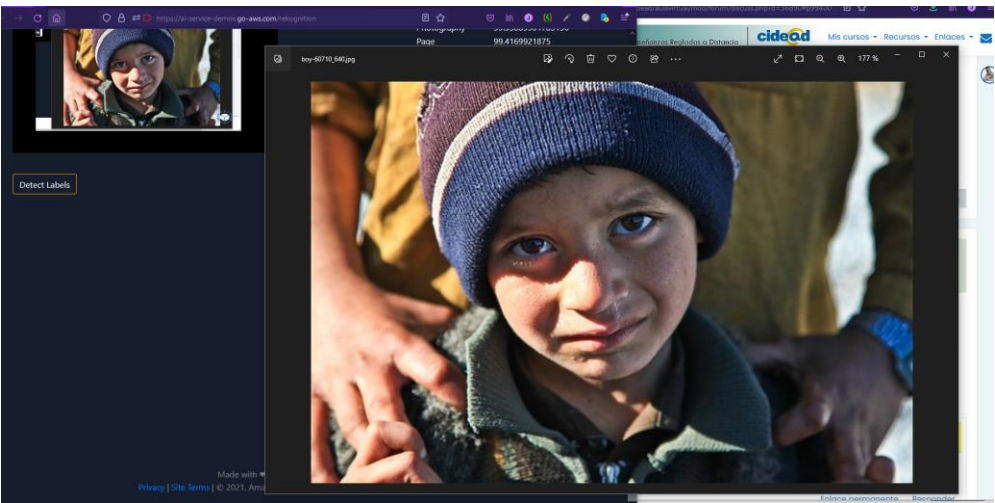
Detect Labels

Results: Classification successful!

Label	Confidence
Cup	99.93594360351562
Glove	99.7524642944336
Clothing	99.7524642944336
Person	99.48271179199219
Man	99.48271179199219
Adult	99.48271179199219
Male	99.48271179199219
Couch	99.20512390136719
Furniture	99.20512390136719
Woman	99.19364166259766
Female	99.19364166259766
Shoe	99.085205078125
Footwear	99.085205078125
Face	94.48101043701172
Head	94.48101043701172
Pants	93.4391860961914
Interior Design	92.41876220703125
Indoors	92.41876220703125
Photography	89.00714111328125
Canine	71.38993835449219
Mammal	71.38993835449219
Animal	71.38993835449219
Pet	66.06005096435547
Smile	56.4412956237793
Happy	56.4412956237793
Sitting	56.25095748901367
Pug	55.49605941772461
Dog	55.49605941772461
Scarf	55.47076416015625

- Repite el proceso con varias emociones y analiza si la herramienta es capaz de detectar que la persona está sonriendo.

Probamos con imagen en la que fallaba Vision AI.



En el primer escaneo, a pesar de haber dado mayor presencia a la imagen, no detecta emociones.


Amazon Rekognition makes it easy to add image and video analysis to your applications. You just provide an image or video to the Amazon Rekognition API, and the service can identify objects, people, text, scenes, and activities. It can detect any inappropriate content as well. Amazon Rekognition also provides highly accurate facial analysis and facial recognition. You can detect, analyze, compare faces for a wide variety of use cases, including user verification, cataloging, people counting, and public safety.

In this example, we're going to show how easy it is to send an image to Amazon Rekognition to perform object identification.

Methods:

`detectLabels()`: Detect object labels from an input image!

Step 1: Take picture




Detect Labels

Results:

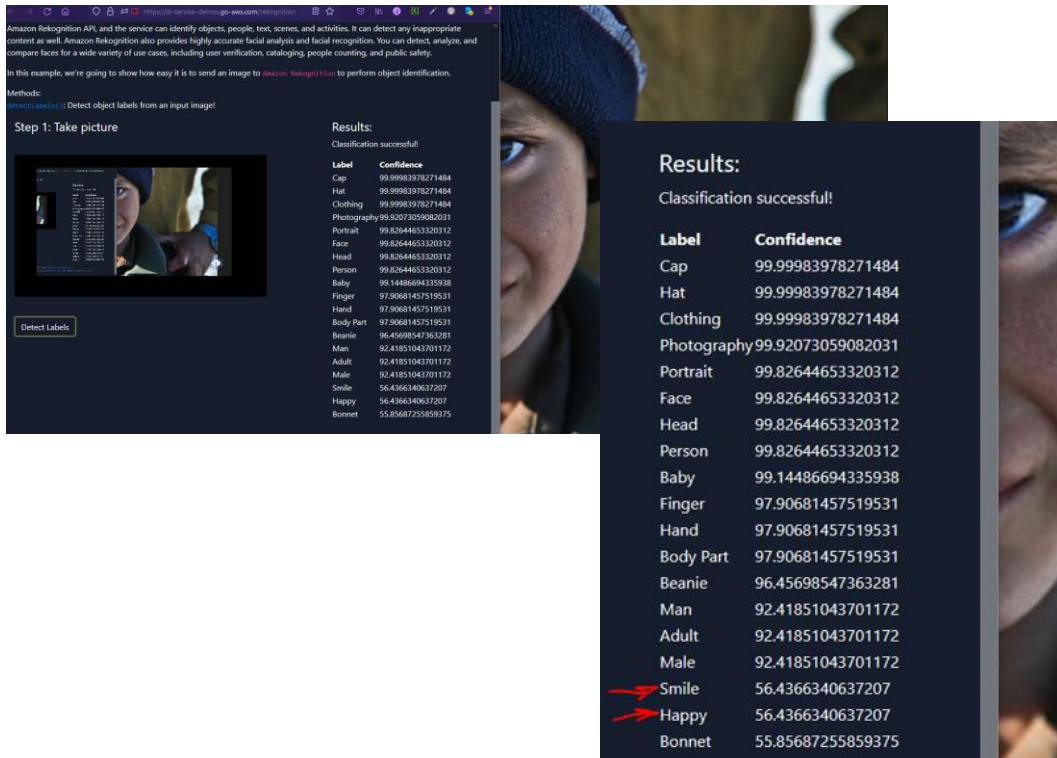
Classification successful!

Label	Confidence
Cap	99.99994659423828
Hat	99.99994659423828
Clothing	99.99994659423828
Finger	99.99993133544922
Person	99.99993133544922
Hand	99.99993133544922
Body Part	99.99993133544922
Photography	99.98265075683594
Portrait	99.97147369384766
Face	99.97147369384766
Head	99.97147369384766
Beanie	99.66898345947266
Baby	98.98002624511719
Baseball Cap	57.28898239135742
Bracelet	56.74080276489258
Jewelry	56.74080276489258
Accessories	56.74080276489258
Bonnet	55.15446472167969



Label	Confidence
Cap	99.99994659423828
Hat	99.99994659423828
Clothing	99.99994659423828
Finger	99.99993133544922
Person	99.99993133544922
Hand	99.99993133544922
Body Part	99.99993133544922
Photography	99.98265075683594
Portrait	99.97147369384766
Face	99.97147369384766
Head	99.97147369384766
Beanie	99.66898345947266
Baby	98.98002624511719
Baseball Cap	57.28898239135742
Bracelet	56.74080276489258
Jewelry	56.74080276489258
Accessories	56.74080276489258
Bonnet	55.15446472167969

Debo ampliar como en el anterior caso a pantalla completa para conseguir algo, aunque como en el caso de Google, AWS falla en la detección y nos muestra happy y Smile.



Amazon Rekognition API and the service can identify objects, people, text, scenes, and activities. It can detect any inappropriate content as well. Amazon Rekognition also provides highly accurate facial analysis and facial recognition. You can detect, analyze, and compare faces for a wide variety of use cases, including user verification, cataloging, people counting, and public safety.

In this example, we're going to show how easy it is to send an image to **Amazon Rekognition** to perform object identification.

Method:
[DetectLabels](#) Detect object labels from an input image!

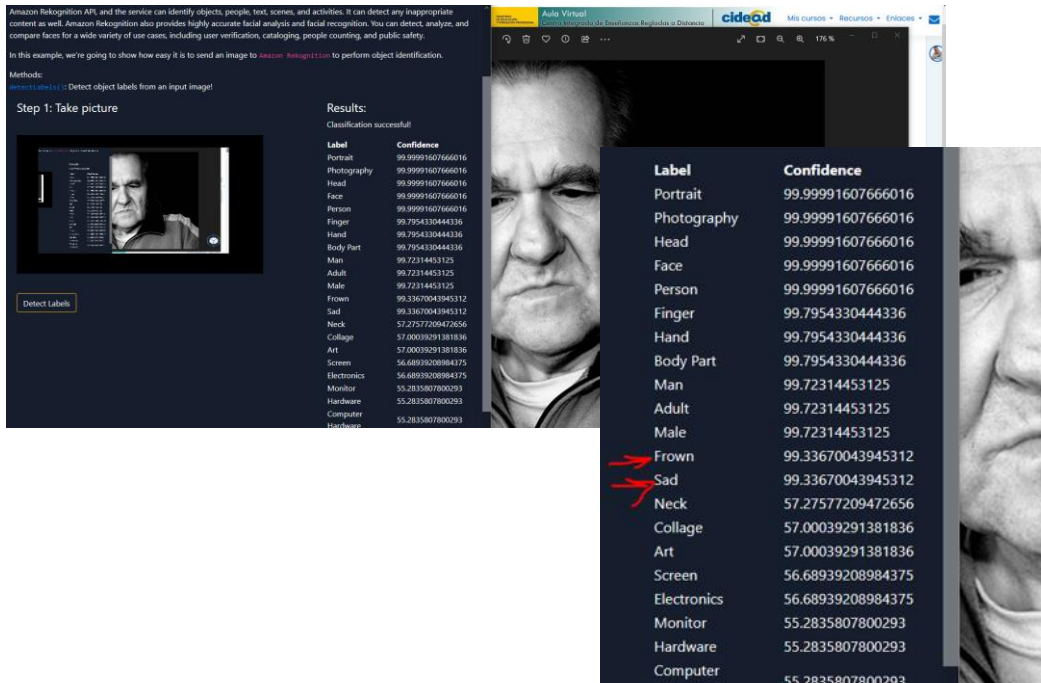
Step 1: Take picture

Detect Labels

Results:
Classification successful!

Label	Confidence
Cap	99.99983978271484
Hat	99.99983978271484
Clothing	99.99983978271484
Photography	99.92073059082031
Portrait	99.82644653320312
Face	99.82644653320312
Head	99.82644653320312
Person	99.82644653320312
Baby	99.14486694335938
Finger	97.90681457519531
Hand	97.90681457519531
Body Part	97.90681457519531
Beanie	96.45698547363281
Man	92.41851043701172
Adult	92.41851043701172
Male	92.41851043701172
Smile	56.4366340637207
Happy	56.4366340637207
Bonnet	55.85687255859375

Con la última imagen, el resultado es más claro desde la primera prueba.



Amazon Rekognition API and the service can identify objects, people, text, scenes, and activities. It can detect any inappropriate content as well. Amazon Rekognition also provides highly accurate facial analysis and facial recognition. You can detect, analyze, and compare faces for a wide variety of use cases, including user verification, cataloging, people counting, and public safety.

In this example, we're going to show how easy it is to send an image to **Amazon Rekognition** to perform object identification.

Method:
[DetectLabels](#) Detect object labels from an input image!

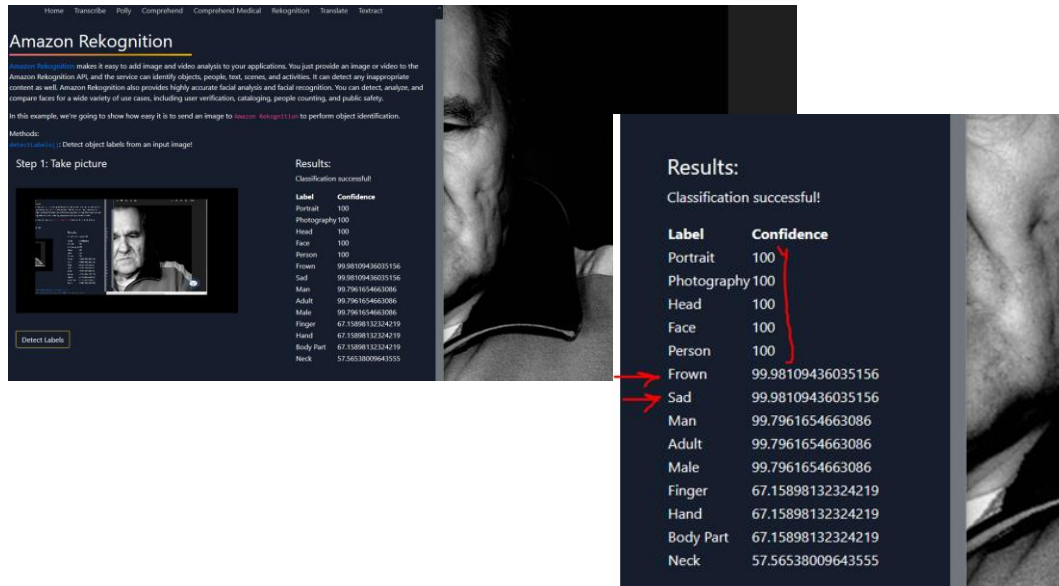
Step 1: Take picture

Detect Labels

Results:
Classification successful!

Label	Confidence
Portrait	99.99991607666016
Photography	99.99991607666016
Head	99.99991607666016
Face	99.99991607666016
Person	99.99991607666016
Finger	99.7954330444336
Hand	99.7954330444336
Body Part	99.7954330444336
Man	99.72314453125
Adult	99.72314453125
Male	99.72314453125
Frown	99.33670043945312
Sad	99.33670043945312
Neck	57.27577209472656
Collage	57.00039291381836
Art	57.00039291381836
Screen	56.68939208984375
Electronics	56.68939208984375
Monitor	55.2835807800293
Hardware	55.2835807800293
Computer	55.2835807800293

Si “limpiamos” un poco la captura para evitar que recoja datos como los navegadores que tenemos en la anterior, el resultado es mucho más concreto.



Amazon Rekognition

Amazon Rekognition makes it easy to add image and video analysis to your applications. You just provide an image or video to the Amazon Rekognition API and the service can identify objects, people, text, scenes, and activities. It can detect any inappropriate content as well. Amazon Rekognition also provides highly accurate facial analysis and facial recognition. You can detect, analyze, and compare faces for a wide variety of use cases, including user verification, cataloging, people counting, and public safety.

In this example, we're going to show how easy it is to send an image to Amazon Rekognition to perform object identification.

Methods:

`detectLabels()` Detect object labels from an input image

Step 1: Take picture

Results:

Classification successful!

Label	Confidence
Portrait	100
Photography	100
Head	100
Face	100
Person	100
Frown	99.98109436035156
Sad	99.98109436035156
Man	99.7961654663086
Adult	99.7961654663086
Male	99.7961654663086
Finger	67.15898132324219
Hand	67.15898132324219
Body Part	67.15898132324219
Neck	57.56538009643555

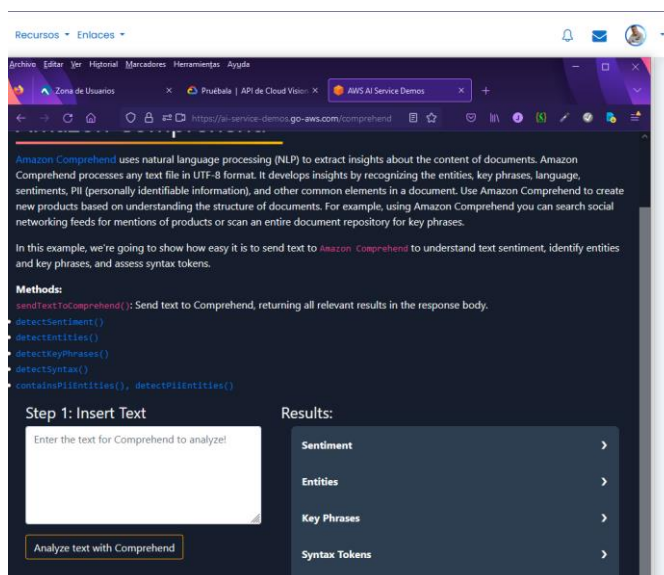
- Finalmente, escribe una reflexión en pocas líneas, indicando cuál sería la mejor herramienta para analizar que un cliente está satisfecho o feliz con el servicio prestado.

El hecho de no poder “inyectarle” una imagen a Rekognition y ver cómo según acotas el área de reconocimiento, acierta en mayor grado lo relativo a los rostros, creo que la hace una herramienta muy útil para campos donde quizás deba centrar su análisis en otras cuestiones, como la comparación y coincidencia u otros datos.

Sin embargo, para la cuestión concreta de analizar la satisfacción de un cliente, la forma que ha tenido en mi caso de analizar Vision AI cada imagen, separando cada rostro y haciendo un análisis individual y sencillo de interpretar, la hacen más idónea para esta cuestión respecto a AWS.

- ✓ **Apartado 2: Comparar dos herramientas de análisis de lenguaje natural para análisis de sentimiento.**

- Accede a la interfaz de prueba de AWS Comprehend a través de [este enlace](https://ai-service-demos.go-aws.com/comprehend).



Recursos **Enlaces**

Amazon Comprehend

Amazon Comprehend uses natural language processing (NLP) to extract insights about the content of documents. Amazon Comprehend processes any text file in UTF-8 format. It develops insights by recognizing the entities, key phrases, language, sentiments, PII (personally identifiable information), and other common elements in a document. Use Amazon Comprehend to create new products based on understanding the structure of documents. For example, using Amazon Comprehend you can search social networking feeds for mentions of products or scan an entire document repository for key phrases.

In this example, we're going to show how easy it is to send text to Amazon Comprehend to understand text sentiment, identify entities and key phrases, and assess syntax tokens.

Methods:

`sendTextToComprehend()` Send text to Comprehend, returning all relevant results in the response body.

`detectSentiment()`

`detectEntities()`

`detectKeyPhrases()`

`detectSyntax()`

`containsPIIEntities(), detectPIIEntities()`

Step 1: Insert Text

Enter the text for Comprehend to analyze

Analyze text with Comprehend

Results:

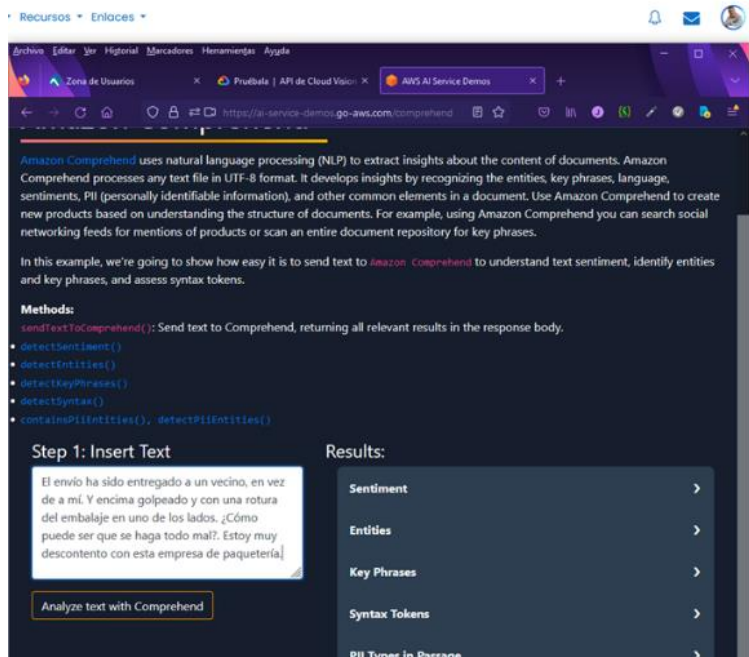
Sentiment

Entities

Key Phrases

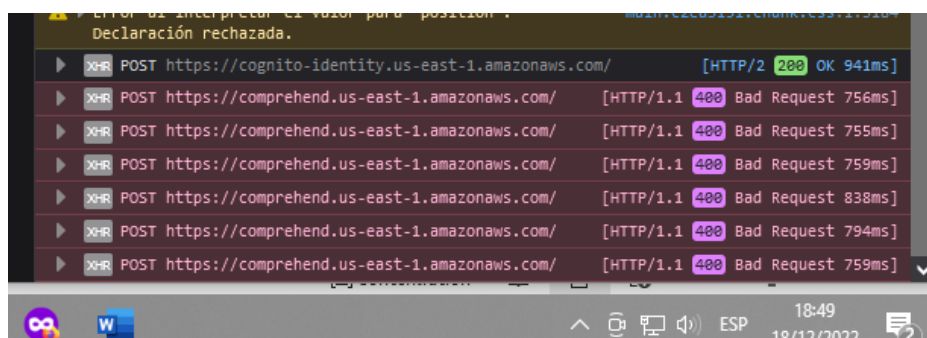
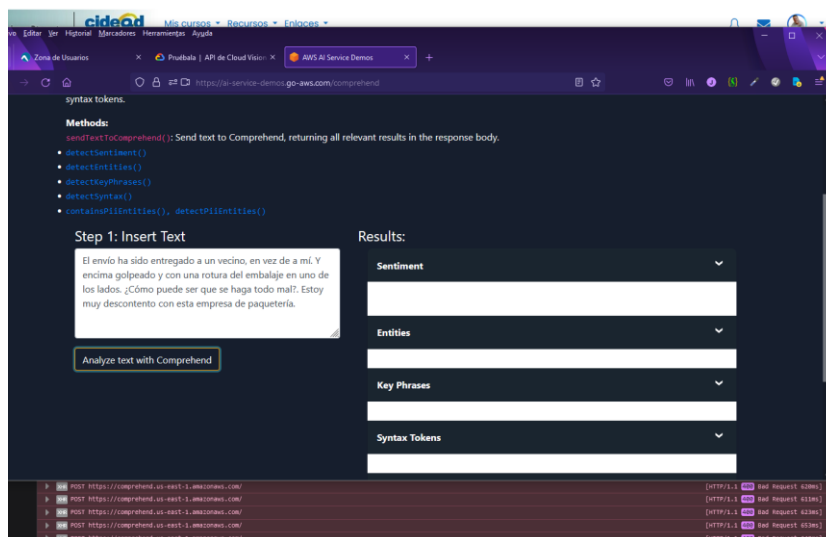
Syntax Tokens

- Escribe un comentario como si estuvieses descontento por cómo te han entregado un paquete que estaban esperando y pulsa el botón para analizarlo.

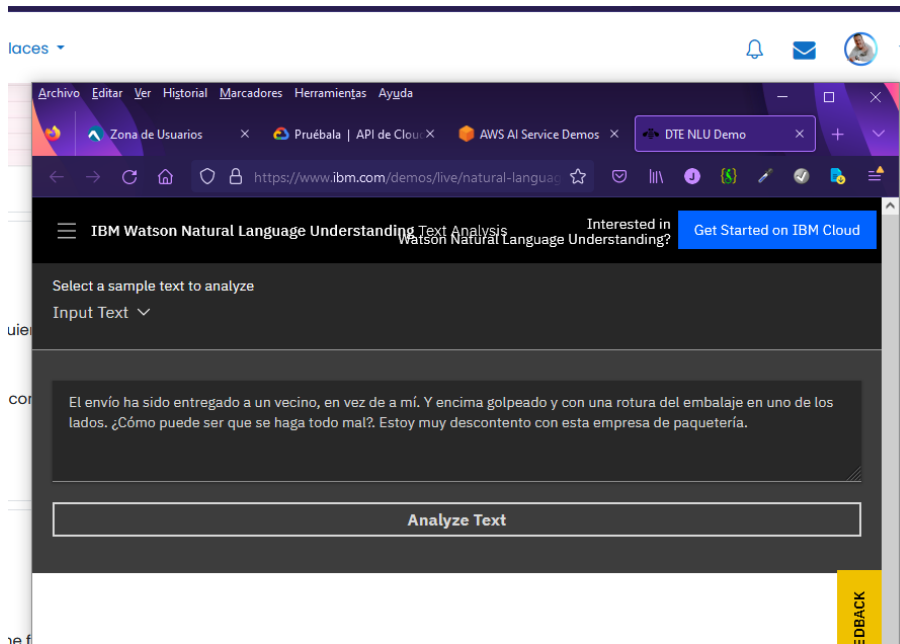


- Revisa los resultados que se muestran en el apartado "Sentiment" para comprobar si detecta negatividad y en qué grado.

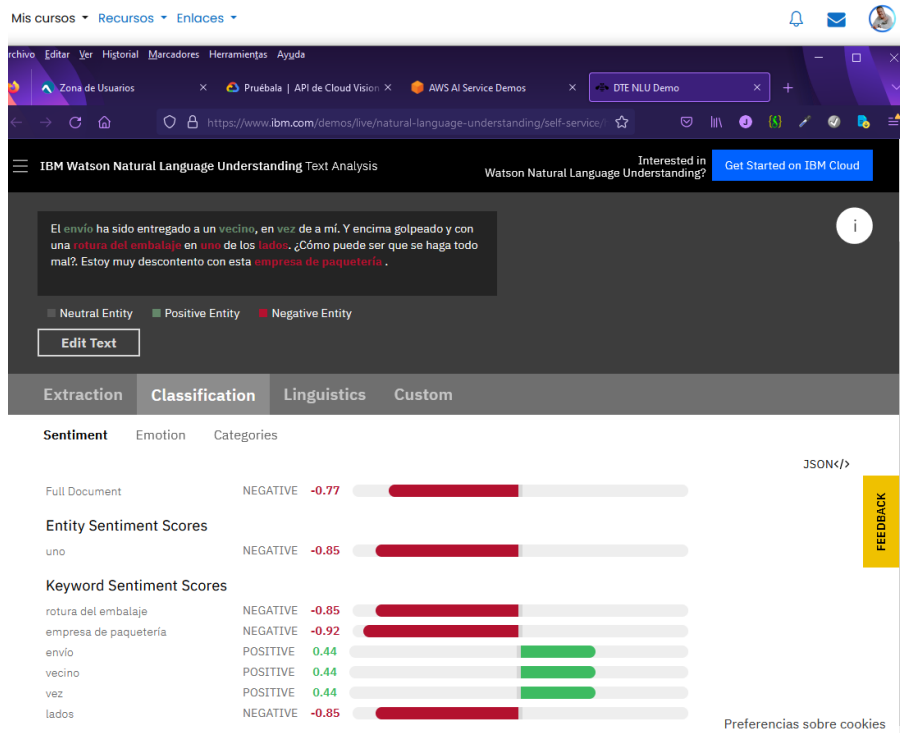
Tras varios intentos, el servicio, a 18/12/2022 no funciona y no devuelve datos por un error del propio servidor.



- Accede a la interfaz de prueba de IBM Watson a través de [este enlace](#).
- Pulsa en la opción "TRY YOUR OWN" > "Input Text" e introduce en el campo de texto el mismo comentario que en la interfaz anterior. Pulsa el botón para analizar dicho texto.



- Revisa los resultados en el apartado "Classification" > "Sentiment" para comprobar si detecta la negatividad del comentario y en qué grado.



The screenshot shows the IBM Watson Natural Language Understanding interface. The text input is: "El envío ha sido entregado a un vecino, en vez de a mí. Y encima golpeado y con una rotura del embalaje en uno de los lados. ¿Cómo puede ser que se haga todo mal?. Estoy muy descontento con esta empresa de paquetería." The sentiment analysis results are displayed as JSON:

```

{
  "sentiment": {
    "document": {
      "score": -0.769172,
      "mixed": "1",
      "label": "negative"
    }
  }
}

```

Legend: Neutral Entity (grey), Positive Entity (green), Negative Entity (red). Buttons: Edit Text, Classification, Sentiment, JSON, FEEDBACK.

Se observa como realiza el análisis de forma correcta (-0.77) y a pesar de no haber utilizado de forma voluntaria expresiones o lenguaje que pudieran facilitar el reconocimiento de este, e incluso lo asigna de forma adecuada a cada keyword detectada, como puede observarse también en la captura de extracción de datos.

The screenshot shows the IBM Watson Natural Language Understanding interface with the 'Extraction' tab selected. The text input is the same as in the previous screenshot. The keyword extraction results are displayed as a table:

Keyword	Relevance
rotura del embalaje	0.854163
empresa de paquetería	0.754269
envío	0.432277
vecino	0.30519

Buttons: Edit Text, Extraction, Classification, Linguistics, Custom, JSON, FEEDBACK.

- Finalmente, escribe una breve reflexión comparando ambos análisis y cuál podría ser la herramienta más adecuada para analizar el grado de satisfacción de clientes finales de una empresa que se dedica a la logística y reparto.

Lamentablemente, al no funcionar el servicio para AWS, no puedo comparar de forma fiel las dos herramientas.

Si nos fijamos para ello de la información e imágenes que nos muestra el temario, pueden observarse grandes similitudes en el uso de ambas.

Nos muestran las distintas recogidas de datos de forma similar, y en ambas disponemos de la información en formato JSON para evaluarlas.

Bajo estas pautas, yo me inclinaría por Watson, no solo por su bagaje, que presumo puede pesar muy a su favor el tiempo de aprendizaje, sino también por la propia interfaz con una representación gráfica más clara a mi parecer.