

Taller 2 – Supermercado Inteligente.

Lizeth Viviana Perdomo Castañeda - lv.perdomoc1@uniandes.edu.co

David Esteban Fajardo Torres - de.fajardo@uniandes.edu.co

En el siguiente informe, se presenta el proceso de desarrollo de un modelo de clasificación de Machine Learning el cual consiste en un sistema de automatización que permite a clientes de un supermercado tomar productos y salir sin pasar por caja, mientras las cámaras y sensores registran automáticamente los artículos seleccionados.

1. Entendimiento y preparación de los datos

El dataset se basa en un conjunto de imágenes las cuales se encuentran clasificadas en *train*, *test* y *val*, dentro de esta clasificación se encuentran las etiquetas principales que corresponden a *Fruit*, *Packages* y *Vegetables*; para el segundo nivel de etiquetas se encuentra el tipo de producto, por ejemplo, *apple*, *asparagus* y *milk*. Finalmente, algunas de estas categorías cuentan con un tercer nivel de etiquetado que corresponde a la variedad del producto.

Al realizar el análisis de las categorías en el tercer nivel de etiquetado, se encuentra que en el conjunto *train* y *test* se tienen 81 clases y el conjunto *val* tiene 60 clases, lo que corresponde al 74% de la totalidad, por lo tanto, para el desarrollo del modelo se utilizan las 60 clases en los tres conjuntos de datos, obteniendo 1996 imágenes en *train*, 1889 en *test* y 296 en *val*.

En cuanto a las características de las imágenes, de las 60 clases, 22 (36%) corresponden a frutas, 20 (34%) corresponden a paquetes y 18 (30%) corresponden a vegetales, por lo que se evidencia una distribución uniforme en el tipo de productos utilizados para el desarrollo del modelo.

Todas las imágenes son a color, se encuentran en formato jpg y tienen un tamaño entre 16 y 30 KB.

Para el aumento de datos, se aplican rotaciones, desplazamientos y zoom de manera aleatoria al conjunto de datos *train* generando versiones variadas de cada imagen en cada época de entrenamiento, mejorando así la capacidad del modelo para generalizar.

2. Entrenamiento del modelo de Machine Learning

Para el desarrollo de la solución se prueban dos modelos de Machine Learning, Convolutional Neural Network (CNN) y MobileNetV2 como modelo basado en Transfer Learning utilizando imageNet de dataset.

Para el entrenamiento del modelo se utiliza el conjunto de imágenes de *train*, para seleccionar el modelo con mejores resultados se utiliza el conjunto de imágenes de *test* y, finalmente, para el análisis de los resultados del modelo seleccionado se utiliza el conjunto de imágenes de *val*.

3. Análisis de resultados del modelo

Para la evaluación de los modelos, se utilizaron las métricas de accuracy, loss, precisión, recall y f1_score, obteniendo los siguientes resultados:

Modelo CNN

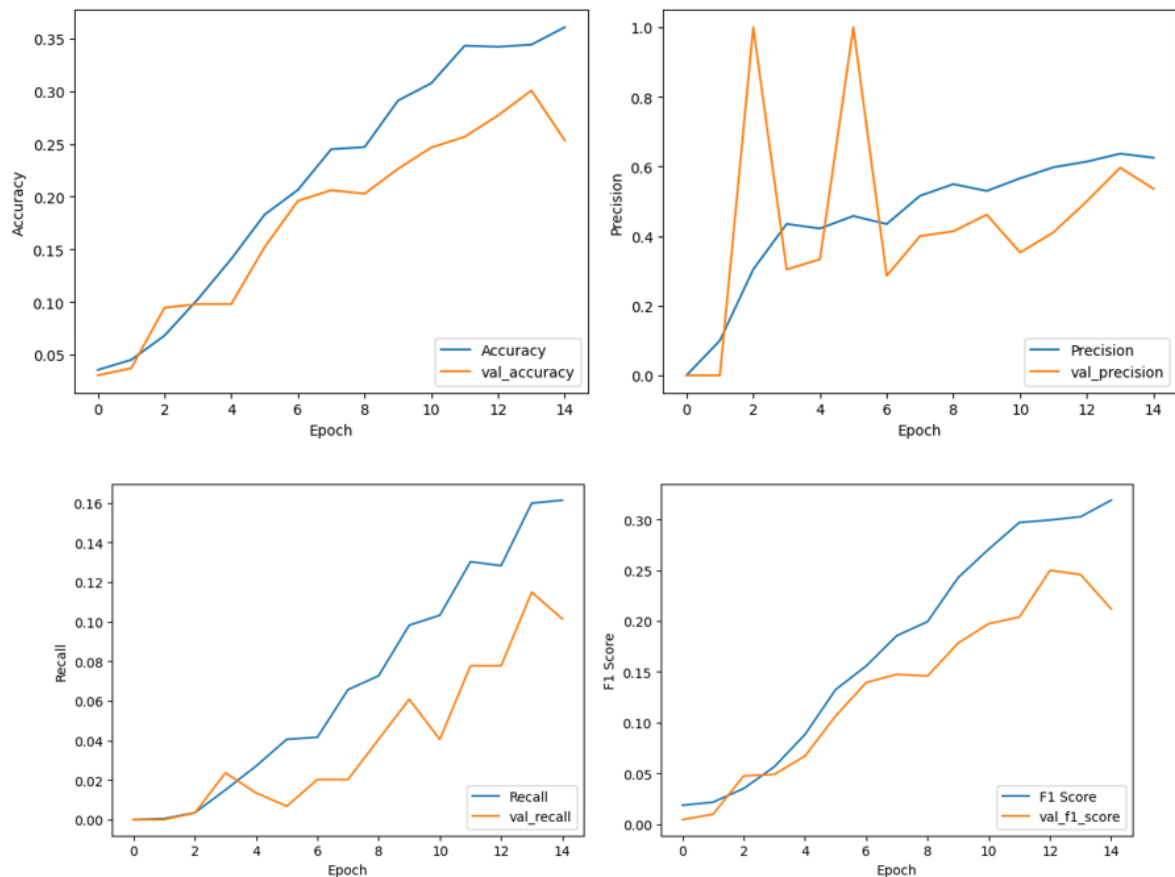


Ilustración 1. Métricas modelo CNN

- Accuracy (0.3486): El modelo clasifica correctamente el 34.86% de las muestras.
- Loss (2.0464): La pérdida representa el error del modelo en términos de sus predicciones frente a las etiquetas reales.
- Precision (0.5810): El 58.10% en realidad pertenecen a la clase predicha. Aunque el modelo logra cierta precisión en las predicciones positivas, este valor no está acompañado de un *recall* alto, lo cual sugiere un desbalance.
- Recall (0.1669): El *recall* mide la capacidad del modelo de detectar correctamente las muestras de cada clase. En este caso, un 16.69% indica que el modelo solo identifica correctamente una pequeña fracción de las instancias reales de cada clase, lo que generalmente significa que está pasando por alto la mayoría de los ejemplos positivos.
- F1 Score (0.2823): El F1 Score es una medida de la precisión y el *recall*. Por lo tanto, el modelo tiene problemas tanto para clasificar correctamente (*recall*) como para ser preciso en las predicciones.

Modelo de Transfer Learning

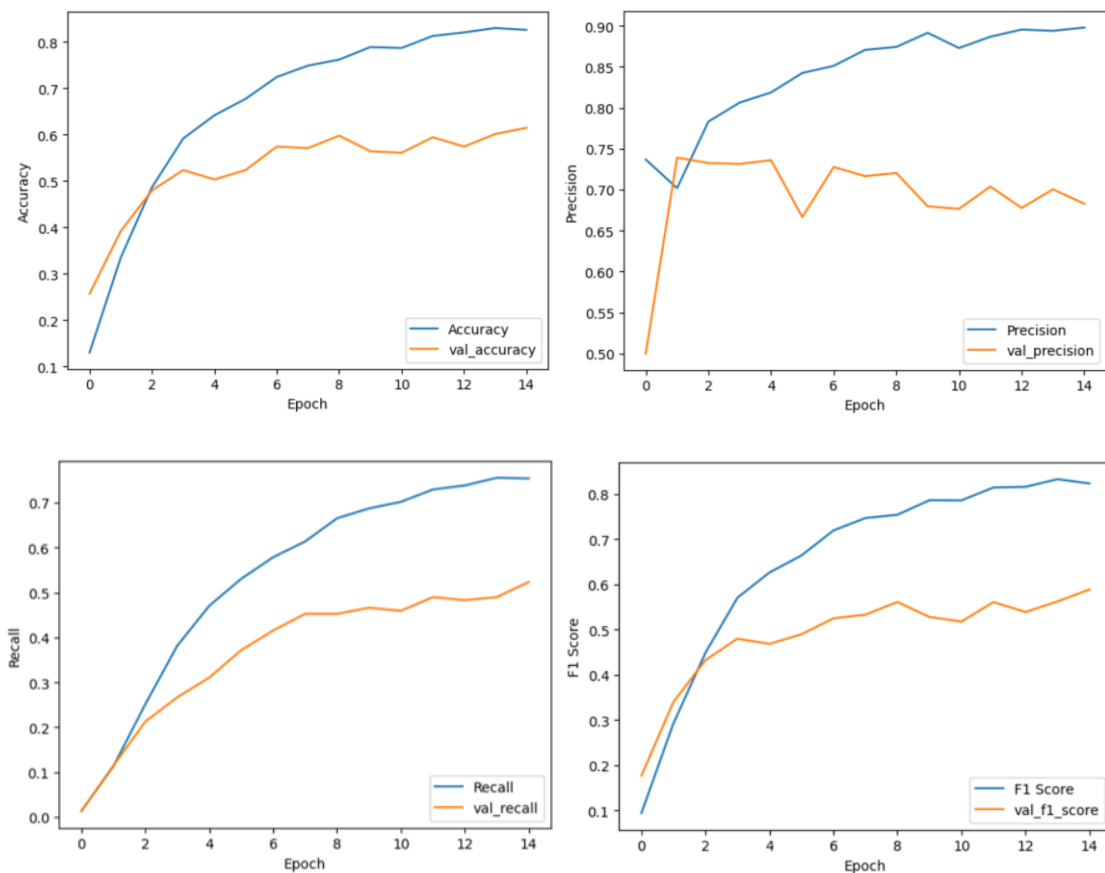


Ilustración 2. Métricas modelo Transfer Learning

- Accuracy (0.6638): Con una precisión global del 66.38%, el modelo está clasificando correctamente aproximadamente dos tercios de las muestras, lo cual es una mejora en comparación con un modelo aleatorio.
- Loss (0.9987): El modelo aún tiene una discrepancia entre las predicciones y las etiquetas verdaderas, pero es significativamente menor en comparación con el modelo CNN.
- Precision (0.7568): El modelo es bastante bueno en predecir la clase correcta cuando hace una predicción positiva.
- Recall (0.5850): El modelo está pasando por alto bastantes ejemplos de las clases correctas.
- F1 Score (0.6183): El modelo muestra un balance entre la precisión y el recall, es decir, que en general las predicciones que hace tienen a ser correctas y muestra una mejora significativa con el modelo CNN.

De acuerdo con los resultados obtenidos, se evidencia un mejor desempeño en el modelo de Transfer Learning, por esto, se hace un análisis más detallado a este modelo y se obtienen los siguientes resultados:



Ilustración 3. Comparación EPOCHS - Porcentaje de clasificación de cada clase

En la ilustración 3 se evidencia el porcentaje de clasificación de cada categoría, en la imagen de la izquierda se realiza el entramiento del modelo con 10 épocas y en la de la derecha con 15 épocas; se logra evidenciar que con 15 épocas el modelo mejoró, ya que más categorías pudieron ser predichas y aumentó el número de categorías con un 100% de predicciones. El detalle de esta visualización se encuentra en el repositorio.

4. Generación de valor

Con la implementación del modelo de Transfer Learning se estima la siguiente generación de valor al supermercado:

- a. El tiempo promedio para registrar un producto en un supermercado varía, pero generalmente toma entre 3 y 5 segundos por producto cuando es manual y de 1 a 2 segundos en sistemas automatizados como el autoservicio con escaneo. Estos datos provienen de estudios de eficiencia operativa en retail realizados por investigadores de la National Retail Federation (NRF).

Para efectos de cálculo, se asume un tiempo de 3 segundos por producto. En sistemas tradicionales, este tiempo se traduce en costos operativos directos e indirectos. Con el modelo, el proceso se automatiza, y los productos se registran automáticamente, eliminando la intervención humana y reduciendo costos asociados a personal.

- b. El modelo de Transfer Learning optimiza el proceso de registro al reconocer el producto en menos de un segundo por imagen. Asumimos un tiempo de procesamiento de 0.5 segundos por producto.

Teniendo en cuenta la precisión del modelo, se procesará de manera correcta aproximadamente el 66.38% de los productos de un cliente en el tiempo estimado de 0.5 segundos por producto. Para los productos que el modelo clasifique erróneamente, se requerirá intervención adicional, lo cual añadirá tiempo de procesamiento.

Considerando una compra promedio de 15 productos:

- Productos correctamente clasificados: $15 \text{ productos} * 66.38\% * 0.5 \text{ seg} \approx 4.97 \text{ segundos}$.

- Productos incorrectamente clasificados: $15 \text{ productos} \times 33.62\% \times 3 \text{ seg} \approx 15.13 \text{ segundos}$ (suponiendo 3 segundos de corrección manual por producto).

Tiempo total por cliente:

- Implementando el modelo: $4.97 \text{ seg} + 15.13 \text{ seg} = 20.1 \text{ segundos}$
- Proceso manual: $15 \text{ productos} \times 3 \text{ seg} = 45 \text{ segundos}$

Esto representa una reducción de aproximadamente 55% en el tiempo de procesamiento por cliente en comparación con el método manual. Este ahorro en tiempo implica que el flujo de clientes aumenta y se reducen los cuellos de botella, mejorando la eficiencia operativa del supermercado.

- c. Con una precisión del 66.38%, el modelo tiene una tasa de error del 33,62% en la clasificación de productos.

Se asume un costo promedio de \$2 USD por error, debido a ajustes en inventario o devoluciones. Este valor se basa en datos sobre pérdidas por inexactitudes de inventario en el sector minorista, especialmente en supermercados.

Para cada 1,000 productos procesados, se clasificarán incorrectamente el 33,62%, lo que resulta en: $1000 \times 0.3362 \times 2 = 672.4 \text{ USD}$

Este es un costo considerable y sugiere que podrían requerirse mejoras en el modelo para reducir el impacto financiero y mejorar la precisión.

- d. El ahorro real por predicción acertada se basa en la reducción de personal necesario para registrar productos y en el aumento de la satisfacción del cliente al reducir el tiempo de espera.

Se asume un salario de \$15 USD por hora para el personal en caja, considerando que es el valor promedio del salario mínimo por hora en los Estados Unidos. Si se reduce el tiempo total de registro en 55% y cada empleado puede atender a más clientes en el mismo tiempo, se podría operar con menor personal o reasignarlo.

Con un flujo de 100 clientes por hora y un ahorro de 24.9 segundos por cliente, el ahorro por hora sería:

$$100 \times (24.9 \text{ seg}) = 2490 \text{ segundos} = 41.5 \text{ minutos}$$

Este ahorro se traduce en un ahorro de personal de aproximadamente \$10 USD por hora.

- e. Para calcular el retorno de inversión (ROI), se consideran tanto los costos de desarrollo como los ahorros anuales del modelo, asumiendo que los costos de instalación e infraestructura ya fueron considerados en otra etapa del proyecto.

Costos de Desarrollo:

- Entrenamiento del Modelo y Configuración de Hardware: \$5,000 USD.
- Mantenimiento Anual y Ajustes: \$1,000 USD.

Estos valores se asumen basados en revisión bibliográfica en donde se establece que los costos de mantenimiento, que incluyen ajustes y optimizaciones periódicas del modelo para mejorar su precisión o adaptarse a cambios en el catálogo de productos, generalmente equivalen a entre el 15% y el 25% de los costos de desarrollo iniciales.

Por otro lado, la creación de una solución de IA personalizada para un proyecto básico de clasificación de imágenes suele ubicarse en un promedio de \$5,000 USD. Este costo incluye la preparación de datos, el desarrollo y entrenamiento del modelo, y la validación para asegurar precisión y rendimiento adecuado, ideal para un prototipo o proyecto piloto con alcance reducido.

Beneficio anual:

Con un ahorro de \$10 USD por hora, estimamos un ahorro anual de:

$$8 \text{ horas/día} \times 365 \text{ días/año} \times 10 \text{ USD/hora} = 29,200 \text{ USD/año}$$

Por lo tanto, el ROI es calculado como la diferencia entre el beneficio anual y los costos totales sobre los costos totales:

$$\text{ROI} = \frac{29200 - (5000 + 1000)}{5000 + 1000} \times 100 = 386.67\%$$

Con una precisión de 66.38%, el modelo aún ofrece un ROI positivo del 386.67%.

Teniendo en cuenta los cálculos realizados, el punto de equilibrio es el momento en el cual el beneficio acumulado iguala la inversión inicial y de mantenimiento. Para determinar este punto, sumamos los costos iniciales y los costos de mantenimiento y los comparamos con el beneficio acumulado anual.

$$\text{Punto de equilibrio} = \frac{5000+1000}{29200} = 0.205 \text{ años}$$

Esto indica que el modelo comenzaría a generar dividendos después de aproximadamente 2.46 meses de implementación. A partir de este momento, el beneficio neto empezaría a superar los costos acumulados, y el modelo comenzaría a ser rentable para el supermercado.

Insights

Con base en el análisis realizado y los resultados obtenidos, se destaca lo siguiente:

1. El modelo de Transfer Learning superó significativamente al modelo CNN en todas las métricas clave: accuracy, precision, recall, y F1 Score.
2. La precisión del 64.53% del modelo de Transfer Learning muestra que aún hay un margen significativo de optimización, especialmente en clases con menor frecuencia.
3. El modelo automatiza el registro de productos, reduciendo el tiempo de procesamiento en un 53.78% en comparación con el método manual, mejorando el flujo de clientes y reduciendo los cuellos de botella.
4. Los productos en las categorías de vegetales tuvieron una mejor clasificación en comparación con las categorías de frutas y paquetes.
5. Productos como zanahorias, lechuga, banano, aguacate, pepino, berenjenas, tomates y la leche de la marca Gran Ecological obtuvieron las mejores predicciones, alcanzando un 100% en algunas de ellas.

6. Fue complejo diferenciar productos como papa, jengibre y champiñones, debido a su similitud en color y forma, lo que resultó en un bajo desempeño en estas categorías.
7. A pesar de los costos asociados a los errores de clasificación, el modelo sigue siendo rentable, generando un ROI positivo del 386.67% y alcanzando el punto de equilibrio a los 2.46 meses de implementación.

Recomendaciones

Se recomienda optimizar el modelo de Transfer Learning entrenándolo con imágenes individuales y ampliando el dataset para mejorar la precisión en categorías visualmente similares, como los tubérculos. Configurarlos con 15 épocas asegura un mejor desempeño en comparación con 10 épocas, y se sugiere aplicar ajuste fino en las capas superiores de MobileNetV2 para incrementar la precisión. Además, implementar un sistema de revisión manual para productos con baja confianza en la predicción ayudará a reducir los costos asociados a errores en inventario.

Antes de una implementación total, se aconseja realizar un piloto en un entorno controlado, asegurando que el modelo opere adecuadamente bajo las condiciones reales del supermercado. Monitorear continuamente el desempeño y recalculando el ROI permitirá ajustar costos de mantenimiento y evaluar la sostenibilidad del sistema, maximizando así su impacto operativo y financiero.

Adicionalmente, una vez implementado el modelo, se recomienda entrenarlo regularmente ya que en los supermercados hay una rotación considerable de productos, así como cambios en la presentación.