**Implementación de Aplicaciones de Aprendizaje Automático en la Nube**

Identificación de etiquetas de medicamentos a través de imágenes

**ITBA - Master in Management & Analytics**

Profesor: Ramiro Savoie

1 de Octubre 2024

**Equipo 3:**

Gonzalo Martínez Ruiz Moreno

Matias Premio Prado

Maximiliano Coltro

Santiago Granda

**Índice:**

[**Bloque 1: Propuesta de valor 5**](#_heading=h.30j0zll)

[1. ¿A qué objetivo de negocio estamos sirviendo? 5](#_heading=h.wsjmep68bao9)

[2. ¿Quién es el usuario final de nuestro sistema? 5](#_heading=h.wvpuu6mpqo77)

[3. ¿Cuál es el problema que estamos intentando resolverle? 5](#_heading=h.pjenwl5aos4c)

[4. ¿Por qué es importante hacerlo? 5](#_heading=h.eyx2mea5bu5i)

[5. ¿Qué valor le entrega el proyecto de ML al usuario final? 6](#_heading=h.u6xzm05i7dg5)

[6. ¿Cómo va a usar nuestras salidas o predicciones? 6](#_heading=h.ec6dwqakl4ry)

[**Bloque 2: Fuentes de datos 7**](#_heading=h.1fob9te)

[**1. ¿Con qué datos históricos vamos a alimentar el modelo? 7**](#_heading=h.641795jroa3w)

[Fuentes de datos internas: 7](#_heading=h.6p96zc27qd6c)

[Fuentes de datos externas: 7](#_heading=h.bb2txcy5djo9)

[2. ¿Cuáles son los costos ocultos de la aplicación de ML? 7](#_heading=h.pfmg46sz69jm)

[3. ¿Cuánto podrían crecer los datos? 8](#_heading=h.ulzw881noj8l)

[4. ¿Qué tan costoso se puede volver el almacenamiento de los datos? 9](#_heading=h.3qr33jk5ff42)

[**Bloque 3: Tarea de predicción 9**](#_heading=h.3znysh7)

[1. ¿Es supervisada o no supervisada? 9](#_heading=h.dd9zytj1zjrq)

[2. ¿Es detección de anomalías? 9](#_heading=h.4eawsirbv00b)

[3. ¿El problema es sobre qué opción debería ser tomada? 9](#_heading=h.izhgrrfqsskk)

[4. ¿Necesitamos predecir un valor continuo? 10](#_heading=h.a5o9msl2t0l1)

[5. ¿Qué categoría debe ser predicha? ¿Es binaria o multiclase? 10](#_heading=h.qao9avrb0bjo)

[6. ¿Necesitamos agrupar nuestra data? 10](#_heading=h.kennc0ynlqse)

[7. ¿Cuál sería la definición de un sample? 10](#_heading=h.145386dhwir9)

[8. ¿Será un modelo de Machine Learning o de Deep Learning? 10](#_heading=h.k1zchw7drnal)

[**Bloque 4: Ingeniería de features 11**](#_heading=h.2et92p0)

[1. ¿Cómo se extraen los features de las fuentes crudas? 11](#_heading=h.multg7e7qwox)

[2. ¿Hay que cruzar muchas fuentes de datos? 11](#_heading=h.o71pugm40zjn)

[3. ¿Hay que hacer un procesamiento complejo para que sean útiles? 11](#_heading=h.sbai6hg1c0as)

[4. ¿Hay personal o herramientas para hacer esta labor? 12](#_heading=h.xga6uy230a3t)

[5. ¿Se considera incluir a expertos del dominio para especificar qué aspectos de la data son los más importantes para esta tarea de ML en particular? 12](#_heading=h.p97h30vbkjr0)

[**Bloque 5: Evaluación offline 12**](#_heading=h.tyjcwt)

[1. Métricas específicas del dominio que justifiquen el desarrollo del modelo 12](#_heading=h.8unqa52uo3o0)

[2. Objetivo del negocio 13](#_heading=h.e6hlzylqv1kq)

[3. ¿Cuál será la performance mínima con la que se autorizará la puesta productiva del modelo? 13](#_heading=h.eelg9xi7sjka)

[4. ¿Cuáles son las consecuencias medibles en los errores de predicción del modelo como los Falsos Positivos y los Falsos Negativos? 13](#_heading=h.q9ir5ouoi4qh)

[**Bloque 6: Toma de decisiones 14**](#_heading=h.3dy6vkm)

[1. ¿Qué decisiones se toman con las predicciones del sistema? 14](#_heading=h.qclxgyivau34)

[2. ¿Estas decisiones proveen el valor propuesto al usuario final? 14](#_heading=h.pw1nefa18kcu)

[3. ¿Cómo interactúa el usuario final u otro sistema con las predicciones del modelo? 14](#_heading=h.i5kbmlp9vexw)

[4. ¿Cuáles son los costos ocultos en la toma de decisiones como los human-in-the-loop? 14](#_heading=h.475b7tj4hrtu)

[**Bloque 7: Realizando predicciones 15**](#_heading=h.1t3h5sf)

[1. ¿Cuándo deberían estar las predicciones disponibles? 15](#_heading=h.ajjgc7tohyqo)

[2. ¿Nuevas predicciones son hechas a demanda? 15](#_heading=h.j8r51gvhg3u6)

[3. ¿Nuevas predicciones son calendarizadas? 15](#_heading=h.8nh3v8acd8ja)

[4. ¿Las predicciones son hechas sobre la marcha en cada punto de datos o para un batch de datos de entrada? 16](#_heading=h.3dan1u7cc16n)

[5. ¿Hay un human-in-the-loop en la salida de estas predicciones? 16](#_heading=h.3fl71kp873we)

[6. ¿Se dispone de hardware para predecir? 16](#_heading=h.j4t8a4yt1azg)

[7. ¿Se utiliza algún servicio de Cloud para predecir? 16](#_heading=h.tj8hap9f3qjv)

[**Bloque 8: Recolectando datos 17**](#_heading=h.4d34og8)

[1. ¿Se dispone de datos para hacer un entrenamiento inicial? 17](#_heading=h.bz1k4d1q8ky5)

[2. ¿Cómo etiquetamos los datos nuevos? 17](#_heading=h.6rl5oetrr1b9)

[3. ¿Hay que procesar datos multimedia de tipo imagen, sonido o video? 17](#_heading=h.lfs2urga0uyp)

[4. ¿Hay un human-in-the-loop en la limpieza manual y el etiquetado de la data entrante? 17](#_heading=h.vwo7d7uni1s3)

[**Bloque 9: Construyendo modelos 18**](#_heading=h.2s8eyo1)

[1. ¿Qué tan seguido debería ser reentrenado el modelo? A cada hora, cada semana, o con cada punto nuevo de datos que ingresa. Por ejemplo, por inflación. 18](#_heading=h.a6ih2cuvpgsi)

[2. ¿Se dispone de hardware para entrenar? 18](#_heading=h.w9lrpz2jb6w0)

[3. ¿Se utiliza algún servicio de Cloud para entrenar? 18](#_heading=h.rltsfqlo9ww6)

[4. ¿Cómo vamos a lidiar con los asuntos de escala en la medida que la operación se vuelva más compleja y costosa? 19](#_heading=h.nitg5dlbeecg)

[5. ¿Cuál es el stack de tecnologías usado? 19](#_heading=h.xc1zyakssm6y)

[**Bloque 10: Monitoreo y evaluación en vivo 19**](#_heading=h.17dp8vu)

[1. ¿Cómo vamos a hacer seguimiento de la performance del sistema? 19](#_heading=h.652voh49jac3)

[2. ¿Cómo evaluamos la creación de valor? 20](#_heading=h.akmxnmn0voxm)

# 

# **Bloque 1: Propuesta de valor**

Para las farmacias y empresas de medicamentos quienes necesitan optimizar la gestión de stock y reducir errores en la entrega de los mismos, nuestro modelo de clasificación es un identificador de imágenes que permite automatizar el proceso de selección y disponibilidad de los medicamentos de forma segura

### 1. ¿A qué objetivo de negocio estamos sirviendo?

El objetivo que estamos sirviendo es el de optimizar la gestión de stock y reducir errores en la entrega de medicamentos en farmacias. Esto se traduce en una mejora de la eficiencia operativa, reducción de tiempos de espera y una disminución de riesgos relacionados con la incorrecta distribución de medicamentos. Adicionalmente, el sistema también busca mejorar la seguridad del paciente y la satisfacción de los profesionales de la salud.

### 2. ¿Quién es el usuario final de nuestro sistema?

El usuario final de nuestro sistema son los responsables de la gestión del stock de medicamentos en farmacias, como farmacéuticos o personal de farmacia. También se pueden beneficiar otros actores en la cadena de suministro farmacéutico, como almacenes o centros de distribución, donde el stock de medicamentos necesita ser gestionado con precisión.

### 3. ¿Cuál es el problema que estamos intentando resolverle?

Estamos resolviendo el problema de los errores humanos en la identificación y entrega de medicamentos. Esto incluye situaciones donde un medicamento almacenado de forma incorrecta se transporta al farmacéutico o se entrega al paciente, lo que puede generar pérdidas de tiempo, deterioro de los medicamentos, e incluso riesgos para la salud del paciente si se entregan productos no aptos para el consumo.

### 4. ¿Por qué es importante hacerlo?

Es importante hacerlo porque los errores en la entrega de medicamentos pueden tener graves consecuencias, tanto económicas como en la salud de los pacientes. Un sistema que automatice la verificación de los medicamentos puede evitar devoluciones, minimizar pérdidas por deterioro de productos sensibles, y reducir la posibilidad de causar daño a los pacientes. Además, mejora la productividad del personal farmacéutico y aumenta la satisfacción del cliente al recibir un servicio más eficiente y confiable.

### 5. ¿Qué valor le entrega el proyecto de ML al usuario final?

El valor clave es la reducción de errores en la identificación de medicamentos, lo que optimiza la gestión del stock y permite una entrega más rápida y precisa. El sistema de deep learning añade una capa de seguridad que evita que el farmacéutico entregue un medicamento incorrecto, lo cual ahorra tiempo y recursos. Además, mejora la fiabilidad del servicio y disminuye el riesgo de errores humanos.

### 6. ¿Cómo va a usar nuestras salidas o predicciones?

El sistema se utilizará mediante un flujo de trabajo simple: cuando el usuario tome una foto del medicamento seleccionado, el sistema de su comercio llamará a nuestro modelo vía endpoint disponibilizado en AWS Sagemaker. El mismo verificará si el medicamento coincide con el solicitado (medicamento a validar). Si es correcto, responderá con un código de aprobación (pasa); si no es correcto, mostrará uno de desaprobación (no pasa), aclarando detalles de error para permitir su resolución.

# **Bloque 2: Fuentes de datos**

### 1. ¿Con qué datos históricos vamos a alimentar el modelo?

El modelo será alimentado por diversas fuentes de datos históricas con información relevante para la identificación de medicamentos y su gestión en farmacias:

#### **Fuentes de datos internas:**

* **Imágenes de medicamentos**: Fotografías históricas de los envases y etiquetas de los medicamentos disponibles en Argentina.
* **Etiquetas y bases de datos de medicamentos**: Bases de datos farmacéuticas que contengan información detallada sobre medicamentos aprobados, sus presentaciones y nombres comerciales

#### **Fuentes de datos externas:**

* **Sistemas de inventarios (OLTP)**: Datos transaccionales de las farmacias sobre el inventario y la rotación de medicamentos, que podrían incluir códigos de barras, nombres, fechas de vencimiento y lotes.
* **APIs REST de enriquecimiento:** APIs que proporcionan información farmacéutica, como el Vademecum o servicios de identificación de medicamentos que puedan enriquecer el sistema con datos actualizados sobre nuevas presentaciones o productos.

### 2. ¿Cuáles son los costos ocultos de la aplicación de ML?

* **Preparación de los datos**: Limpiar, curar y disponibilizar eficientemente los datos para el modelo puede requerir tiempo y recursos significativos, especialmente si los datos provienen de múltiples fuentes no estructuradas.
* **Etiquetado manual**: Si se requiere un etiquetado manual de imágenes de medicamentos, el costo en términos de horas de trabajo puede ser alto, particularmente para crear un dataset inicial para entrenar el modelo.
* **Mantenimiento del modelo**: El modelo necesitará mantenimiento continuo para ajustarse a cambios en las etiquetas, nuevos productos o formatos de presentación.
* **Infraestructura:** Cada reentrenamiento requiere mucho procesamiento por tratarse de una solución de deep learning. Si los tiempos comprometidos del reentrenamiento y la frecuencia son exigentes, la infraestructura a contratar será costosa.
* **Almacenamiento**: Si los datos crecen exponencialmente, los costos de almacenamiento también lo harán, particularmente por contener datasets de imágenes.

### 3. ¿Cuánto podrían crecer los datos?

Los datos pueden crecer significativamente con el tiempo, debido a:

* **Aumento de las imágenes de medicamentos:** A medida que la farmacia adquiera más productos o cambien las presentaciones, el número de imágenes a procesar y almacenar crecerá.
* **Actualizaciones constantes de catálogos de medicamentos:** Los laboratorios farmacéuticos suelen cambiar sus productos y etiquetas con frecuencia, lo que implica una necesidad continua de actualización de datos.
* **Datos de uso y rendimiento:** A medida que el sistema se implemente en múltiples farmacias, también se generarán datos de uso , rendimiento del sistema y tableros de control, lo que implicará un aumento progresivo en el almacenamiento.

### 4. ¿Qué tan costoso se puede volver el almacenamiento de los datos?

El almacenamiento de datos puede volverse costoso dependiendo de la cantidad de imágenes de alta resolución que se almacenen, el número de transacciones registradas. Utilizar servicios en la nube con estrategias de almacenamiento eficientes, como almacenar imágenes comprimidas y utilizar soluciones de almacenamiento frío para datos históricos, podría ayudar a mitigar estos costos.

# **Bloque 3: Tarea de predicción**

### 1. ¿Es supervisada o no supervisada?

La tarea de predicción en este proyecto es **supervisada**. El sistema necesitará ser entrenado con un conjunto de datos etiquetados, donde cada imagen de un medicamento esté asociada con su nombre y características correctas. Esto permitirá que el modelo aprenda a identificar correctamente cada medicamento en función de su etiqueta visual.

### 2. ¿Es detección de anomalías?

El problema no se enfoca en la **detección de anomalías** per se, aunque el sistema podría detectar situaciones fuera de lo esperado (como un medicamento con una etiqueta incorrecta o una etiqueta ilegible). El enfoque principal es más bien la **clasificación**, es decir, determinar si el medicamento en cuestión coincide o no con el solicitado.

### 3. ¿El problema es sobre qué opción debería ser tomada?

Sí, el problema se refiere a la decisión binaria de si el medicamento es correcto o no. El sistema debe tomar una decisión entre dos opciones: **"pasa" (el medicamento es el correcto)** o **"no pasa" (requiere revisión manual)**.

### 4. ¿Necesitamos predecir un valor continuo?

No, no es necesario predecir un valor continuo. El problema aquí es de **clasificación binaria**, donde se debe determinar si el medicamento es el correcto o no, en lugar de predecir un valor numérico continuo.

### 5. ¿Qué categoría debe ser predicha? ¿Es binaria o multiclase?

El problema es de **clasificación binaria**, donde las dos categorías a predecir son:

1. **Correcto (Pasa)**: La imagen del medicamento coincide con el solicitado.
2. **Incorrecto (No pasa)**: La imagen del medicamento no coincide con el solicitado.

### 6. ¿Necesitamos agrupar nuestra data?

No es necesario agrupar los datos en este caso, ya que se trata de un problema de clasificación directa.

### 7. ¿Cuál sería la definición de un sample?

Un **sample** sería cada imagen de un medicamento, que va acompañada de una etiqueta que identifica el nombre del medicamento correcto (y posiblemente otros atributos como el número de lote o la fecha de vencimiento). Este sample se usa para entrenar el modelo de deep learning en la tarea de reconocimiento y clasificación de imágenes.

### 8. ¿Será un modelo de Machine Learning o de Deep Learning?

Dado que el proyecto implica **identificación visual** de etiquetas en medicamentos, un **modelo de Deep Learning** es la opción más adecuada. Las redes neuronales convolucionales (CNN) son particularmente efectivas para la tarea de clasificación de imágenes, entrenándose en un conjunto de datos de imágenes etiquetadas de medicamentos.

# 

# **Bloque 4: Ingeniería de features**

### 1. ¿Cómo se extraen los features de las fuentes crudas?

Con **EfficientNetB1**, la extracción de características se realiza directamente a partir de las imágenes de las etiquetas de los medicamentos. Este modelo de deep learning está optimizado para identificar patrones visuales como formas, colores, disposición de los elementos en la etiqueta y el texto impreso. Además, se sigue utilizando **OCR** para extraer el texto de las etiquetas, como el nombre del medicamento, el número de lote o la fecha de vencimiento, lo que complementa el análisis de las imágenes.

### 2. ¿Hay que cruzar muchas fuentes de datos?

Aunque la fuente principal siguen siendo las imágenes, en algunos casos es necesario cruzarlas con otros datos tabulares como los registros de inventario o la base de datos de medicamentos solicitados. EfficientNetB1 se encargará de la identificación visual, mientras que el cruce con bases de datos permite verificar que el medicamento procesado es el correcto según la solicitud.

### 3. ¿Hay que hacer un procesamiento complejo para que sean útiles?

El procesamiento de las imágenes es menos complejo debido a la capacidad optimizada de **EfficientNetB1** para extraer características de forma eficiente. Sin embargo, sigue siendo necesario hacer algunos pasos de preprocesamiento como la normalización de imágenes (tamaño, brillo, contraste) y la eliminación de ruido. EfficientNetB1 también se beneficia de técnicas de **data augmentation**, como rotaciones o ajustes en la iluminación, lo que permite entrenar el modelo de forma robusta. El uso de **OCR** continúa siendo esencial para capturar el texto de las etiquetas.

### 4. ¿Hay personal o herramientas para hacer esta labor?

Si, el equipo de Data & AI está compuesto por ingenieros de Machine Learning y expertos en Computer Vision, que poseen amplia experiencia en el procesamiento de imágenes y entrenamiento de modelos de reconocimiento.

### 5. ¿Se considera incluir a expertos del dominio para especificar qué aspectos de la data son los más importantes para esta tarea de ML en particular?

La participación de **expertos en farmacia** es fundamental. Estos expertos señalan los aspectos más importantes para diferenciar medicamentos similares y colaboran en la validación del sistema. Gracias a su experiencia, se identifican las características visuales críticas y se anticipan posibles problemas como presentaciones similares o errores en las etiquetas. Esto asegura que el sistema diseñado con **EfficientNetB1** cumpla con las expectativas del sector y minimice riesgos en la identificación.

# **Bloque 5: Evaluación offline**

Antes de implementar cualquier tipo de modelo debemos especificar y establecer la metodología y métrica para evaluar el **sistema completo**

### 1. Métricas específicas del dominio que justifiquen el desarrollo del modelo

Para el sistema de identificación de etiquetas de medicamentos, las métricas más relevantes incluyen precisión (Accuracy), exactitud (Precision), sensibilidad (Recall) y F1-Score. Estas métricas se enfocan en reducir los errores en la entrega de medicamentos, asegurando que se cumplan los requerimientos de seguridad del paciente y eficiencia en el manejo del inventario.

### 2. Objetivo del negocio

Garantizar que, en un plazo de 6 meses, el sistema logre una **precisión mínima del 98%** y un **recall mínimo del 95%** en la identificación de medicamentos, asegurando así la entrega correcta de productos y la seguridad del paciente.

### 3. ¿Cuál será la performance mínima con la que se autorizará la puesta productiva del modelo?

Para autorizar la implementación del modelo en producción, se establecerán los siguientes umbrales de rendimiento: **precisión mínima del 98%**, **recall mínimo del 95%** y **F1-Score > 0.96**. Si el modelo no alcanza estas métricas en pruebas offline, no se procederá a la implementación productiva.

### 4. ¿Cuáles son las consecuencias medibles en los errores de predicción del modelo como los Falsos Positivos y los Falsos Negativos?

Los errores de predicción pueden tener consecuencias significativas. Los **Falsos Positivos** ocurren cuando el sistema identifica incorrectamente un medicamento como el solicitado, lo que puede llevar a la entrega de un producto incorrecto, afectando la salud del paciente y aumentando los costos operativos. Por otro lado, los **Falsos Negativos** se producen cuando el sistema no reconoce un medicamento que es el solicitado, retrasando el proceso de entrega y aumentando la carga de trabajo manual en el farmacéutico. Para minimizar estos errores, es que las métricas de performance del modelo son tan exigentes, garantizando la calidad del servicio o identificando claramente la excepción, evitando riesgos.

# **Bloque 6: Toma de decisiones**

### 1. ¿Qué decisiones se toman con las predicciones del sistema?

Las predicciones del modelo determinan si el medicamento que se va a entregar al farmacéutico es correcto o requiere revisión. Si el sistema indica que la identificación es correcta (señal verde), se procede a la entrega del medicamento. En caso de una identificación incorrecta (señal roja), se inicia un proceso de verificación adicional, donde el farmacéutico revisa el medicamento antes de la entrega al paciente.

### 2. ¿Estas decisiones proveen el valor propuesto al usuario final?

Sí, las decisiones tomadas basadas en las predicciones del sistema proporcionan un valor significativo al usuario final al minimizar errores en la entrega de medicamentos. Esto no solo mejora la eficiencia operativa de la farmacia, sino que también asegura la seguridad del paciente al garantizar que reciban el medicamento correcto, reduciendo así riesgos potenciales para la salud.

### 3. ¿Cómo interactúa el usuario final u otro sistema con las predicciones del modelo?

El usuario interactúa con el sistema a través de una interfaz fácil de usar. Esta interfaz permite capturar imágenes de los medicamentos y recibir retroalimentación inmediata sobre si el medicamento coincide con el solicitado. Además, el sistema puede integrarse con software de gestión de inventario, facilitando la actualización automática del stock y generando alertas en caso de discrepancias.

### 4. ¿Cuáles son los costos ocultos en la toma de decisiones como los human-in-the-loop?

Aunque el sistema automatiza gran parte del proceso, se deben considerar costos ocultos relacionados con la necesidad de un **human-in-the-loop**. Esto incluye el tiempo que el personal dedica a la verificación de medicamentos cuando se producen errores o medicamentos sin reconocimiento automático, así como la capacitación necesaria para usar el sistema de manera efectiva. También puede haber costos asociados a la revisión de casos especiales, el manejo de devoluciones de medicamentos incorrectos y la pérdida de confianza en el sistema si ocurren errores recurrentes. Estos costos pueden afectar la eficiencia general y el retorno sobre la inversión del proyecto.

# **Bloque 7: Realizando predicciones**

Este bloque nos permite saber cuando hacer las predicciones basadas en nuevas entradas.

### 1. ¿Cuándo deberían estar las predicciones disponibles?

Las predicciones deben estar disponibles en tiempo real, es decir, inmediatamente después de que se capture la imagen del medicamento. Esto asegura una respuesta rápida y eficiente en el proceso de verificación antes de la entrega al farmacéutico.

### 2. ¿Nuevas predicciones son hechas a demanda?

Las nuevas predicciones se generan a demanda. Cada vez que un farmacéutico o personal del inventario toma una foto de un medicamento, el sistema procesa la imagen y devuelve la predicción en tiempo real.

### 3. ¿Nuevas predicciones son calendarizadas?

No se requiere un enfoque calendarizado para las predicciones, ya que el sistema está diseñado para funcionar de manera reactiva, respondiendo a las solicitudes inmediatas en el flujo de trabajo diario de la farmacia.

### 4. ¿Las predicciones son hechas sobre la marcha en cada punto de datos o para un batch de datos de entrada?

Las predicciones se realizan sobre la marcha, en cada punto de datos. Esto permite una interacción fluida y a demanda, garantizando que cada medicamento se verifique de forma eficiente antes de ser entregado.

### 5. ¿Hay un human-in-the-loop en la salida de estas predicciones?

Aunque el sistema automatiza la identificación, un human-in-the-loop está presente en la etapa de verificación. Si la predicción indica un error, el farmacéutico revisa manualmente el medicamento para asegurar que no se cometa un error en la entrega.

### 6. ¿Se dispone de hardware para predecir?

El sistema requiere hardware adecuado, como cámaras de calidad para capturar imágenes nítidas de las etiquetas de los medicamentos. Además, se necesitará una estación de trabajo con capacidad de procesamiento suficiente para ejecutar el modelo **EfficientNetB1** y realizar predicciones en tiempo real.

### 7. ¿Se utiliza algún servicio de Cloud para predecir?

El modelo fue entrenado con herramientas, algoritmos y arquitecturas independientes de los servicio cloud. Al mismo tiempo, se utilizan las facilidades de los entornos cloud para simplificar el desarrollo y consumo del modelo desarrollado.

# **Bloque 8: Recolectando datos**

### 1. ¿Se dispone de datos para hacer un entrenamiento inicial?

Se dispone de un conjunto de datos inicial que incluye imágenes de etiquetas de medicamentos y sus correspondientes etuiquetas. Estos datos se utilizarán para entrenar el modelo EfficientNetB1 y establecer una base sólida para su funcionamiento.

### 2. ¿Cómo etiquetamos los datos nuevos?

Los nuevos datos se etiquetan mediante un proceso que involucra la verificación manual por parte del personal de la farmacia por única vez. Cada imagen capturada de un medicamento se compara con la base de datos existente para asegurar que la etiqueta sea correcta. Se puede implementar un sistema de anotación semiautomático, donde el modelo hace sugerencias de etiquetas que luego son confirmadas o corregidas por el farmacéutico. Cada nuevo medicamento agregado será aprovechable por todos los usuarios al mismo tiempo gracias a ser consumido como servicio y tener una fuente única de conocimiento.

### 3. ¿Hay que procesar datos multimedia de tipo imagen, sonido o video?

Si, la mayoría de los datos procesados son imágenes, específicamente imágenes de las etiquetas de medicamentos para luego poder identificar las nuevas capturas enviadas

### 4. ¿Hay un human-in-the-loop en la limpieza manual y el etiquetado de la data entrante?

Sí, se implementa un human-in-the-loop durante el proceso de limpieza y etiquetado de los nuevos datos. El personal de la farmacia participa activamente en la revisión de los datos entrantes para asegurar que estén correctamente etiquetados y limpios de errores, al mismo tiempo que el modelo registra todos los casos que no son identificados o son marcados como erroneos. Este enfoque no solo mejora la calidad de los datos, sino que también ayuda a mantener la precisión del modelo a lo largo del tiempo.

# **Bloque 9: Construyendo modelos**

Referido a la **actualización** del modelo, dado que diferentes tareas requieren diferentes frecuencias de reentrenamiento:

### 1. ¿Qué tan seguido debería ser reentrenado el modelo? A cada hora, cada semana, o con cada punto nuevo de datos que ingresa. Por ejemplo, por inflación.

El modelo debería ser reentrenado bajo demanda o cada vez que se necesite la incorporación de un nuevo medicamento a la base de reconocidos. Este enfoque permite mantener la precisión del modelo frente a posibles cambios en la presentación de etiquetas o nuevos medicamentos que ingresen al mercado.

### 2. ¿Se dispone de hardware para entrenar?

Sí, se dispone de hardware adecuado, incluyendo GPU y amplia memoria RAM y en caso de necesidad de escalar, se cuenta con recursos elásticos en la nube de AWS. Al no tener tiempos acotados de entrenamiento, no se ve necesario más hardware que el disponible.

### 3. ¿Se utiliza algún servicio de Cloud para entrenar?

No, se utilizan servicios cloud para disponibilizar de forma eficiente y segura principalmente.

### 4. ¿Cómo vamos a lidiar con los asuntos de escala en la medida que la operación se vuelva más compleja y costosa?

Según la estrategia planteada y lo mencionado anteriormente, el procesamiento no debería escalar de forma significativa en el mediano plazo y en caso de hacerlo de forma excepcional, se puede ejecutar el entrenamiento en un entorno cloud con dimensionamiento elástico prácticamente sin límite. Lo que podría escalar con más probabilidad es el consumo del servicio, por lo que se debería tener alta disponibilidad de ancho de banda y recurrencia, otro asunto fácilmente escalable en entornos cloud.

### 5. ¿Cuál es el stack de tecnologías usado?

Python, GitHub, Amazon Sagemaker, Docker, Amazon Lambda, Amazon S3, EfficentNETB1 (base de modelo)

**6. ¿Qué hacemos si aparece un modelo que supera ampliamente el que estamos desarrollando?**

Si aparece un modelo que supera significativamente al desarrollado, se evaluará la viabilidad de lograr resultados similares con el nuestro para luego aplicarlo. La estrategia de implementar en dockers en un entorno cloud, permite que el reemplazo de una versión de modelo por otra mejorada sea parte de lo planificado.

# **Bloque 10: Monitoreo y evaluación en vivo**

### 1. ¿Cómo vamos a hacer seguimiento de la performance del sistema?

Se implementará un sistema de monitoreo continuo que evalúe la precisión y el rendimiento del modelo en tiempo real. Esto incluirá la recopilación de métricas como la tasa de falsos positivos y negativos, así como todo el feedback proporcionado por el usuario final. Además, se realizarán auditorías periódicas para revisar el rendimiento general del sistema.

### 2. ¿Cómo evaluamos la creación de valor?

La creación de valor se evaluará mediante indicadores clave de rendimiento (KPI) que reflejen la efectividad del sistema. Por ejemplo:

* Cantidad de errores en la entrega de medicamentos
* Cantidad de medicamentos estropeados
* Tiempo promedio de atención
* Satisfacción de los clientes finales

Los mismos permiten evaluar la generación de valor, generación de eficiencia o mejora en la experiencia de los usuarios

# **Arquitectura Implementada**

