



Informe final del proyecto

Detección de fraude en los avisos de cobranzas de los servicios básicos

Diciembre 2024

Kauvin Lucas

1. Motivación

Con la creciente digitalización de los servicios financieros en el mundo, la identificación en línea de los consumidores mediante los medios electrónicos se ha hecho común, así como las preocupaciones por los fraudes por adulteración o alteración maliciosa de los documentos de verificación.

La motivación para el presente proyecto se centró en digitalizar el proceso de verificación del lugar de domicilio de los clientes de una institución bancaria boliviana que soliciten algún producto financiero mediante aplicación móvil en su teléfono inteligente.

En Bolivia, un método común para verificar el domicilio consiste en solicitar al cliente un aviso de cobranza de algún servicio básico (donde se muestre el nombre del cliente y la dirección de su vivienda) para luego contrastarla con la dirección declarada.

El hito que se abordó con el presente proyecto consistió en la detección de fraude durante el proceso de verificación domiciliaria empleando el enfoque de *computer vision*. Se diseñó un producto mínimo viable (MVP) con interfaz de programación de aplicaciones (API) que detecta indicios de alteración o manipulación de los avisos de cobranzas subidos en línea en tiempo real.

2. Desarrollo del proyecto

2.1. Requerimientos de negocio

Se requirió la construcción de una aplicación de Inteligencia Artificial que sea capaz de analizar los avisos de cobranzas para detectar los posibles patrones de manipulación o alteración en los documentos. La solución debe cumplir con los siguientes criterios de aceptación:

- **Criterio 1:** API consumible que procese y detecte fraude en los documentos en tiempo real;
- **Criterio 2:** tiempo de respuesta máximo por fotografía de hasta 1 minuto;
- **Criterio 3:** exactitud (*accuracy*) por lo menos 75% sobre los datos de prueba.

Para el presente proyecto, se asumió que todas las personas sujetas al proceso de verificación de domicilio subirían la fotografía de su aviso de cobranza siguiendo las siguientes instrucciones:

- Las 4 esquinas del documento deben ser claramente visibles y contrastar con el fondo;

- No subir fotocopia o escaneado del documento;
- Todos los textos en el documento deben ser claramente legibles;
- La iluminación debe ser clara, pero evitando deslumbramientos en el documento;
- El documento no debe presentar los siguientes defectos: tachaduras, mutilaciones y perforaciones;
- El documento debe ser reciente (fecha de emisión dentro de los últimos 3 meses).

Por lo tanto, la verificación de los puntos anteriores estaría fuera de alcance del MVP.

2.2. Conjunto de datos

Uno de los grandes desafíos del proyecto fue la obtención o generación del conjunto de datos. Como los avisos de cobranzas contienen información de identificación personal (PII, en sus siglas en inglés), no suelen ser accesibles y su obtención fue considerada inviable.

Como alternativa, se buscó generar sintéticamente un conjunto de datos compuesto por imágenes de avisos de cobranzas similares a los reales.

Como primer paso, se diseñaron los templates “reales” (*ground truth*) que serían la base para generar los avisos de cobranzas no manipulados, y los templates falsos (*fakes*) para la generación de los documentos con indicios de fraude.

Los templates falsos buscaron incorporar algunos de los posibles indicios de fraude comunes, como:

- Diseño antiguo o en desuso,
- Logos antiguos o pegados desde otra fuente,
- Fuente y tamaño de texto inadecuados,
- Diseño replicado digitalmente y
- Colores y marcas de agua inconsistentes.

Luego, se construyó un script denominado `generate.py` para automatizar la generación del contenido de texto y su impresión en los templates.

AVISO DE COBRANZA POR CONSUMO DE ENERGÍA ELÉCTRICA – COOPERATIVA DE ELECTRIFICACIÓN DE EJEMPLO R.L.

MES DE FACTURACIÓN: Septiembre 2024
FECHA EMISIÓN: 05/09/24
FECHA VENCIMIENTO: 07/10/24
FECHA PRÓXIMA MEDICIÓN: 06/10/24
FECHA PRÓXIMA EMISIÓN: 06/10/24

CÓDIGO FIJO: 645711

NOMBRE: T - PERSONA TEST BARBERY
DIRECCIÓN: AVENIDA TERCER ANILLO EXTERNO, CASA 711
UV: 0045-
DIST: 4

NRO 711
CIUDAD: SANTA CRUZ DE LA SIE
MZA: 0003- SECCIÓN: 1

Nº MEDIDOR: MEDI005932
TIPO DE MEDICIÓN: Leído
CAUSA: DOMICILIARIA-PD-BT
CATEGORÍA: 01.001.
CÓDIGO DE UBICACIÓN: 045.00008515.0035

ESCÁNEA PARA PAGAR

CORTE EN PROCESO

FECHA			FECHA			FECHA			FECHA		
ACTUAL	ANTERIOR	DÍAS CONSUMO	ACTUAL	ANTERIOR	FACTOR	CONSUMO	CONS. ACUMULADO	CONSUMO TOTAL			
05/09/24	07/08/24	29	34472	34413	1.000	59	0	59			

PERÍODO	CONSUMO-KWH	IMPORTE-Bs	ESTADO
09-24	59	44.30	Impaga
08-24	75	72.30	IMPAGA
07-24	59	115.10	VENCIDA
06-24	49	41.00	18/06/2024
05-24	76	72.60	18/06/2024
04-24	89	83.50	18/06/2024
03-24	101	80.90	27/03/2024
02-24	91	72.00	27/03/2024
01-24	88	72.00	27/03/2024
12-23	126	117.40	06/12/2023
11-23	86	78.50	06/12/2023
10-23	86	151.50	06/12/2023
09-23	97	87.60	06/12/2023

FACTURACIÓN - CRE		TASAS MUNICIPALES - ALCALDÍA MUNICIPAL	
DETALLE	IMPORTE - Bs	DETALLE	IMPORTE - Bs
Importe por energía	46.70	Aseo Urbano	8.00
Benef. x T. Dignidad	-11.70	Alumbrado publico	4.10
Rest control calidad	-3.30		
Tasa AFCOOP	0.50		
TOTAL	32.20	TOTAL	12.10

CORTE A PARTIR DE: 12/09/24

INFORMACIÓN AL CONSUMIDOR:
Los reclamos por el servicio deben ser presentados por el consumidor o un tercero en el sistema TEST de CEP, de forma verbal, escrita o comunicarse al teléfono gratuito 175 o al número de emergencia 3351574, dentro de los 20 días hábiles de conocido el hecho.
En caso de no obtener respuesta en el plazo establecido o que ésta no sea satisfactoria para usted, puede acudir a la Autoridad de Fiscalización de Electricidad y Tecnología Nuclear o comunicarse a la línea gratuita 800-10-2407. ESTA EMPRESA ES FISCALIZADA, CONTROLADA, SUPERVISADA Y REGULADA POR LA AUTORIDAD DE FISCALIZACIÓN DE ELECTRICIDAD Y TECNOLOGÍA NUCLEAR (ARTN). LÍNEA GRATUITA 800-10-2407

TOTAL A CANCELAR 44.30

EVITE EL CORTE CANCELANDO SUS FACTURAS PUNTUALMENTE

Imagen 1 - Ejemplo de un aviso de cobranza real (con template ground truth)

AVISO DE COBRANZA POR CONSUMO DE ENERGÍA ELÉCTRICA – COOPERATIVA DE ELECTRIFICACIÓN DE EJEMPLO R.L.

MES DE FACTURACIÓN: Noviembre 2024
FECHA EMISIÓN: 08/11/24
FECHA VENCIMIENTO: 10/12/24
FECHA PRÓXIMA MEDICIÓN: 09/12/24
FECHA PRÓXIMA EMISIÓN: 09/12/24

CÓDIGO FIJO: 610027

NOMBRE: LAURA MORALES LOPEZ FERNANDEZ
DIRECCIÓN: BARRIO ESTACIÓN ARGENTINA CALLE EL CARMEN, UV: 0021
DIST: 6

DEPTO: 16
CIUDAD: SANTA CRUZ DE LA SIERRA
MZA: 0001 SECCION: 1

Nº MEDIDOR: MEDI7874
TIPO DE MEDICIÓN: Leído
CAUSA: DOMICILIARIA-PD-BT
CATEGORÍA: 01.001.
CÓDIGO DE UBICACIÓN: 026.00007186.0077

ESCÁNEA PARA PAGAR

CORTE EN PROCESO

FECHA			FECHA			FECHA			FECHA		
ACTUAL	ANTERIOR	DÍAS CONSUMO	ACTUAL	ANTERIOR	FACTOR	CONSUMO	CONS. ACUMULADO	CONSUMO TOTAL			
08/11/24	10/10/24	29	3734	3891	1.10	513	0	513			

PERÍODO	CONSUMO-KWH	IMPORTE-Bs	ESTADO
11-24	513	563.04	Impaga
10-24	670	735.70	03/11/2024
09-24	867	952.39	06/08/2024
08-24	943	1034.80	28/06/2024
07-24	907	995.81	28/06/2024
06-24	890	977.10	04/04/2024
05-24	889	976.03	04/04/2024
04-24	876	961.99	06/04/2024
03-24	676	742.56	06/03/2024
02-24	717	787.68	04/02/2024
01-24	690	757.08	23/12/2023
12-23	531	583.49	01/12/2023
11-23	731	802.68	01/12/2023

FACTURACIÓN - CRE		TASAS MUNICIPALES - ALCALDÍA MUNICIPAL	
DETALLE	IMPORTE - Bs	DETALLE	IMPORTE - Bs
Importe por energía	563.04	Aseo Urbano	7.63
Benef. x T. Dignidad	-11.70	Alumbrado publico	5.14
Rest control calidad	-3.30		
Tasa AFCOOP	0.50		
TOTAL	548.54	TOTAL	12.77

CORTE A PARTIR DE: 10/11/24


INFORMACIÓN AL CONSUMIDOR:
Los reclamos por el servicio deben ser presentados por el consumidor o un tercero en el sistema TEST de CEP, de forma verbal, escrita o comunicarse al teléfono gratuito 175 o al número de emergencia 3351574, dentro de los 20 días hábiles de conocido el hecho.
En caso de no obtener respuesta en el plazo establecido o que ésta no sea satisfactoria para usted, puede acudir a la Autoridad de Fiscalización de Electricidad y Tecnología Nuclear o comunicarse a la línea gratuita 800-10-2407. ESTA EMPRESA ES FISCALIZADA, CONTROLADA, SUPERVISADA Y REGULADA POR LA AUTORIDAD DE FISCALIZACIÓN DE ELECTRICIDAD Y TECNOLOGÍA NUCLEAR (ARTN). LÍNEA GRATUITA 800-10-2407

TOTAL A CANCELAR 561.31

EVITE EL CORTE CANCELANDO SUS FACTURAS PUNTUALMENTE

Imagen 2 - Ejemplo de un aviso de cobranza falso (con template fake)

En algunas imágenes, también se aplicaron las siguientes manipulaciones para simular los otros posibles escenarios de fraude, como el tachado y reemplazo del texto en las secciones de nombre completo y dirección domiciliaria en el ejemplo a continuación.



AVISO DE COBRANZA
 Por Servicios de Agua Potable y Alcantarillado Sanitario
 Cooperativa de Alcantarillado Santa Cruz RL "COOPACRUZ"
 Av. San Aurelio No 1547 Teléfono 362-2324
 NO VALIDO PARA CREDITO FISCAL

CÓDIGO ASOCIADO
35209788

NOMBRE: ROBERT IBÁÑEZ GONZALES RODRIGUEZ

DIRECCIÓN: CALLE COMBATE BAGRE, N/A, SANTA CRUZ DE LA SIERRA

DISTRITO:

FECHA LECTURA	LECTURA	CONSUMO m³	OBS.
05/09/2024	811	2	6.22
05/08/2024	811	2	6.22

DEUDA ATRASADA Bs: 160.38

N° DE MESES: 11

CORTE
 SI
SI PAGO, IGNORE ESTE MENSAJE

MES	CONSUMO m³	IMPORTE Bs	ESTADO
2024-09	2	11.78	
2024-08	2	12.91	
2024-07	3	16.96	
2024-06	3	17.20	
2024-05	3	15.98	
2024-04	3	15.92	
2024-03	3	17.74	
2024-02	3	15.64	
2024-01	2	11.67	
2023-12	2	10.39	
2023-11	2	11.45	
2023-10	2	14.53	


ITEM	DETALLE	IMPORTE Bs
1	SERV. AGUA POTABL	11.78
2	SERV. ALC. SANITA	4.51
3	APORTE DE SOCIO	1.96

INFORMACIÓN AL CONSUMIDOR

* Amigo Cooperativista, para pagar su factura diríjase a nuestros Agentes de Cobranza autorizados, indicados al reverso del presente.

* El reclamo por alguna factura podrá ser realizado hasta la fecha de vencimiento de ésta (30 días posterior a la fecha de su emisión)

Podrás escanear y pagar con código QR ->



TOTAL DEUDA Bs: 172.16

IMPORTE TOTAL FACTURA Bs: 18.25

Imagen 3 - Ejemplo de un aviso de cobranza donde se evidencia el tachado y reemplazo de texto en las secciones “nombre” y “dirección”

Las imágenes generadas y/o manipuladas fueron impresas, recortadas y fotografiadas con diferentes modelos de teléfonos inteligentes, variando en perspectiva y en los niveles de iluminación y resolución. En total, se generaron 400 imágenes, de las cuales 200 correspondían a documentos manipulados (fraude) y las otras 200 a los documentos no manipulados (no fraude).

2.3. Procesamiento de los datos



Se subieron las fotografías al Google Drive y se utilizó el notebook `preprocess.ipynb` en Google Colaboratory (Colab) para aplicar las siguientes transformaciones a las fotografías mediante el script `image_transform.py`:

- Redimensionamiento de la imagen,
- Identificación de las esquinas y recortado y

- Transformación de perspectiva.

Una vez identificados los contornos del documento en cada imagen, se procedió con el recorte de cada documento. El 89% de estas imágenes transformadas fue almacenado en la carpeta denominada “train”, y el 11% fue almacenado en “test”.

Las fotografías almacenadas en la carpeta “train” fueron utilizadas como el conjunto de datos de entrada del modelo para la etapa de experimentación.

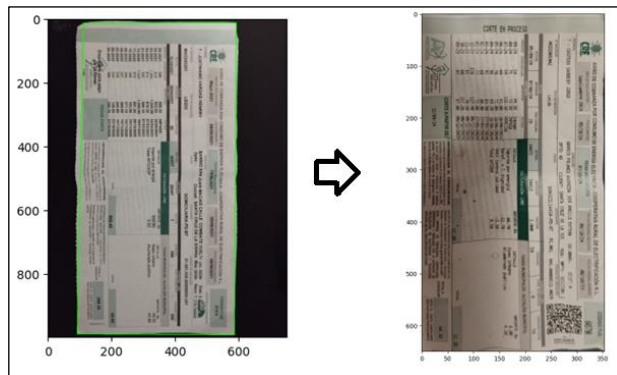


Imagen 4 – Identificación de los bordes del aviso de cobranza y recorte del documento

2.4. Experimentación



Para la etapa de experimentación, se utilizó el notebook `train.ipynb` en Colab, en un ambiente virtual con GPU T4 de 16 GB de RAM. El modelo fue construido con Keras del Tensorflow. La métrica utilizada durante el entrenamiento fue la **exactitud** (accuracy) y se utilizó el **Cross Entropy** como función de pérdida (SparseCategoricalCrossentropy).

El conjunto de datos de entrada, compuesto por 356 imágenes, fue dividido entre 80% datos de entrenamiento y 20% datos de validación.

El modelo construido tiene la clásica arquitectura de una red neuronal convolucional, conformada por 3 capas convolucionales con capas entremedias de *Max Pooling*, seguidas por una capa oculta lineal completamente conectada y una capa de salida, con funciones de salida *ReLU* en cada capa (salvo la última). Como optimizador, se optó por

Adaptive Moment Estimation (Adam). Se utilizó el regularizador *Dropout* entre las últimas capas.

Layer (type)	Output Shape	Param #
rescaling_1 (Rescaling)	(None, 500, 1024, 3)	0
conv2d (Conv2D)	(None, 498, 1022, 16)	448
max_pooling2d (MaxPooling2D)	(None, 249, 511, 16)	0
conv2d_1 (Conv2D)	(None, 247, 509, 32)	4,640
max_pooling2d_1 (MaxPooling2D)	(None, 123, 254, 32)	0
conv2d_2 (Conv2D)	(None, 121, 252, 64)	18,496
max_pooling2d_2 (MaxPooling2D)	(None, 60, 126, 64)	0
dropout (Dropout)	(None, 60, 126, 64)	0
flatten (Flatten)	(None, 483840)	0
dense (Dense)	(None, 256)	123,863,296
dropout_1 (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 2)	514

Total params: 123,887,394 (472.59 MB)

Trainable params: 123,887,394 (472.59 MB)

Imagen 5 – Resumen del modelo construido

Se entrenó el modelo en lotes de datos de 32 imágenes, haciendo un recorrido en todo el conjunto de datos en un total de 6 veces. La exactitud alcanzada fue del 97.18% sobre los datos de validación.

```

Epoch 1/6
9/9 ██████████ 65s 5s/step - accuracy: 0.4477 - loss: 19.9000 - val_accuracy: 0.4366 - val_loss: 1.0177
Epoch 2/6
9/9 ██████████ 27s 901ms/step - accuracy: 0.6219 - loss: 0.7476 - val_accuracy: 0.8732 - val_loss: 0.3895
Epoch 3/6
9/9 ██████████ 6s 674ms/step - accuracy: 0.8488 - loss: 0.3249 - val_accuracy: 0.8732 - val_loss: 0.3414
Epoch 4/6
9/9 ██████████ 12s 885ms/step - accuracy: 0.9105 - loss: 0.2146 - val_accuracy: 0.9437 - val_loss: 0.2289
Epoch 5/6
9/9 ██████████ 10s 923ms/step - accuracy: 0.9494 - loss: 0.1350 - val_accuracy: 0.7606 - val_loss: 0.4060
Epoch 6/6
9/9 ██████████ 8s 851ms/step - accuracy: 0.9009 - loss: 0.2164 - val_accuracy: 0.9718 - val_loss: 0.1286

```

Imagen 6 – Métricas reportadas durante el entrenamiento

A continuación, se muestran las curvas de exactitud y error reportados por el modelo entre durante el entrenamiento entre datos de entrenamiento y de validación, por cada recorrido en el conjunto de datos (eje horizontal).

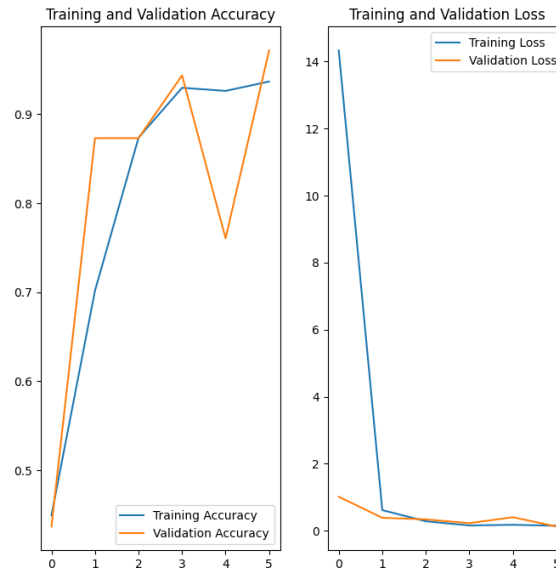


Imagen 7 – Gráficos de exactitud y error entre datos de entrenamiento y validación

2.5. Medición

El modelo se midió sobre las 44 fotografías transformadas en la carpeta “test”, reportando una exactitud del 93,37%.

2.6. Inferencia

Para la inferencia, se construyó el script `serve.py` para desplegar una aplicación con API diseñada con el marco FastAPI. Dicha aplicación, mediante el endpoint `invocations`, recibiría una imagen, la preprocesaría (identificando y recortando los bordes del aviso de cobranza) y aplicaría el modelo para clasificarlo entre fraude y no fraude, retornando el resultado y el porcentaje de confianza del modelo. El tiempo de respuesta es de 4 a 5 segundos en un servidor local.

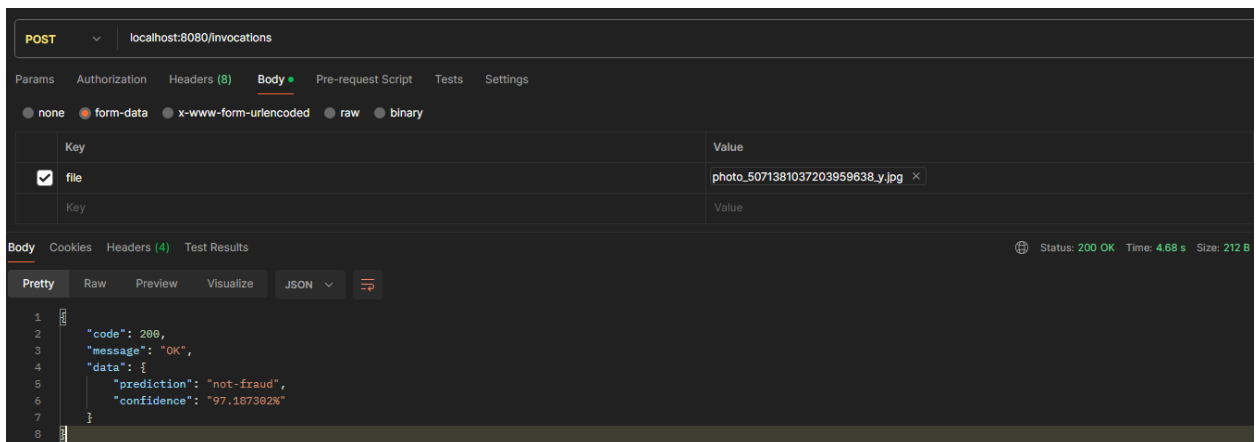
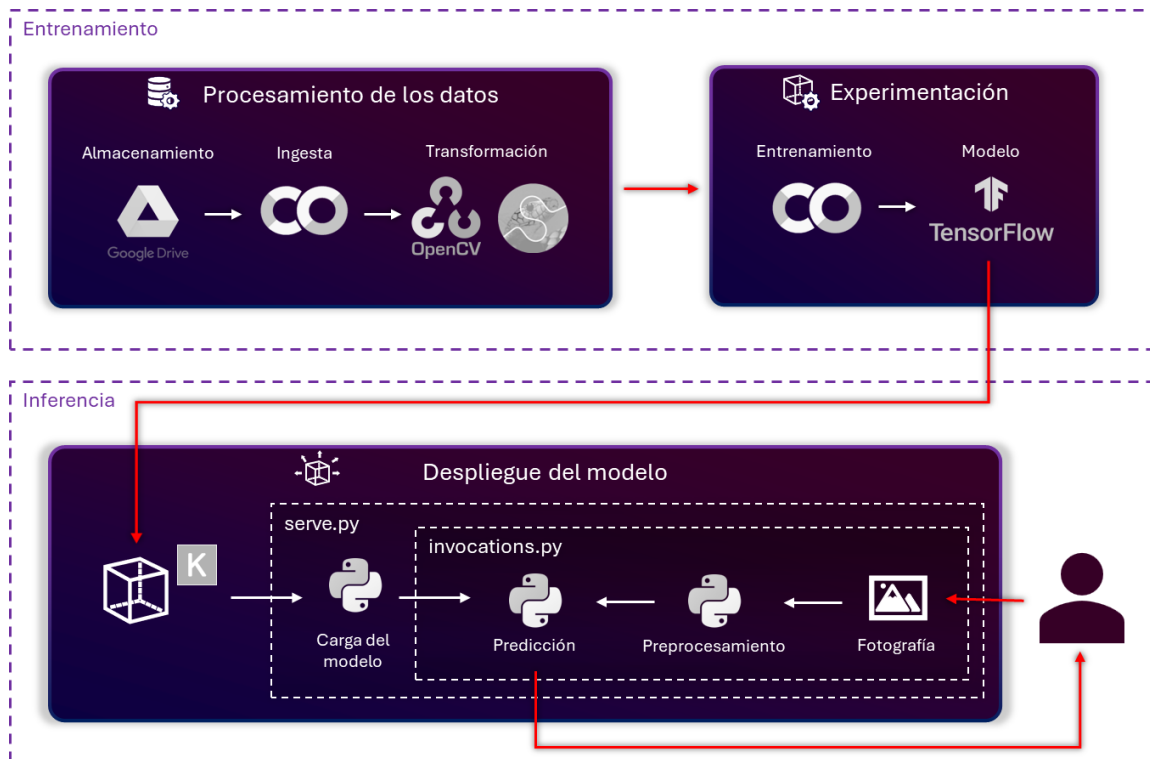


Imagen 7 – Ejemplo de respuesta del endpoint `invocations`

3. Arquitectura final de la solución



4. Línea de tiempo de la implementación de la solución

	Sprint				
	1	2	3	4	5
	23/09 - 06/10	07/10 - 20/10	21/10 - 03-11	04/11 - 17/11	18/11 - 01/12
Requerimientos de negocio					
Arquitectura inicial de la solución					
Generación del conjunto de datos					
Experimentación					
Empaquetado del código					
Despliegue					
Informe final					

5. Enlaces

- Repositorio del proyecto: [GitHub](#)