

## TALLER 2

**Objetivo:** evaluar las capacidades del estudiante para aplicar técnicas de machine learning, las cuales permita descubrir insights, sugerir accionables al negocio y calcular el valor ganado.

**Contexto del negocio:** Apoyo a un Supermercado Inteligente

El concepto de supermercado inteligente está transformando la manera en que interactuamos con los entornos de retail, combinando la inteligencia artificial, la automatización y la ciencia de datos para ofrecer una experiencia de compra más eficiente, personalizada y sin fricciones. Estos supermercados están diseñados para mejorar todos los aspectos del proceso de compra, desde la entrada del cliente hasta la salida, utilizando tecnologías avanzadas para automatizar tareas, gestionar inventarios en tiempo real y analizar el comportamiento del consumidor.

Usted ha sido contratado como científico de datos para un nuevo supermercado inteligente, que ya ha implementado tecnologías avanzadas como sensores, cámaras y para iniciar la automatización de procesos, ahora necesita aprovechar mejor los datos generados a diario para tomar decisiones más informadas y eficaces.

El supermercado le ha dado la responsabilidad de desarrollar integrado un sistema de automatización que permite a los clientes tomar productos y salir sin pasar por caja, mientras las cámaras y sensores registran automáticamente los artículos seleccionados.

Para esto le ha proporcionado un dataset abierto de varios productos de mercado que además tiene etiquetas suficientes para identificar el tipo de productos:

- [Dataset](#)

Mediante el uso de modelos de Machine Learning, en conjunto con técnicas de preparación de datos, se espera que usted esté en capacidad de construir el modelo que identifique los productos, y argumente el valor que generará al supermercado los resultados que obtenga.

## ACTIVIDADES Y ENTREGABLES

1. **[20%] Entendimiento y preparación de los datos:** Genere un breve reporte de entendimiento de datos en donde destaque las dimensiones del dataset, las características de las imágenes e indicadores que considere importante. Integre técnicas de aumento de datos y determine qué productos y que categorías empleará.
2. **[20%] Entrenamiento del modelo de Machine Learning:** Utilizando el conjunto de imágenes, construya un modelo que sea capaz de predecir el tipo de producto. Pruebe al menos dos modelos diferentes. Recuerde utilizar de la siguiente manera cada conjunto:
  - a. El conjunto de entrenamiento (*train*) para ajustar los hiperparámetros de los modelos.
  - b. El conjunto de prueba (*test*) para seleccionar el mejor modelo.
  - c. El conjunto de validación (*valid*) para realizar el análisis de resultados del modelo.

Aproxime su solución en el segundo nivel de categorías (Tipos de productos, i.e manzanas, aguacates, bananas, etc).

3. **[20%] Análisis de resultados del modelo:** Calcule las métricas de evaluación para su modelo, y explique su significado y su valor para la empresa. Justifique la calidad de su modelo, qué elementos impactaron positivamente dichas métricas, y qué oportunidades de mejora encuentra.
4. **[20%] Generación de valor:** y determine la ganancia esperada para el Supermercado con la implementación del modelo por cada predicción acertada, así como el punto a partir del cual la inversión en el modelo empezará a dejar dividendos. Realice una estimación paramétrica, haciendo explícitos todos los supuestos usando datos, incluyendo fuentes aplicables y estadísticas recientes. Para este punto debe considerar:
  - a. Los costos de tiempo asociado al registro de productos.
  - b. El ahorro de tiempo teórico de su modelo.
  - c. El costo asociado a los errores de su modelo.
  - d. El ahorro real por predicción acertada.
  - e. El ROI tomando en cuenta los costos de desarrollo y despliegue de su modelo.
5. **[20%] Insights:** Redacte un informe ejecutivo o una presentación corta en donde muestre los hallazgos o insights más importantes en términos del modelo y los resultados. Ofrezca una recomendación final para el negocio.
6. **[10%] Bono:** Realice su modelo usando el nivel detallado de tipos de productos y sus marcas. (por ej. Manzanas Golden-Delicious, Granny-Smith, etc).

## **FORMATO DE ENTREGA**

- El taller debe desarrollarse en parejas.
- Todos los recursos generados deben entregarse mediante un repositorio de GitHub con su estructura apropiadamente documentada mediante el archivo Readme.
- El repositorio de GitHub debe ser público.
- El o los notebooks generados deben poderse ejecutar secuencialmente sin ningún error. En el caso de haber creado varios notebooks, se debe clarificar el orden en que deben ser ejecutados. No olvide también incluir el dataset utilizado.
- Se penalizará a los estudiantes que no cumplan con este formato de entrega