

CIENCIA DE DATOS APLICADA



Grocery Store

INTEGRANTES

Nombre	Código	Correo Electrónico
Juan Nicolás Estepa Guzmán	202324933	j.estepa@uniandes.edu.co
Santiago Rodríguez Cruz	202013371	s.rodriguez52@uniandes.edu.co

INGENIERÍA DE SISTEMAS Y COMPUTACIÓN
MAESTRÍA EN INGENIERÍA DE LA INFORMACIÓN
UNIVERSIDAD DE LOS ANDES
BOGOTÁ D. C.

Table of Contents

1. Análisis Inicial de la Información.....	3
1.1. Situación actual.....	¡Error! Marcador no definido.
2. Analítica Realizada.....	¡Error! Marcador no definido.
2.1. Cantidad de usuarios y visitas	¡Error! Marcador no definido.
3. Resultados Obtenidos	¡Error! Marcador no definido.
4. Conclusiones.....	3

1. Análisis Inicial de la Información

A partir de la información brindada del almacén de cadena, se realiza un preprocesamiento a todas las imágenes de los tres datasets para que no se generen conflictos en el modelo. También se va a realizar una aumentación de datos modificando su rotación, aplicando zoom aleatorios y normalización de los datos. Se obtiene la distribución de los datos del set de entrenamiento para conocer información de las categorías a utilizar.

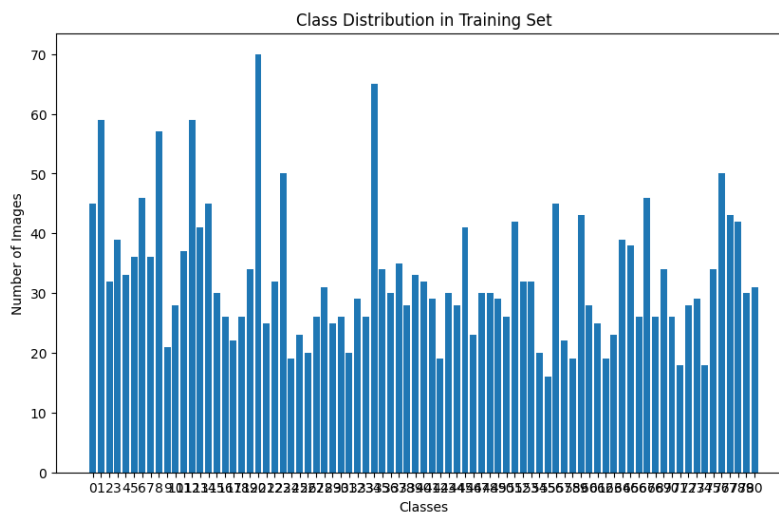


Figura 1. Distribución de categorías de alimentos, dataset de entrenamiento.

Con la información de la distribución, se observa que existen algunas clases con pocos datos, por lo que al momento de entrenar el modelo pueden existir fallos detectando estas clases. Cabe aclarar que se esta realizando la ejecución del modelo con todas las clases disponibles de los elementos del dataset, no en el segundo nivel de jerarquía disponible.

2. Modelo

Se realiza un modelo de redes neuronales convolucionales, donde se obtuvieron los siguientes resultados:

Conjunto de Entrenamiento:

- Accuracy: 0.41
- Loss: 2.50

Conjunto de Prueba:

- Accuracy: 0.408
- Loss: 2.49

Estos valores muestran una consistencia entre los conjuntos de entrenamiento y prueba, lo que indica que el modelo no está sobreajustado ni subajustado de forma significativa. Sin embargo, la precisión general sugiere que el modelo tiene dificultades para diferenciar correctamente las clases, y el alto valor de **loss** apunta a que las predicciones están lejos de las etiquetas reales en muchos casos.

Para obtener este resultado ayudo significativamente el preprocesamiento del dataset, la normalización del dataset y la modificación del formato "básico" del modelo de red neuronal convolucional, donde se utilizar 4 capas internas con una normalización a sus resultados de salida y el uso de una capa ****global average pooling**** para acelerar y estabilizar el proceso de aprendizaje. También en la capa de salida se aplicó una regularización ****L2**** para evitar el sobreajuste y problemas que se puedan presentar en zonas muertas.

Como oportunidades de mejora, se pueden agregar más datos al dataset de entrenamiento, aplicando más modificaciones a las imagenes originales para que la cantidad de datos sea mayor, y en clases que tengan poca información su entrenamiento pueda mejorar. Como segunda oportunidad se puede realizar la búsqueda de hiperpárametros que optimicen el entrenamiento y mejoren el resultado de este. Finalmente se puede realizar la implementación de modelos preentrenados para esta clase de problemas, que pueda mejorar los resultados al generar mejores características.

3. Generación de valor

Para el cálculo de la generación de valor vamos a tener en cuenta varios supuestos:

1. Costo total del modelo: 10000 €
2. Operación mensual y mantenimiento del modelo: 2000 €/mes
3. Beneficios mensuales: 7000 €

Con estos datos vamos a calcular lo siguiente

1. Beneficio actual

$$\text{Beneficio actual} = (\text{porcentaje del modelo} * \text{beneficios mensuales})$$

$$\text{Beneficio actual} = 2856 \text{ €/mes}$$

2. ROI

$$\text{ROI} = (\text{Beneficio actual} - \text{costos mensuales}) / (\text{costos totales de inversión}) * 100$$

$$\text{ROI} = 7.13\%$$

Al tener un ROI bajo, se espera que se puedan realizar actualizaciones al modelo que puedan generar un mayor valor a la empresa al momento de implementarlo, esto se consigue al momento en que se logre entregar un modelo con una precisión más alta, al menos de un 0.7%, así como también reducir los costos de desarrollo del modelo y de mantenimiento de este.

4. Insights

- Al tener un ROI positivo pero pequeño, se está empezando a generar valor y recuperar los costos de inversión realizados, pero es necesario que se mejore más el modelo, para que el retorno sea más significativo a corto plazo.
- Aunque los errores son significativos, el modelo ya contribuye a reducir costos relacionados con la clasificación manual, optimización del inventario, o mejoras en la calidad de las recomendaciones.
- El retorno actual sugiere que el modelo está en una etapa temprana. Si se mejora la precisión o se ajustan las operaciones, el ROI podría multiplicarse.
- Implementarse el modelo en este momento puede presentar problemas si se llegan a generar una subida de precios en los productos, lo que disminuya el valor actual del retorno de inversión, pudiendo llegar a ser hasta un valor negativo.

Como recomendaciones finales, se sugiere al cliente que en este momento haga uso del modelo para la segmentación de los alimentos dentro del negocio, dado que está generando un beneficio, teniendo en cuenta que si se reducen sus costos operacionales se puede lograr un mayor beneficio a corto plazo. Su implementación al principio puede costar, dado que existe la posibilidad de que los empleados no estén familiarizados con el sistema y les cueste adaptarse a él.