



## **PROYECTO DE GRADO**

Modelo de detección de productos en catálogos de venta directa

## **AUTOR**

Felipe Clement

## **ASESORES**

Javier Diaz Cely, PhD

Facultad De Ingeniería

Maestría en Ciencia de Datos

Santiago de Cali

2020

# Tabla de Contenidos

<b>Resumen</b> .....	3
<b>1. Introducción</b> .....	4
<b>1.1. Contexto y Antecedentes</b> .....	4
<b>1.2. Justificación</b> .....	5
<b>1.3. Descripción del Problema</b> .....	6
<b>1.4. Objetivo General</b> .....	6
<b>1.5. Objetivos Específicos</b> .....	6
<b>2. Antecedentes</b> .....	7
<b>2.1. Marco Teórico</b> .....	7
2.1.1. <i>Aprendizaje profundo y redes neuronales convolucionales</i> .....	7
2.1.2. <i>Familia de modelos R-CNN</i> .....	9
2.1.4. ResNet .....	11
2.1.5. Aumento de datos (data augmentation) .....	12
<b>2.2. Estado del Arte</b> .....	12
<b>3. Metodología</b> .....	14
<b>4. Preparación del conjunto de datos</b> .....	15
<b>5. Modelamiento mediante Faster R-CNN con detectron2</b> .....	18
<b>6. Validación</b> .....	19
<b>7. Resultados</b> .....	21
<b>8. Conclusiones</b> .....	23
<b>Bibliografía</b> .....	25
<b>Anexos</b> .....	27

## Resumen

Para la empresa Valienta S.A.S, es de gran importancia proveer una buena experiencia de usuario y un conjunto de características a sus vendedoras, con el propósito de lograr que tengan éxito en sus labores. Para resolver el problema específico de clasificar productos en catálogos, y hacerlos mas visibles a las usuarias, se decide etiquetar los productos dentro de 13 categorías. Para tal fin, se propone el uso del método de detección conocido como Faster R-CNN como método de detección automático de los productos, entrenado sobre un conjunto de datos con 1250 imágenes y alrededor de 4400 instancias, distribuidas de manera desbalanceada en las 13 clases. La metodología utilizada consistió en el preprocesamiento de catálogos en PDF a imágenes JPEG, las cuales fueron posteriormente etiquetadas con cajas alrededor de cada producto en el formato propuesto por MSCOCO. Se modela, mediante detectron2, distintas configuraciones de hiperparametros sobre el modelo, utilizando como validación un conjunto de datos que contiene el 20% de las instancias. Los resultados obtenidos por medio de la métrica de precisión promedio tienen una varianza alta entre categorías, y naturalmente demostrando métricas de validación inferiores en las clases con menor número de instancias. Para aquellos productos con mayor cantidad de instancias (ropa, ropa interior, y accesorios), la precisión de predicción fue considerablemente mas alta que para los objetos con menor número de instancias, lo cual para estas clases visualmente lleva a predicciones utilizables en contextos prácticos. Sin embargo, se encuentra una oportunidad para mejoras, particularmente en la cantidad de datos y modalidad de etiquetado, que lleven a la aplicación práctica de la detección de objetos sobre todas las clases en el conjunto de datos.

# 1. Introducción

## 1.1. Contexto y Antecedentes

En Colombia, la venta directa – o venta por catálogo – es una manera de distribución minorista a distancia que se lleva cabo por medio de intermediarios (vendedores) que se suscriben para ser distribuidoras de una empresa, quién les otorga un catálogo físico (o virtual) utilizado como soporte de ventas a clientes finales, a cambio de un porcentaje de comisión. Solamente en Colombia, mas de dos millones de vendedores componen esta forma de distribución, vendiendo más de 4.8 billones de pesos al año en cosméticos y elementos de cuidado personal, ropa, accesorios, y bienestar/alimentos, que componen el 45%, 30% y 7% de todas las ventas respectivamente (Mouthón, 2019). De esta fuerza de venta, el 87% son mujeres (para propósitos prácticos de este documento usamos el termino vendedora para referirnos a esta fuerza de venta de manera general), y aproximadamente dos tercios de estas personas tienen edad entre 25 y 44 años.

La relación entre vendedoras por catálogo y empresas es de naturaleza simbiótica, ambos interesados encuentran valor en tenerla. Para las empresas, la venta directa representa un canal de distribución económico y de gran potencial, debido a que las labores logísticas realizadas por las empresas se concentran solamente en el contacto con las vendedoras. Estas últimas se encargan de etapas posteriores del proceso, como la distribución y el mercadeo hacia sus redes de clientes finales. Por esta razón, las empresas obtienen un gran beneficio al utilizar este canal, permitiéndoles ampliar sus fuerzas de venta a un costo menor que el de operar locales, mantener una plataforma web e-commerce o vender por medio de mayoristas (los cuales pueden cobrar comisiones mas allá del 55% del total de la venta, un número mucho mayor a los porcentajes que se suelen otorgar a las vendedoras, que se concentran entre 20-50%), esta información es tomada empíricamente por Valienta mediante conversaciones con las empresas proveedoras. Por otro lado, la venta directa representa una manera de trabajo independiente accesible por muchas personas potencialmente sin un empleo formal, pues no requiere de un nivel educativo mínimo o experiencia de algún tipo en otras labores.

A pesar de las ventajas que trae la venta directa para las empresas y vendedoras, ambas partes sufren de problemas prácticos al momento de su implementación. Si bien la venta por catálogo brinda a las empresas ventajas sobre otros métodos de distribución, igualmente implementar este canal involucra costos iniciales altos y algunos futuros costos fijos. Principalmente, se deben realizar esfuerzos de mercadeo para atraer y mantener una fuerza de venta de vendedoras en el tiempo, diseñar y distribuir catálogos, y mantener medios de comunicación con las vendedoras (lo cual a su vez podría involucrar contratar empleados dedicados únicamente a esto). Para las vendedoras, las cuales generalmente trabajan al tiempo con mas de dos catálogos de distintas empresas,

manejar la logística de la distribución de los productos a los clientes finales es un reto técnico demandante de tiempo y propenso a errores que terminan costándole a ambas partes. Las vendedoras, además, en la mayoría de los casos deben de pagar un costo de inscripción a las empresas y comprar cada catálogo, enfrentarse con valores de orden mínimos prohibitivamente altos o comisiones por debajo del promedio. Adicionalmente, algunas empresas pueden implementar medidas restrictivas para las vendedoras al enterarse que éstas venden catálogos de otras empresas, incrementando el valor de la orden mínima que debe ser pedida, reduciendo los porcentajes de comisión o inclusive prohibiendo expresamente a la vendedora de usar otros catálogos.

Las vendedoras por catálogo y las empresas componen un segmento del mercado que históricamente ha sido olvidado por la tecnología, específicamente, por alguna tecnología que busque resolver los problemas de la venta directa de manera general. Por esta razón, en el año 2019 se crea la empresa Valienta S.A.S, cuyo propósito principal es empoderar a las vendedoras (denominadas Valientas) mediante la tecnología, buscando una solución a sus problemas al momento de vender por catálogo. De manera complementaria, Valienta busca también solucionar los problemas que afrontan las empresas al implementar el canal. La empresa se materializa como marketplace que atiende a este mercado bilateral, partiendo de una idea sencilla: conectar a las vendedoras con las empresas y solucionar los problemas mencionados anteriormente para ambas partes.

## **1.2. Justificación**

De manera empírica, la empresa (Valienta) se ha dado cuenta que la experiencia de usuario de las vendedoras es un factor clave para el éxito de ellas al utilizar la plataforma. De hecho, al ser utilizada en su mayoría por públicos no expertos en el uso de tecnologías como computadores o smartphones, para la empresa es especialmente retador implementar una plataforma fácil de usar para las vendedoras. Por esta razón, la empresa busca implementar cambios en la plataforma constantemente para mejorar la UX de las vendedoras, además de agregar nuevas funcionalidades. Los impactos de una mala UX para las vendedoras se reflejan en su éxito en la plataforma, particularmente, en las vendedoras nuevas o potenciales. Es común para Valienta recibir preguntas de las vendedoras sobre como realizar acciones aparentemente básicas dentro de la plataforma. Particularmente, las vendedoras potenciales tienden a rechazar a la empresa pues en ocasiones no pueden encontrar fácilmente catálogos que ellas están acostumbradas a vender.

Adicionalmente, los catálogos pueden ser compartidos (por las vendedoras a sus clientes) solamente en su presentación completa, limitando sus opciones al tratar de visualizar y compartir productos específicos dentro de los catálogos, una funcionalidad que está en el core de otras empresas que implementan ventas directas, tales como

Elenas y Meesho. Debido a lo anterior, la empresa ha decidido invertir tiempo en implementar mejores maneras de visualizar catálogos y productos. Sin embargo, al tratar colocarse de intermediario entre empresas y vendedoras, Valienta no cuenta con la información completa de los productos de un catálogo, y obtener esta información de las empresas proveedoras es imposible en la práctica. Por este motivo, este trabajo de grado busca utilizar la información que si está disponible para Valienta (los archivos de catálogos de venta) con el propósito de implementar una solución basada en ciencia de datos al problema de categorización de los productos para mejorar la experiencia de usuario de las vendedoras.

### **1.3. Descripción del Problema**

Actualmente no es posible visualizar ordenadamente mediante categorías la variedad de los catálogos y productos individuales en la plataforma de Valienta. Esto se debe a la falta de metadatos descriptivos en las imágenes de los productos en catálogos, como efecto de la dificultad de obtener aquellos datos de las empresas proveedoras y la imposibilidad de realizar una labor de etiquetado manual. Como consecuencia, la información sobre los catálogos disponibles en la plataforma no es observada por usuarios existentes/potenciales, y la posibilidad de introducir nuevas funcionalidades necesarias en la plataforma se ve limitada, impactando negativamente la experiencia de usuario y potencialmente disminuyendo el éxito de las vendedoras.

### **1.4. Objetivo General**

Desarrollar un modelo de detección de productos en catálogos de venta directa, automatizando la categorización de catálogos y segmentación de productos para el descubrimiento y uso por parte de los usuarios de la plataforma.

### **1.5. Objetivos Específicos**

- a) Crear un conjunto de datos etiquetado con imágenes de productos en catálogos de venta directa extraídos de la plataforma Valienta
- b) Construir un modelo capaz de detectar productos en catálogos
- c) Validar las predicciones del modelo basado en un conjunto de catálogos de prueba

## 2. Antecedentes

### 2.1. Marco Teórico

#### 2.1.1. Aprendizaje profundo y redes neuronales convolucionales

El aprendizaje automático es el estudio de algoritmos computacionales que aprenden a partir de la experiencia (datos) (Mitchell, 1997). Puede ser utilizado en una gran variedad de dominios para resolver problemas complejos basado en los datos disponibles. Las herramientas empleadas pueden utilizarse para cumplir tareas de manera supervisada o no supervisada. El aprendizaje supervisado se basa en datos, de los cuales algunos son determinados por el usuario o problema como variables objetivo, donde la meta es lograr predecir dichas variables dado un conjunto de variables predictores. Cuando los datos no cuentan con una o varias variables objetivo, se emplean técnicas de aprendizaje no supervisado para entender la estructura y relaciones de los datos (James et al., 2000).

El aprendizaje supervisado encuentra una función entre un conjunto de datos con variables predictoras y un conjunto de variables de respuesta, por ejemplo, tratar de determinar el salario (variable respuesta) de una persona dado sus años de experiencia y años de estudio (variables predictoras). Dos grupos de tareas comunes para las cuales es posible aplicar aprendizaje supervisado son la predicción y la inferencia. La predicción se preocupa por determinar valores futuros para una o varias variables dado un conjunto de variables predictores. La inferencia trata de aumentar el conocimiento de los datos a través del entendimiento de la función de mapeo encontrada (por ejemplo, cuales variables son mas importantes en un modelo de regresión lineal) (James et al., 2000). En el caso de este trabajo, la tarea a realizar es la detección (identificación) de objetos. Se busca, marcar las regiones rectangulares (cajas) de una imagen que pertenecen a un objeto de una determinada categoría (Chollet, 2017).

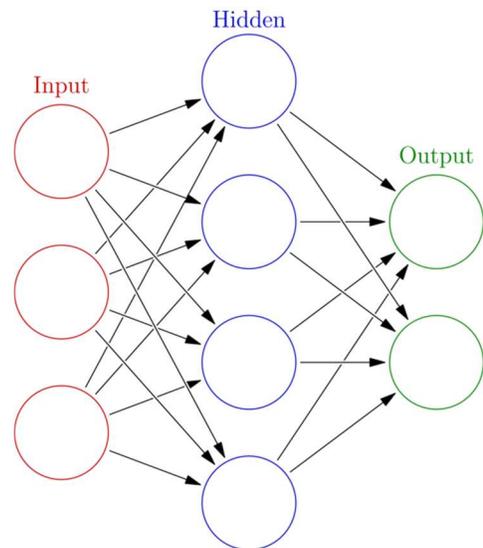


Figura 1 Red neuronal

Dentro del aprendizaje automático, se encuentra la sub disciplina del aprendizaje profundo, el cual busca aproximar a la función mencionada anteriormente mediante capas de extracción de características cuyo poder representativo aumenta proporcionalmente a la profundidad de la capa (Chollet, 2017). En la gran mayoría de los casos, los modelos que habilitan el aprendizaje profundo son las redes neuronales,

donde cada capa se compone de varias “neuronas” (termino inspirado en la biología humana). Para las redes neuronales tradicionales (densas) esto significa que cada neurona de una capa es una función de todas las neuronas de la capa anterior (Figura 1), que se calcula realizando el producto-suma de las conexiones con las neuronas anteriores (pesos) por los valores de las neuronas anteriores, sumándoles opcionalmente un termino de sesgo y finalmente pasando el resultado por una función de activación para otorgar no linealidad (Chollet, 2017). Estas redes se entrenan para aprender a realizar mejor su tarea – minimizar una función de medida del error, comúnmente denominada función de perdida – mediante un algoritmo llamado back propagation (Rumelhart et al., 1986), el cual ajusta los pesos de la red tomando como entrada el error generado al comparar una predicción de la red contra el resultado real sobre cada uno de los datos de entrada.

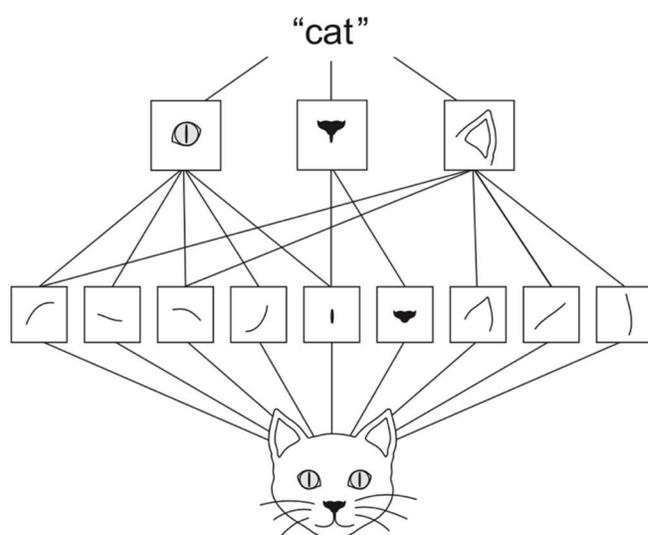


Figura 2 Jerarquía de características aprendidas por deep learning (Chollet, 2017)

Desde su creación, las redes neuronales han evolucionado para tomar una gran cantidad de formas (arquitecturas) dependiendo de la labor a realizarse. La forma tradicional es la red neuronal densa, pero existen otras arquitecturas comunes como las redes neuronales convolucionales (CNN) (Fukushima, 1980), o las redes neuronales recurrentes (RNN) (Hopfield, 2018), entre otras. Esta gran variedad de arquitecturas, combinado con el crecimiento del poder computacional y la cantidad de los datos disponibles han convertido al aprendizaje automático en la subdisciplina de la inteligencia artificial más popular y exitosa (Chollet, 2017).

Las redes neuronales convolucionales se diferencian de las redes neuronales densas debido a su capacidad de aprender patrones locales a través de operaciones de convolución sobre los datos de entrada, a diferencia de una red neuronal densa, que aprende patrones globales sobre todo el dato de entrada (Chollet, 2017). Esta diferenciación con las redes neuronales densas hace a estas redes eficientes en cuanto a su entrenamiento, pues patrones locales en los datos no deben ser reaprendidos cuando se presenten en diferentes lugares de la serie de datos, es decir, son invariantes a la translación. Las convoluciones en estas redes se entienden como “filtros” (ventanas deslizantes) de tamaño fijo que realizan operaciones matemáticas sobre los datos para detectar la presencia de una característica, dichos filtros actúan como las neuronas de estas redes. Las CNN son ampliamente usadas en el campo de la visión por computadora, donde los filtros entrenados buscan características locales en las

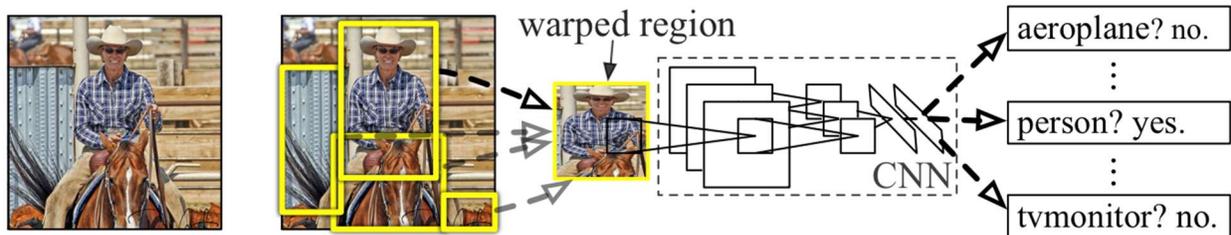


Figura 3 Metodología de R-CNN (Girshick et al., 2014)

imágenes, las cuales pueden posteriormente ser combinadas en características de mas alto nivel representativo. Por este motivo, estas redes cumplen la propiedad de generar jerarquías de características (Figura 2) que pueden ser posteriormente la entrada a redes densas que decidan la relevancia de las características y tomen la decisión final.

### 2.1.2. Familia de modelos R-CNN

Para la tarea de detección de objetos en imágenes, la cual consiste en marcar regiones de interés (RoI) de una imagen como pertenecientes a una categoría (clase) se ha desarrollado una familia de modelos basados en Region Convolutional Neural Networks (R-CNN) (Girshick et al., 2014). Estos modelos pueden ser utilizados para detectar regiones rectangulares (también conocidas como cajas) en las que se encuentra un objeto de una categoría – esta tarea es denominada detección de objetos –, o para clasificar cada píxel individual de cada objeto perteneciente a una clase en una imagen, esta tarea es conocida como segmentación de instancias. A continuación, se describen los modelos R-CNN, Fast R-CNN y Faster R-CNN, los tres primeros modelos de esta familia de modelos, los cuales dieron origen a otros modelos para otras tareas, como Mask R-CNN (He et al., 2020) y Cascade R-CNN (Cai & Vasconcelos, 2018) para segmentación semántica/panóptica.

#### 2.1.2.1. Region Convolutional Neural Network (R-CNN)

Este modelo consiste en utilizar una combinación de varias técnicas: un algoritmo de generación de RoI, CNN's y maquinas de soporte vectorial (SVM) (Cortes & Vapnik, 1995). Primero, utilizando el algoritmo selective search (Uijlings et al., 2013), se extraen de cada imagen en el conjunto de datos múltiples RoI – que representan regiones candidato a contener un objeto de alguna de las clases – cuya forma se compone de: coordenadas de la esquina superior izquierda, altura y ancho. En el siguiente paso, cada RoI es recortada y redimensionada a un tamaño cuadrado fijo, posteriormente, cada RoI redimensionada es pasada por una CNN para extracción de sus características. Finalmente, las características resultantes de la región son clasificadas por una SVM especifica para la categoría de la imagen (Figura 3) (Girshick et al., 2014).

### 2.1.2.2. Fast R-CNN

Fast R-CNN (Figura 4) ofrece mejoras a R-CNN en cuanto a velocidad de entrenamiento/inferencia, simplicidad de entrenamiento y exactitud. Esto se logra reutilizando las características de una imagen para cada una de sus regiones propuestas y eliminando del todo el uso de SVM específicas a cada clase. En vez de realizar una pasada por toda la CNN para cada región como R-CNN, Fast R-CNN realiza una sola pasada de la imagen y posteriormente cada RoI es proyectado al mapa de características resultante. Después, la RoI proyectada es pasada por una capa de Max Pooling especial para extraer las características en un tamaño fijo. Posterior a esto, cada mapa de características es expandido en un vector de características por medio de dos capas densas, cuyo resultado es entregado finalmente a las capas finales de clasificación y regresión de las coordenadas de la región final (Girshick, 2015).

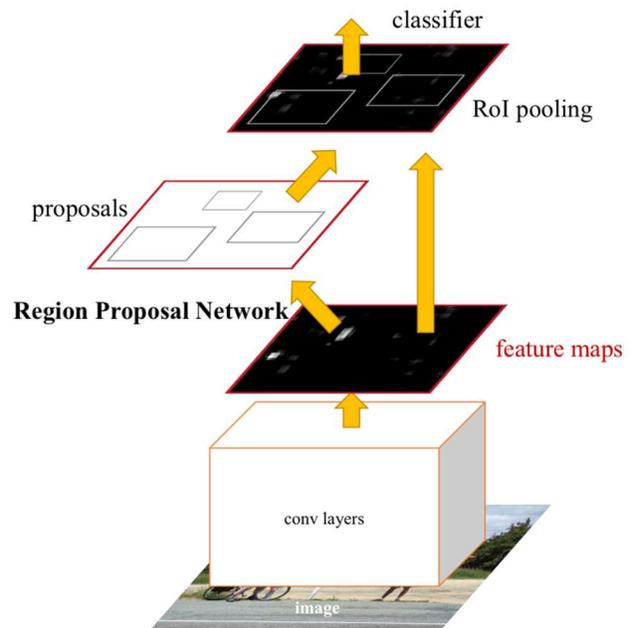


Figura 4 Metodología de fast R-CNN (Girshick, 2015)

### 2.1.2.3. Faster R-CNN

Mejorando en los mismos aspectos en los que Fast R-CNN mejoro sobre R-CNN, Faster R-CNN se convierte en el modelo baseline para las tareas de detección y segmentación de instancias. La diferencia principal con el modelo anterior radica en la generación de las RoI, mientras que las versiones anteriores de R-CNN usaban selective search, Faster R-CNN incorpora la generación de RoI dentro de la arquitectura del modelo. Para generar las RoI, Faster R-CNN utiliza una CNN pequeña denominada Region Proposal Network (RPN) (Figura 5), la cual es entrenada junto con el resto del modelo para producir RoI de manera más rápida y exacta que los modelos anteriores. Posterior a generar las RoI, estas son utilizadas junto con un mapa de características producido por la primera capa CNN de la red y son entrada para la capa de RoI pooling de Fast R-CNN, después de la cual se sigue el proceso de igual forma que con Fast R-CNN (Ren et al., 2017).

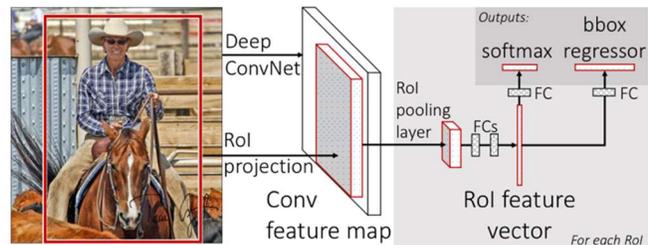


Figura 5 Metodología de Faster R-CNN (Ren et al., 2017)

Faster R-CNN mejora la exactitud, velocidad y simplicidad de

utilizar Fast R-CNN, demostrando mejores métricas de evaluación sobre conjuntos de datos utilizados en los proyectos de MSCOCO (Lin et al., 2014), PASCAL VOC (Everingham et al., 2010) e ILSVRC (Russakovsky et al., 2015). Adicionalmente, la primera CNN de la red, también conocida como el backbone, puede ser intercambiada por redes entrenadas para otros conjuntos de datos con el objetivo de proveer preentrenamiento.

### 2.1.3. Aprendizaje por transferencia

La técnica del aprendizaje por transferencia (transfer learning) se emplea con el motivo de mejorar la exactitud de un modelo a entrenarse, aprovechándose de modelos entrenados para problemas relacionados. Dentro del aprendizaje por transferencia, existen dos técnicas: extracción de características y afinamiento. La extracción de características consiste en utilizar las capas de una red neuronal entrenada como extractores de características, sobre las cuales, un clasificador nuevo es colocado. Lo anterior cumple el objetivo de entrenar solamente el clasificador nuevo, reduciendo el número de parámetros y disminuyendo el sobreajuste del modelo. El afinamiento es similar a la extracción de características, diferenciándose solo en que algunas capas de la red previamente entrenada son “desbloqueadas” para ser entrenadas junto con el clasificador nuevo, permitiendo la reformación de los detectores de características más abstractas, afinándolos al nuevo caso de uso (Chollet, 2017). Comúnmente, esta técnica es utilizada en arquitecturas como Faster R-CNN para servir de “backbone”, la primera red convolucional pre entrenada, por la cual se producen los mapas de características de las imágenes.

### 2.1.4. ResNet

ResNet es una arquitectura de redes neuronales con conexiones residuales entre capas. La motivación detrás de estas redes neuronales es habilitar el entrenamiento de redes cuya exactitud aumente proporcional al número de capas. Antes de ResNet, existía un problema de “degradación”, por medio del cual redes neuronales empeoraban en rendimiento en cuanto su número de capas aumentaba, imposibilitando el uso de redes con un gran número de capas. Esta arquitectura introduce conexiones residuales entre bloques de capas de la red por medio de sumas de características anteriores a características posteriores a los bloques. Esta técnica introduce el “aprendizaje residual” donde los bloques de la red son motivados a aprender mapeos de características residuales, permitiendo el entrenamiento local de los bloques en vez de un solo mapeo entre los datos de entrada y salidas de la red (He et al., 2016). Como resultado de la aparición del aprendizaje residual, ResNet (particularmente sus variantes convolucionales con 101 y 152 capas) se ha convertido en el backbone por excelencia para tareas de detección de objetos, y por tanto es escogido como el backbone para utilizar con Faster R-CNN en este proyecto.

### 2.1.5. Aumento de datos (data augmentation)

Data augmentation, o aumento de datos, es una técnica de expansión de conjuntos de datos para lidiar con sobreajuste de modelos. Consiste en aumentar artificialmente un conjunto de datos, generando variaciones de los datos levemente distintos a los datos originales. Al hacer esto, un modelo es entregado un mayor número de ejemplos por medio de los cuales puede generalizar características. Particularmente, en imágenes, esto se refleja como transformaciones a las imágenes por medio de rotaciones, cambios de color, tamaño, deformaciones, introducciones de ruido, etc. Como resultado de utilizar esta técnica, es posible incrementar la exactitud de modelos que no disponen de grandes cantidades de datos, pues, comparativamente, el modelo tendrá una menor cantidad de parámetros a entrenar por cada observación del conjunto de datos. En el contexto de este proyecto, se utilizará data augmentation por medio de flips aleatorios a las imágenes.

## 2.2. Estado del Arte

La aplicación de la detección de objetos ha sido aplicada en el contexto de la venta al por menor, como se observará a continuación. Sin embargo, estas aplicaciones se han enfocado específicamente en resolver problemas dentro de tiendas minoristas, o con imágenes de productos específicos no agrupados en un catálogo de venta con múltiples categorías amplias. Los métodos encontrados serán evaluados con respecto a:

- a) Habilidad de detectar productos de múltiples clases amplias (ropa, maquillaje, herramientas, etc.)
- b) Contexto de las imágenes (estantes, fotografías individuales, catálogos, máquinas de venta automática, etc.)
- c) Uso de aprendizaje profundo
- d) Uso de datos textuales adicionales a las imágenes

En 2015, Gül Varol en la Universidad de Oxford propuso una técnica de detección de marcas distintas de cigarrillos en estantes de tiendas. Su técnica hizo uso de una cascada de clasificadores seguidos por una SVM. Este estudio se basó únicamente en imágenes de un tipo de producto (Varol & Kuzu, 2015). Dos años después, en la Universidad Politécnica de Catalunya (UPC), se utilizaron redes neuronales convolucionales y procesamiento de lenguaje natural para detectar el producto principal en imágenes de productos de revistas de moda, tratándose como un problema de clasificación binaria sobre las detecciones de todos los productos en la imagen (Yu et al., 2017). Este estudio logró obtener una exactitud de clasificación de productos principales sobre las imágenes mayor al 90%. En 2019, investigadores de Auto-ID Labs utilizaron CNNs para detectar 90 productos diferentes en 15000 instancias pertenecientes a fotografías de máquinas de venta, obteniendo porcentajes de exactitud cerca al 95% para las clases con mayor cantidad de instancias (Fuchs et al., 2019). Finalmente, en enero

del 2020, la empresa ParallelDots creó un detector de productos sobre estantes utilizando Faster R-CNN, capaz de obtener una precisión promedio de 56% para distintos datasets de productos disponibles al público (Varadarajan et al., 2020).

Como se puede observar, no existe un método en el estado del arte que resuelva la detección de productos con múltiples categorías sobre catálogos de venta. Esta diferencia es significativa debido a que los catálogos presentan particularidades comparados con estantes, máquinas de venta e imágenes de productos individuales aumentadas con descripciones textuales. En un catálogo de ventas no se cuenta con datos textuales separados de la imagen, como en el método de la UPC, por lo tanto, esto implica tener que extraerlos por medio de OCR, algo que se escapa del alcance de este trabajo, adicional a que el problema resuelto por la UPC (el más similar al problema de este trabajo) no es compatible para clasificar los distintos productos, solo para detectar el principal. La tabla 1 presenta un cuadro comparativo entre los distintos métodos.

<b>Característica\Método</b>	<b>Oxford</b>	<b>UPC</b>	<b>Auto-ID Labs</b>	<b>ParallelDots</b>	<b>Este trabajo</b>
Múltiples clases amplias	No	No	<u>Si</u>	<u>Si</u>	<u>Si</u>
Contexto de las imágenes	Estantes	Revistas de moda	Maquinas de venta	Estantes	<u>Catálogos</u>
Uso de aprendizaje profundo	No	<u>Si</u>	<u>Si</u>	<u>Si</u>	<u>Si</u>
Uso de datos adicionales a las imágenes	<u>No</u>	Si	<u>No</u>	<u>No</u>	<u>No</u>

Tabla 1 Comparación del estado del arte

### 3. Metodología

El modelo de ciclo de vida de proyecto de analítica escogido es CRISP-DM. Es un proceso de estándar abierto que describe los acercamientos comunes en la analítica por los científicos de datos. Consiste en iterar continuamente sobre un conjunto de fases que parten del entendimiento de un dominio específico hasta el despliegue final de la solución (Wirth, 2000). Sus fases consisten en (Figura 6):

- Entendimiento del negocio: Entendimiento de los objetivos y requerimientos desde un punto de vista de negocio, conversión en una definición de problema de ciencia de datos
- Entendimiento de los datos: Recolección de datos y familiarización (búsqueda de problemas y conocimientos iniciales)
- Preparación de los datos: Preparación del conjunto de datos final, este servirá como entrada a la etapa de modelamiento; incluye limpieza, selección de atributos y observaciones e ingeniería de características
- Modelamiento: Selección y aplicación de diversas técnicas de modelamiento, calibración óptima de parámetros
- Evaluación: Evaluación meticulosa de los pasos ejecutados en la construcción del modelo, comparación contra los objetivos de negocio
- Despliegue: Presentación final de los resultados, puede involucrar la creación de un proceso repetible de minería de datos, o ser tan sencillo como la generación de un reporte

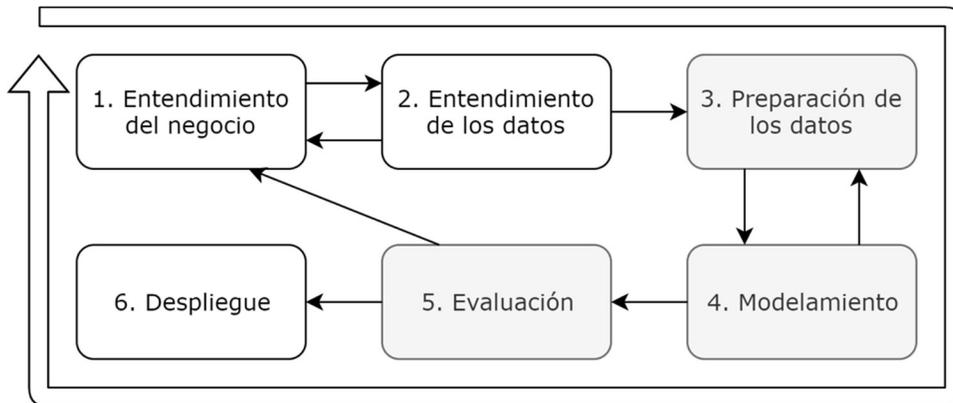


Figura 6 Ciclo de vida de un proyecto de analítica según CRISP

Para la realización del proyecto, se iterará en el ciclo completo (exceptuando el despliegue) de CRISP-DM de una manera ad-hoc, hasta cumplir los objetivos específicos. Las primeras etapas tendrán el propósito de implementar la infraestructura y flujo automatizado para la recolección y preparación de los datos; su entrada al modelo y subsecuente evaluación serán realizadas manualmente. Las siguientes etapas se enfocarán principalmente en el modelamiento y evaluación, buscando mejorar el modelo al máximo nivel; posiblemente iterando adicionalmente en la etapa de preparación de datos.

## 4. Preparación del conjunto de datos

Para cumplir el primer objetivo específico se exploraron las imágenes y categorías presentes en los catálogos de venta y la automatización de su transformación a imágenes utilizables para etiquetar. Previo al procesamiento de los catálogos, fue necesario responder las preguntas:

- ¿Cuántos catálogos y productos existen?
- ¿Qué categorías de productos se encuentran actualmente en los catálogos?
- ¿Cómo se distribuyen los productos en las distintas categorías?
- ¿Cuáles variedades de presentación grafica de productos existen?
- ¿En que formato se encuentran los catálogos y que consideraciones especiales tienen?

Mediante análisis exploratorio de datos sobre los PDF de los catálogos – único formato en el cual se almacenan – se logró encontrar un número total de 33 catálogos. El total de productos dentro de estos catálogos es de 4381, en un total de 1250 páginas, resultando en un promedio de 132 productos por catálogo. El número de productos encontrado corresponde al número de instancias de productos dentro de una página de catálogo asociados con una referencia a la venta, en otras palabras, aquello que sea un producto, pero no esté asociado a la referencia del producto de la página del catálogo no es considerado para la venta, y por ende no es considerado para ser detectado. En la Figura 7 se puede observar un ejemplo, en el cual dos productos están a la venta: una blusa y un pantalón, sin embargo, en esta imagen están presentes otros productos que podrían ser vendidos, pero no son considerados, como un bolso, zapatos, aretes y una pulsera. Adicionalmente, como se observa al lado izquierdo de la Figura 8, productos que se componen de dos o mas “subproductos” se deben considerar como una sola referencia y por lo tanto se etiquetan como si fuesen un solo producto. También, existen casos en los que un producto puede



Figura 7 Ejemplo de imagen con multiples productos pero solo un par a la venta

ser



Figura 8 Izquierda: Ejemplo de imagen con dos productos que cuentan como uno. Derecha: Distintas variantes de un producto

representado por varias imágenes de sus diferentes variaciones (Figura 8 lado derecho), en este caso cada variación individual se considera un producto.

Dentro de los catálogos, se encontró un total de 13 categorías, las cuales fueron asignadas para categorizar de manera “amplia” los productos. Un ejemplo de esta categorización es la clase *clothes* (ropa), donde se incluyen prendas de todo tipo, como pantalones, camisetas, chaquetas, etc, o la clase *accessories* (accesorios) en la cual se incluyen aretes, collares, pulseras, entre otras. La motivación detrás de estas categorías fue crear clases donde cada una tuviese la mayor cantidad de ejemplos cohesivos, debido a la poca disponibilidad de datos. En la

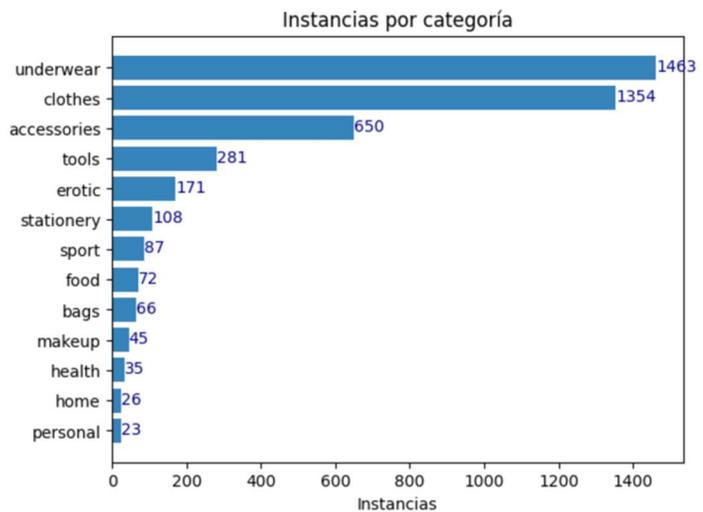


Figura 9 Número de instancias por categoría

Figura 9 es posible visualizar la cantidad de instancias en cada clase.

En total, los catálogos pesan 3.3 GB, y se encontró una alta variabilidad entre los tamaños. Como ejemplo, un catálogo particular de 5 páginas pesa 93.5 MB, mientras que otro de 100 páginas pesa 23.4 MB. Esta discrepancia entre los tamaños se atribuye a la naturaleza de un archivo PDF, pues estos archivos son contenedores de texto, fuentes, imágenes, gráficos vectoriales y hasta código, organizado por medio de capas. Lo anterior significa que un PDF puede almacenar imágenes de tamaños arbitrarios, donde cada una con su propia resolución y espacio de color (Adobe, 2008).

Para el etiquetado y posterior entrada de los catálogos a un modelo de Faster R-CNN era necesario convertirlos en imágenes y etiquetar los productos. Para lograr el preprocesamiento, se diseñó un pequeño pipeline utilizando varios servicios de AWS (Amazon Web Services), incluyendo: Lambda, Step Functions, EC2 y S3, y como herramienta de etiquetado, una aplicación web open source llamada ImgLab. El pipeline completo de preprocesamiento y etiquetado se compone de varias fases (ver figura 10). Primero, los catálogos a convertir en imágenes (una imagen por página) son subidos a un bucket específico de S3, después, la maquina de estados en Step Functions es disparada para obtener los nombres de los catálogos y procesar cada catálogo en paralelo mediante las funciones lambda necesarias, las cuales hacen una petición HTTP a una maquina EC2 que realiza la transformación real y subsecuente subida a S3 como imágenes separadas. La maquina en EC2, la cual se configuró como una instancia t3a.xlarge con 4 núcleos y 16 GB de RAM, es capaz de procesar 2 catálogos al tiempo con utilización de 100% en CPU sin sufrir deterioros en los procesos normales del sistema operativo (como conectividad de red) y completa su trabajo de procesar los 33 catálogos en aproximadamente 20 minutos.

La herramienta utilizada para convertir los catálogos en imágenes es ImageMagick. Esta se configuró para aplanar las capas interiores de los PDF en una sola, con el motivo de estandarizar el espacio de color de las imágenes, labor fundamental para obtener imágenes finales fieles a la intención del creador del PDF, de no hacer esto, ImageMagick

entregaba imágenes con regiones en colores diferentes y con diferencias de luminosidad dependiendo de la capa. Adicionalmente,

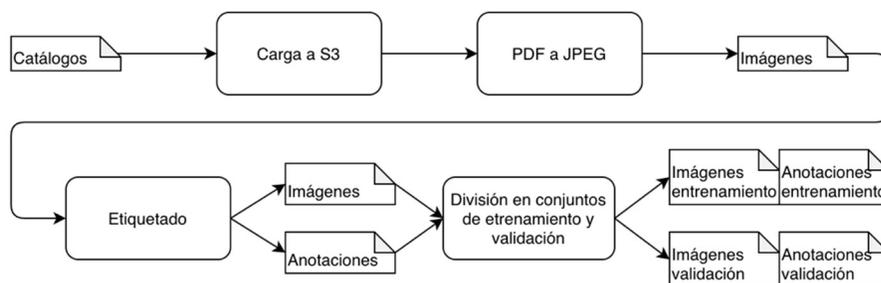


Figura 10 Metodología de preparación de datos

se determinó una densidad de 150 pixeles por pulgada, un redimensionamiento a las imágenes verticales a un ancho fijo de 512 y las imágenes horizontales a ancho fijo de 768. Con esta configuración, fue posible obtener imágenes relativamente pequeñas en peso y con la calidad necesaria para no perder muchos de sus detalles.

El paso final de la preparación del conjunto de datos consistió en etiquetar en formato COO cada imagen utilizando ImgLab. Para completar esta tarea, el etiquetado se realizó catálogo por catálogo, produciendo un JSON de anotaciones para cada catálogo, los cuales fueron posteriormente unificados en un solo JSON de anotaciones de todas las imágenes por medio de una función escrita en Javascript. Finalmente, utilizando otra función de Javascript, se realizó la división estratificada aleatoria no sesgada (utilizando el algoritmo de Fisher-Yates (Odeh & Knuth, 1969)) de todas las anotaciones en dos JSON de anotaciones para entrenamiento y validación con un porcentaje 20% de las anotaciones en el conjunto de validación.

## **5. Modelamiento mediante Faster R-CNN con detectron2**

Partiendo de los datos etiquetados y divididos, se utilizó detectron2 para realizar la experimentación con diferentes variaciones sobre Faster R-CNN. Este framework, liberado al público en octubre del 2019, hace fácil realizar experimentos con diferentes configuraciones de modelos para distintas tareas, desde clasificación de imágenes hasta segmentación semántica, detección de puntos clave o estimación de postura humana (Güler et al., 2018). Comparativamente, detectron2 es mas rápido y exacto en entrenamiento que detectron, parcialmente debido a que su implementación en PyTorch permitió a sus desarrolladores ajustar a bajo nivel el rendimiento de los modelos. Adicionalmente, es extensible a modificaciones por el usuario, lo cual permite reutilizar módulos ya implementados junto con código nuevo.

Los experimentos realizados con Faster R-CNN consistieron en entrenar este modelo con diferentes backbones e hiperparametros. Primero se escogió el mejor backbone y posteriormente se experimentó con distintos hiperparametros. Durante la experimentación, se tuvo cuidado de controlar el sobreajuste, utilizando el conjunto de validación periódicamente durante el entrenamiento para evaluar el desempeño del modelo contra el loss. Adicionalmente, se realizó data augmentation por medio de un flip aleatorio para cada imagen de entrenamiento.

En cuanto a backbones, se utilizaron distintas variantes de ResNet, un modelo desarrollado con el propósito de resolver el deterioro del entrenamiento en las redes

neuronales en cuanto aumenta de número de capas, permitiendo el entrenamiento de redes mas profundas (He et al., 2016). Este tipo de redes son el estado del arte (particularmente ResNeXt (He et al., 2016)) en cuanto a ser utilizados como backbone de los modelos con mejor desempeño en la tarea de detección de objetos, como DetectoRS (Qiao et al., 2020). Por lo tanto, se utilizan tres backbones distintos: ResNet con 101 capas (el baseline original en Faster R-CNN), ResNet con 101 capas y dilataciones en las ultimas capas convolucionales (conv5) y ResNeXt con 101 capas y cardinalidad 8.

Posterior a entrenar los tres modelos baseline con distintos backbones, se procedió a realizar un barrido de hiperparametros considerando:

- Número de iteraciones (teniendo cuidado de observar en cual iteración se empezaba a genera overfitting)
- Tasa de aprendizaje base para descenso de gradiente estocástico con momentum (0.0001, 0.001, 0.005 y 0.01)
- Número de RoI propuestas por la RPN (64, 128, 256, 512 y 1024)
- Tamaño del lote de imágenes (4, 8 y 16)
- Capas congeladas de ResNeXt (desde bloques 2, 3 y 4)

En cuanto a la tasa base de aprendizaje, esta se incremento linealmente hasta terminar la primera época y estabilizándose posteriormente. Una vez encontrada la mejor configuración por medio del AP (average precision, explicado en la siguiente sección) el modelo final fue entrenado con todos los datos por un total de 2000 iteraciones, valor en el cual se alcanzo la mayor precisión promedio (AP). El rastreo de todas las métricas se realizo utilizando la herramienta web de Weights and Biases.

En cuanto al ambiente en el cual se corrieron las pruebas, las primeras tres para decidir cual backbone utilizar fueron realizadas en Google Colab, cuyo entorno provee una GPU Nvidia K80, la cual fue suficiente para las pruebas iniciales. Posterior a esto, las variaciones de hiperparametros y entrenamiento final se realizaron sobre maquinas EC2 p3.8xlarge con 4 Nvidia Tesla V100, 32 nucleos de CPU y 244 GB de RAM. Durante la experimentación, los porcentajes pico de utilización de CPU, Memoria y GPU fueron de 34.4%, 18.4% y 100% respectivamente. Finalmente, el modelo fue exportado en formato pth por PyTorch.

## 6. Validación

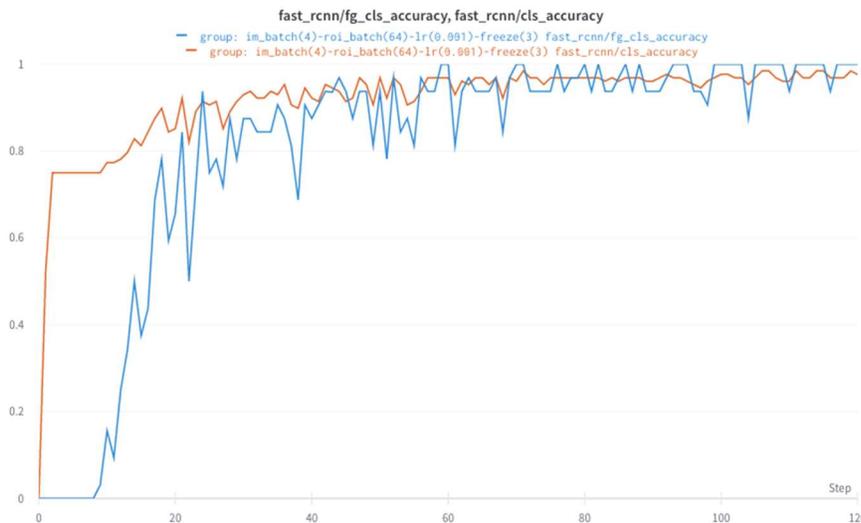


Figura 11 Exactitud de clasificacion en entrenamiento de objetos en primer plano y total. Eje horizontal en Steps, 1 Step = 20 iteraciones

La validación del modelo consistió en el análisis de las métricas del entrenamiento con el conjunto de validación. En cuanto al entrenamiento, las tres métricas más importantes fueron: `loss_box_reg` (loss en regresión de la región rectangular de las instancias), `loss_cls` (loss en clasificación de los objetos de las regiones predichas) y `total_loss`

(`loss` total, definido como la suma de todos los `losses`, incluyendo los de la RPN que no se incluyen en este trabajo), un ejemplo de estas puede ser observado en la Figura 12. Otras métricas utilizadas durante el entrenamiento fueron: `fg_cls_accuracy` (exactitud de clasificación de objetos en primer plano, es decir, aquellos que pertenecen a la clase general “objeto”) y `cls_accuracy` (exactitud de clasificación de los objetos detectados), estas se pueden observar en la Figura 11.

Durante la validación se observó que las métricas de entrenamiento parecieran converger rápidamente, aproximadamente alrededor de 2000 iteraciones. Las curvas producidas por las métricas de entrenamiento se utilizaron con el propósito de observar alrededor de qué valores era posible minimizar el `loss` y cuándo el modelo podría estar aprendiendo más rápido o más lento. En este proceso, se observó que el `loss_cls` siempre convergía rápidamente a valores cercanos a cero, mientras que el `loss_box_reg` se comportaba de maneras diferentes en cada experimento (siendo particularmente sensible al tamaño de lote de imágenes y RoI), por este motivo el afinamiento del modelo se enfocó principalmente en obtener mejores resultados en cuanto a la regresión de las

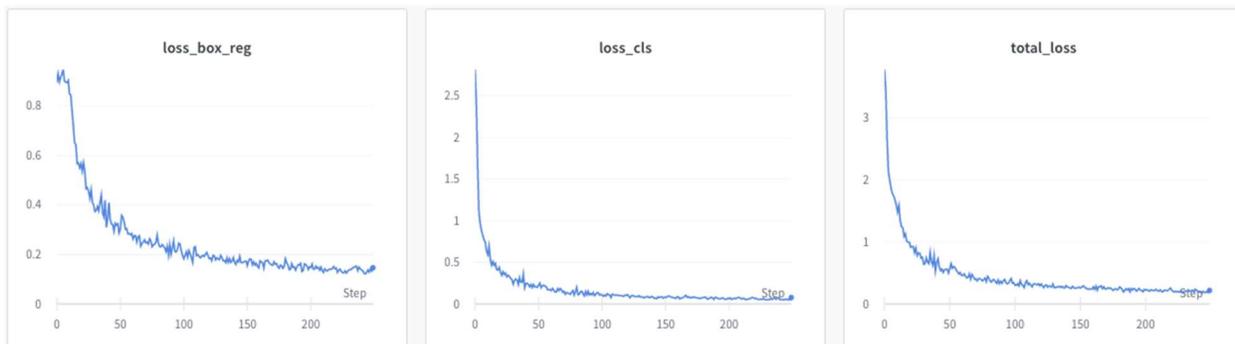


Figura 12 Métricas de evaluación para regresión de la caja, clasificación y loss total

coordenadas de la región rectangular a predecir, pues esta métrica no converge a cero con igual rapidez que *loss\_cls*. También, durante el entrenamiento se observó el comportamiento de las métricas de prueba con el propósito de encontrar alrededor de que número de iteraciones el modelo empezaba a sobreentrenarse.

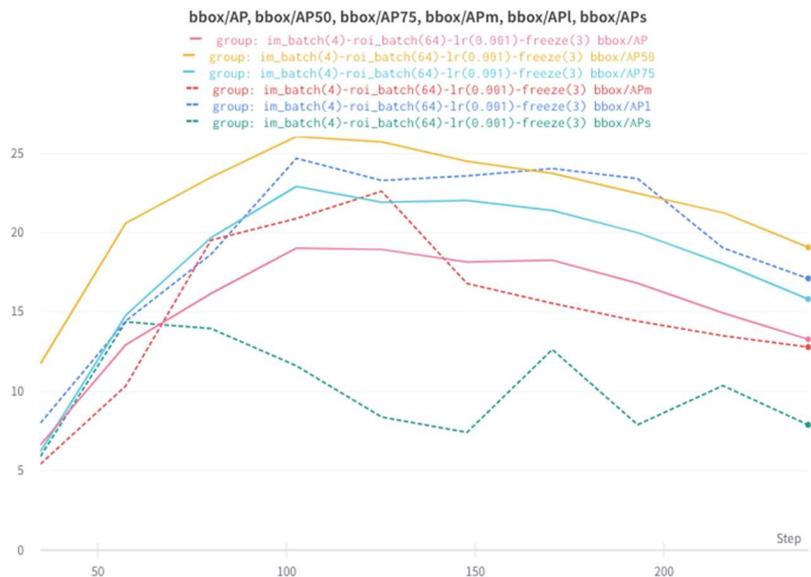


Figura 13 Curvas de AP del modelo entrenado

Las métricas utilizadas con el conjunto de validación son todas las variantes del Average Precision según utilizado en la competencia de COCO (Lin et al., 2014). Esta métrica consiste en encontrar la precisión de predicción basada en la intersección sobre la unión (IoU) de la caja predicha y la caja real, controlado por un umbral del IoU para contar una detección como verdadera. La evaluación de estas métricas se realiza por medio de las distintas variantes de AP, con valores de 0.5 (AP50), 0.75 (AP75), y el promedio entre todas las AP para valores de 0.5 a 0.95 con pasos de 0.05 (AP), así como APs, APm y API para objetos pequeños, medianos y grandes. Como es de esperarse, al evaluar el modelo sobre el conjunto de validación usando estas métricas, es fácil observar que en el step 100 el modelo esta siendo sobreentrenado, al compararlas con las métricas de entrenamiento.

## 7. Resultados

En la Figura 13 se pueden observar los resultados de las métricas de AP sobre el conjunto de validación para el modelo entrenado. Como se puede ver, la mayoría de las curvas tienen un pico en 102 pasos (2040 iteraciones), lo cual equivale a un poco más de 7 épocas. En este es el punto se puede observar una tendencia decreciente en las curvas, significando que el modelo ha comenzado a sobre ajustarse, lo anterior es notable cuando se compara con la curva de *total\_loss* en la Figura 11, la cual sigue decreciendo. La curva que da un mayor valor es la de AP50, la cual es la tradicionalmente usada para las competencias de detección de imágenes como PASCAL VOC (Everingham et al., 2010). Es particularmente interesante la manera en la que esta es mayor a las otras. Una causa posible de esto podría ser el hecho de que es la que tradicionalmente se utiliza en las competencias y por lo tanto los modelos han sido pensados y pre-entrenados principalmente teniendo en cuenta esta métrica, por otro lado, podría ser también una particularidad para el conjunto de datos. Siguiendo a AP50

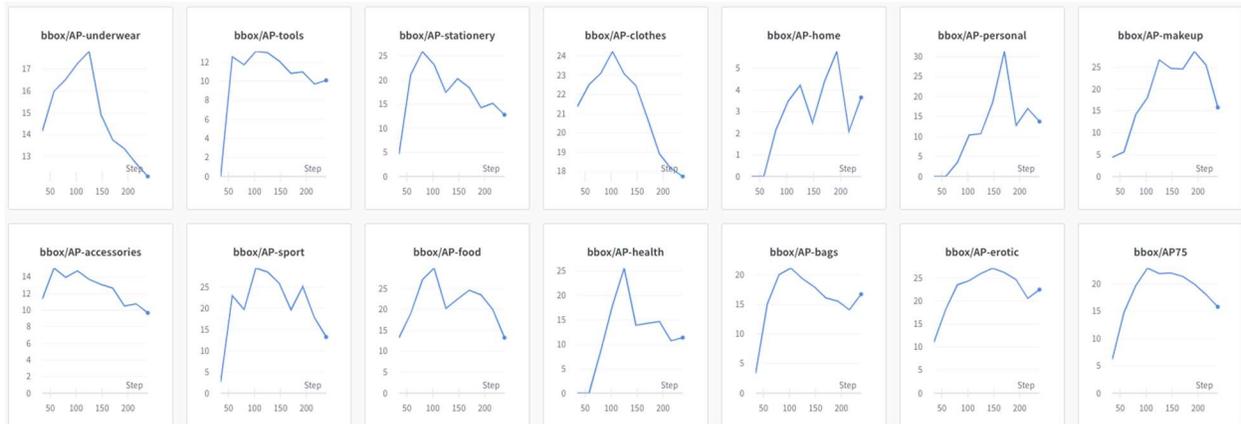


Figura 14 Curvas AP para cada clase

se encuentra AP75, lo cual es natural, pues un IoU de 0.5 es menor a 0.75 y por ende más permisivo en cuanto a marcar una región como una detección exitosa. Después, sigue AP, la cual es menor que AP50 y AP75, sugiriendo que la precisión del modelo es inversamente proporcional al umbral de IoU, lo cual significa que sus detecciones tienden a estar desfasadas con respecto a la región real. Es interesante comparar la precisión entre objetos de diferentes tamaños, en este caso, tanto API y APm son considerablemente mayores a APs, lo cual indica que la detección de objetos pequeños es menos precisa; API es mayor a APm por una pequeña cantidad, sugiriendo que el modelo logra generalizar la detección de objetos de grandes y medianos tamaños.

En el paper original de Faster R-CNN, utilizando VGG-16 (Simonyan & Zisserman, 2015) como backbone y utilizando 300 RoI propuestas por la RPN, se logró un AP de 21.9, comparativamente, un AP de 19 para este conjunto de datos logra acercarse, lo cual indica que todavía hay oportunidades de mejora en el modelo, esto es más evidente comparando el AP50, que es de 42.7 en el paper y 25 en este trabajo. Sin embargo, esta comparación no es del todo justa, pues este conjunto de datos y COCO se diferencian tanto en número de clases como en los detalles de las imágenes de entrenamiento y sus cantidades.

Las discrepancias entre diferentes niveles del umbral de IoU presenta una oportunidad de mejora en el modelo, debido a que curvas de AP con diferentes umbrales deberían tender a parecerse en cuanto aumenta la precisión del modelo, particularmente en la regresión de la caja a predecir. Adicionalmente, la curva de AP de objetos pequeños no logra incrementar hacia un pico, sugiriendo que es difícil para el modelo aprender a reconocer objetos pequeños. Estos problemas con el modelo son atribuibles a la poca cantidad



Figura 15 Objeto con cajas "anidadas"

de datos disponibles para su entrenamiento, el desbalance de las clases y la precisión y cuidado con el cual se etiquetan las regiones en el conjunto de datos de entrenamiento; en la Figura 14 se puede observar las curvas de AP para cada clase, notablemente, las clases con menor cantidad de observaciones (como home, personal, makeup y health) tienen mas varianza y algunas no logran obtener valores tan altos como las clases con más instancias (clothes, underwear).

Las detecciones en accessories es un caso particular para resaltar, pues es la tercera clase mas numerosa, y sin embargo no alcanza valores altos de precisión, esto es parcialmente atribuible a la complejidad de las imágenes y lo inadecuado que es utilizar solo etiquetas de cajas para el entrenamiento. En la Figura 15 se puede observar el problema, para el etiquetado de estas imágenes, habria un total de 4 cajas, para los dos aretes y dos collares, las cajas de los objetos mas grandes contendrian los mas pequeños y un considerable cantidad de espacio vacio, lo cual dificulta el aprendizaje sobre estos objetos. Ademas de esto, los accesorios pueden ser objetos pequeños y grandes. En los anexos se colocan imágenes con predicciones para cada categoria.

## 8. Conclusiones

Por medio del proceso completo de crear un modelo para clasificar productos en catálogos de venta es posible resaltar lo siguiente:

- Los efectos de una cantidad baja de datos y clases desbalanceadas
- La importancia del hardware, el framework y otras herramientas para la rápida experimentación
- El tipo de etiquetas utilizadas para el entrenamiento (cajas vs mascarlas)
- La importancia de la automatización del pipeline de preprocesamiento y etiquetado

Primero, sobre los efectos de la cantidad baja de datos y desbalanceo de clases, para un problema con categorías variadas y desbalanceadas, como el de detectar productos en catálogos de venta, es de gran importancia entrenar el modelo con cantidades mayores de datos, con el propósito de disminuir la varianza en las métricas (las cuales al tener bajas cantidades de instancias dependen de instancias específicas, causando varianza) y obtener una mayor exactitud. Para mitigar este problema en futuros experimentos, será necesario conseguir una mayor cantidad de datos, forma de hacer esto son: solicitar a los proveedores de la empresa sus catálogos pasados, encontrar datasets de catálogos de venta públicos y utilizando datos generados de manera sintética (Nikolenko, 2019). Referente al hardware, framework y otras herramientas, para realizar mas de 90 experimentos con distintas configuraciones del modelo, fueron

fundamentales: el framework de detectron2, que provee una forma fácil de configurar hiperparámetros del modelo sin necesitar modificar el código de las redes neuronales directamente; el hardware utilizado, en el cual se pueden correr experimentos rápidamente, como ejemplo, 2000 iteraciones con un tamaño de lote de 4 imágenes para las 1120 imágenes de entrenamiento y aproximadamente 4400 instancias toman menos de 9 minutos en ejecutarse; finalmente, Weights and Biases como herramienta para rastrear los experimentos provee un conjunto de características esenciales, que superan de manera extrema las capacidades de una herramienta estándar como Tensorboard.

Debido a la naturaleza de los datos, no es suficiente etiquetarlos con cajas, adicionalmente deberían ser etiquetados con mascarar para tener la capacidad de representar de manera mas exacta los objetos y tener la capacidad de usar modelos del estado del arte como DetectoRS en segmentación de instancias, los cuales brindan métricas de AP para COCO de 54.7. Para la agilidad de experimentación, también fue importante automatizar las tareas de transformación y preparación de datos, reduciendo la cantidad de tiempo invertido en actividades repetitivas.

Finalmente, a pesar de las oportunidades de mejora identificadas, el modelo entrenado es capaz de detectar con mayor precisión los objetos de clases numerosas, esto, combinado con un trabajo futuro que implemente las mejoras propuestas demuestra el potencial de utilizar la detección de objetos por medio de Faster R-CNN o modelos mas recientes para una futura aplicación practica.

## Bibliografia

- Adobe. (2008). *Document management-Portable document format-Part 1: PDF 1.7 PDF 32000-1:2008 ii*.  
[http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=51502](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=51502). It is being made available from the website of Adobe Systems Incorporated
- Cai, Z., & Vasconcelos, N. (2018). Cascade R-CNN: Delving into High Quality Object Detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. <https://doi.org/10.1109/CVPR.2018.00644>
- Chollet, F. (2017). Deep Learning with Python. In *Deep Learning with Python* (1st ed.). Manning Publications Co. <https://doi.org/10.1007/978-1-4842-2766-4>
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/bf00994018>
- Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88(2), 303–338. <https://doi.org/10.1007/s11263-009-0275-4>
- Fuchs, K., Grundmann, T., & Fleisch, E. (2019, October 22). Towards identification of packaged products via computer vision: Convolutional neural networks for object detection and image classification in retail environments. *ACM International Conference Proceeding Series*. <https://doi.org/10.1145/3365871.3365899>
- Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4), 193–202. <https://doi.org/10.1007/BF00344251>
- Girshick, R. (2015). Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision, 2015 Inter*, 1440–1448. <https://doi.org/10.1109/ICCV.2015.169>
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 580–587. <https://doi.org/10.1109/CVPR.2014.81>
- Güler, R. A., Neverova, N., & Kokkinos, I. (2018). DensePose: Dense Human Pose Estimation in the Wild. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 7297–7306. <https://doi.org/10.1109/CVPR.2018.00762>
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2020). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2), 386–397. <https://doi.org/10.1109/TPAMI.2018.2844175>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-Decem*, 770–778.

<https://doi.org/10.1109/CVPR.2016.90>

- Hopfield, J. J. (2018). Neural networks and physical systems with emergent collective computational abilities. In *Feynman and Computation* (Vol. 79).  
<https://doi.org/10.1201/9780429500459>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2000). An introduction to Statistical Learning. In *Current medicinal chemistry* (Vol. 7, Issue 10).  
<https://doi.org/10.1007/978-1-4614-7138-7>
- Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8693 LNCS(PART 5), 740–755.  
[https://doi.org/10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48)
- Mitchell, T. M. (1997). *Machine Learning* (1st ed.). McGraw-Hill, Inc.
- Mouthón, L. (2019, April 18). *Negocio de venta directa mueve \$4,8 billones al año | El Heraldo*. <https://www.elheraldo.co/economia/negocio-de-venta-directa-mueve-48-billones-al-ano-621270>
- Nikolenko, S. I. (2019). *Synthetic Data for Deep Learning*.  
<http://arxiv.org/abs/1909.11512>
- Odeh, R. E., & Knuth, D. E. (1969). The Art of Computer Programming. Volume 1: Fundamental Algorithms. In *Journal of the American Statistical Association* (Vol. 64, Issue 325). Addison Wesley Longman Publishing Co., Inc.  
<https://doi.org/10.2307/2283757>
- Qiao, S., Chen, L.-C., & Yuille, A. (2020). *DetectoRS: Detecting Objects with Recursive Feature Pyramid and Switchable Atrous Convolution*.  
<http://arxiv.org/abs/2006.02334>
- Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149.  
<https://doi.org/10.1109/TPAMI.2016.2577031>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536.  
<https://doi.org/10.1038/323533a0>
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3), 211–252. <https://doi.org/10.1007/s11263-015-0816-y>
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. <http://www.robots.ox.ac.uk/>

- Uijlings, J. R. R., Van De Sande, K. E. A., Gevers, T., & Smeulders, A. W. M. (2013). Selective search for object recognition. *International Journal of Computer Vision*, 104(2), 154–171. <https://doi.org/10.1007/s11263-013-0620-5>
- Varadarajan, S., Kant, S., & Srivastava, M. M. (2020). Benchmark for Generic Product Detection: A Low Data Baseline for Dense Object Detection. *Image Analysis and Recognition*, 30–41. [https://doi.org/10.1007/978-3-030-50347-5\\_3](https://doi.org/10.1007/978-3-030-50347-5_3)
- Varol, G., & Kuzu, R. S. (2015). Toward retail product recognition on grocery shelves. *Sixth International Conference on Graphic and Image Processing (ICGIP 2014)*, 9443, 944309. <https://doi.org/10.1117/12.2179127>
- Wirth, R. (2000). CRISP-DM : Towards a Standard Process Model for Data Mining. *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, 24959, 29–39. <https://doi.org/10.1.1.198.5133>
- Yu, L. L., Simo-Serra, E., Moreno-Noguer, F., & Rubio, A. (2017). Multi-modal Embedding for Main Product Detection in Fashion. In *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017* (Vols. 2018-Janua). <https://doi.org/10.1109/ICCVW.2017.261>

## Anexos

Sport

**AB-STRAPS**  
AUMENTO DEL RENDIMIENTO DEPORTIVO

**DESCRIPCIÓN**  
Ideal para realizar trabajos de core en suspensión. Funciona como sostenimiento para el cuerpo.

**ESPECIFICACIONES**  
Fabricado en lana con material acolchado por dentro para mayor comodidad. Mangos de espuma para facilidad en el agarre, costura acanalada. Incluye 2 mosquetones de alpinista.

sport 97% sport 98%

sport 95% sport 93%

\$53.550

COLOR:

**DESCRIPCIÓN**  
 Herramienta para el entrenamiento de fuerza, resistencia y potencia muscular. Permite trabajar la totalidad de los músculos del cuerpo.

**ESPECIFICACIONES**  
 Relleno de hierro y arena. 8kg: 100% arena, 10kg: 70% arena 30% Hierro, 15kg: 20% hierro 30% arena 20kg: 100% hierro Permite peso gravitacional ubicado en 19 botellas distribuidas en toda la superficie del chaleco.

**PEGOS**

8KG	\$141.650	15KG	\$165.450
10KG	\$153.550	20KG	\$177.350

**OTRAS TEXTURAS/COLORES:**

1 2 3 4 5  
6 7 8 9

**CHALECOS DE PESO**  
 RECUPERACIÓN/AUMENTO DEL RENDIMIENTO DEPORTIVO

clothes 76%

sport 96%

Clothes

clothes 97%

clothes 97%

**KIMONO RITA**  
 \$105.000

TALLA: ÚNICA  
 REF: E\_13

© amaria.col

clothes 96%

clothes 96%

**Leggins**  
 SEMI CONTROL

Antes \$35.999 **Ahora**  
**\$32.999** POR SÓLO  
 CADA UNO

EXTRA COMODIDAD  
 MOSTRILLA TALLA DEL

Leggins: 90% alto en  
 100% algodón con elastano  
 20% algodón

Ref. 6961  
 Ref. 6962  
 Ref. 6963  
 Ref. 6964

MA 17

Accessories

accessories 93%

accessories 98%

Anillo Espíritu  
Ref.P01-4621  
Ver Pág. 52

www.MUJERBAMBUROSA.COM  
65

accessories 98%

TOP FÉNIX

Ref. 1000348  
\$41.000  
USD 13.64  
Juego con piedras naturales  
Acero

accessories 94%

accessories 92%

accessories 97%

Ref. 181876  
\$19.000  
USD 5.76  
Pulsera Aborcada Con  
Filas de  
Acero

Ref. 180419  
\$17.000  
USD 5.33  
Aretes Dorsales en ganchos  
Acero

accessories 90%

Ref. 181883  
\$78.000  
USD 24.47  
Pulsera Color Oro-Rosa esmeralda  
Cristal  
Disponibles en Tallas 6-7

accessories 81%

Al comprar este producto,  
siempre que te llega alguno de  
colores disponibles según sexo

www.shoptopsecret.com

5

Underwear

Sin varilla de acero.

accessories 78%

underwear 98%

underwear 95%

underwear 90%

underwear 81%

underwear 98%

underwear 93%

underwear 93%

underwear 92%

BRASIER  
Antes \$25.999  
Ahora  
**\$23.999**

Ref. 53870  
Modelo con varilla de acero y  
cinturón. Disponible en colores  
negro, gris, rojo, verde,  
naranja, azul, rosa, blanco.  
Tallas: M, L, XL, XXL.  
Colores Surtidos.

BRASILERA  
Antes \$9.999  
Ahora  
**\$7.999**

Ref. 53871  
Modelo con varilla de acero y  
cinturón. Disponible en colores  
negro, gris, rojo, verde,  
naranja, azul, rosa, blanco.  
Tallas: M, L, XL, XXL.  
Colores Surtidos.

BRASIER  
Antes \$25.999  
Ahora  
**\$25.999**

Ref. 53878  
Modelo con varilla de acero y  
cinturón. Disponible en colores  
negro, gris, rojo, verde,  
naranja, azul, rosa, blanco.  
Tallas: M, L, XL, XXL.  
Colores Surtidos.

BRASILERA  
Antes \$9.999  
Ahora  
**\$7.999**

Ref. 53879  
Modelo con varilla de acero y  
cinturón. Disponible en colores  
negro, gris, rojo, verde,  
naranja, azul, rosa, blanco.  
Tallas: M, L, XL, XXL.  
Colores Surtidos.

Paquete Especial

underwear 98%

underwear 97%

BRASIER SEÑORERO \$68.900  
PAQUETE X2

Ref. 119110 / Brazer clásico copa ancha / Copa estructurada sin aro / Mayor cubrimiento en cña / Cargueles ajustable / Nylon elastico.  
"Se envían 2 brazers de la misma talla, el color se envía surtido, según disponibilidad de inventario."  
Talla: M, L, XL, XXL  
Colores Surtidos

Paquetes x2

underwear 98%

underwear 98%

PANTY SEÑORERO \$39.900  
PAQUETE X2

Ref. 219194  
Talla media / Control de abollones / Nylon elastico.  
Talla media / Control suave de abdomen / Lencería sola / Escote, nylon elastico.  
"Se envían 2 pantalones de la misma talla, el color se envía surtido, según disponibilidad de inventario."  
Talla: M, L, XL, XXL  
Colores Surtidos

10

Erotic

Lleva cada uno por solo \$19.999

<b>A. LUBRICANTE ÍNTIMO BASSIKA CEREZA</b> X 100ML REF. 10447	<b>B. LUBRICANTE ÍNTIMO BASSIKA WHISKY</b> X 100ML REF. 10448	<b>C. LUBRICANTE ÍNTIMO BASSIKA RÓN</b> X 100ML REF. 10449
<b>D. LUBRICANTE ÍNTIMO BASSIKA OVA</b> X 100ML REF. 10450	<b>E. LUBRICANTE ÍNTIMO BASSIKA TEQUILA</b> X 100ML REF. 10451	<b>F. LUBRICANTE ÍNTIMO BASSIKA CHICLE</b> X 100ML REF. 10452

erotic 89% erotic 80% erotic 78% erotic 94% erotic 87% erotic 89%

*sabores que enloquecen*

El lubricante puede usarse para el juego erótico, sexual y penetración. Te hará vivir momentos de placer increíbles. Base de agua, no mancha, no deja residuos en el cuerpo, seguro con papigay y preservativos.

# erimental

el placer

**erotic 84%**  
**LUBRICANTE ÍNTIMO EFECTO FRÍO**  
El precio por unidad es de \$16.999  
Intensifica el placer y brinda una óptima lubricación. Hidrosoluble, no grasoso. Evita la sequedad. Delicioso sabor a menta.

**erotic 96%**  
**LUBRICANTE ÍNTIMO EFECTO NEUTRO**  
El precio por unidad es de \$16.999  
Incrementa tu placer y te brinda una óptima lubricación. Evita la sequedad. Disuélvase en agua. Delicioso sabor a caramelo.

**erotic 87%**  
**LUBRICANTE ÍNTIMO EFECTO CALIENTE**  
El precio por unidad es de \$16.999  
Disfruta al máximo la íntima parte variación del lubricante caliente. Hidrosoluble. Te hará vivir momentos de placer increíbles. Delicioso sabor a fresa. Aplicar en el área deseada, regular para elevar la temperatura.

**erotic 87%**  
**LUBRICANTE ÍNTIMO EFECTO NEUTRO**  
El precio por unidad es de \$16.999  
Su PH balanceado evita la irritación y la sequedad. Disuélvase en agua. Incluye un sabor Aguardiente.

*Precaución!* Si presenta irritación suspende su uso y consulte al médico. Manténgase en un lugar fresco y seco, fuera del alcance de los niños. Producto solo para adultos.

Stationery

**Calendario 2020 \$19.900**  
Con lindas frases  
with phrases

stationery 94% stationery 94% stationery 94% stationery 86% stationery 92% stationery 90%

**Hojas Repuesto \$10.900**  
Tamaño Carta (Cuaderno Grande)

CUADRICULADA RAYADA

Dot's

<b>stationery 97%</b> ESTUDIANTE El precio por unidad es de \$16.999 Grande   Cuaderno grande   48 páginas Pequeño   Cuaderno grande   48 páginas	<b>stationery 97%</b> INGENIERA El precio por unidad es de \$16.999 Grande   Cuaderno grande   48 páginas Pequeño   Cuaderno grande   48 páginas	<b>stationery 95%</b> MÉDICO El precio por unidad es de \$16.999 Grande   Cuaderno grande   48 páginas Pequeño   Cuaderno grande   48 páginas
<b>stationery 98%</b> MAMA El precio por unidad es de \$16.999 Grande   Cuaderno grande   48 páginas Pequeño   Cuaderno grande   48 páginas	<b>stationery 97%</b> MERCADÓLOGA El precio por unidad es de \$16.999 Grande   Cuaderno grande   48 páginas Pequeño   Cuaderno grande   48 páginas	<b>stationery 97%</b> NUTRICIONISTA El precio por unidad es de \$16.999 Grande   Cuaderno grande   48 páginas Pequeño   Cuaderno grande   48 páginas
<b>stationery 96%</b> ODONTÓLOGA El precio por unidad es de \$16.999 Grande   Cuaderno grande   48 páginas Pequeño   Cuaderno grande   48 páginas	<b>stationery 98%</b> PROFESORA El precio por unidad es de \$16.999 Grande   Cuaderno grande   48 páginas Pequeño   Cuaderno grande   48 páginas	<b>stationery 96%</b> PUBLICISTA El precio por unidad es de \$16.999 Grande   Cuaderno grande   48 páginas Pequeño   Cuaderno grande   48 páginas

Tools



**Abarrotes**

**Coca Cola Sabor Original**  
1500 ml  
~~\$3.505~~  
**\$3.505**

**Pony Malta**  
1000 ml  
**\$2.550**

**Jugo de Fruta Hit**  
1000 ml  
**\$3.100**

**Jugo de Fruta Hit Surtido**  
5x6 | 200 ml  
**\$6.125**

**Abarrotes**

**Leche en Polvo Klim**  
360 g  
**\$11.990**

**Leche en Polvo El Rodeo**  
200 g  
**\$5.790**

**Leche en Polvo El Rodeo**  
380 g  
**\$9.790**

**Milo Econopack**  
250 g  
**\$9.190**

**Chocolate Cruz**  
250 g  
**\$6.390**

**Chiper** Ahora podrás encontrar todo para tu hogar en un solo lugar **ta.**

**Chiper** Ahora podrás encontrar todo para tu hogar en un solo lugar **ta.**

**Bags**

**bags 84%**

**bags 89%**

**bags 96%**

**bags 98%**

**\$27.999**  
Ref. 12163

Manos libres pequeña. Elaborado en sintético y forro interno textil. Color: Negro, rojo, dorado, miel y plata.

Ancho: 22 cm  
Alto: 26 cm  
Profundidad: 6 cm

**bags 92%**

**bags 97%**

**bags 93%**

**bags 95%**

**\$23.999**  
Ref. 12177

Billetera pequeña. Elaborada en sintético y forro interno textil. Color: Negro, amarillo, plata, rojo y azul.

Ancho: 14 cm  
Alto: 18 cm  
Profundidad: 2 cm

**BOLSO FRATTINI**  
\$89.899  
EXTRALARGA  
TALLA ÚNICA  
PROBES

**bags 71%**

**BOLSO FRATTINI**  
\$69.899  
EXTRALARGA  
TALLA ÚNICA  
PROBES

**bags 93%**

**BOLSO FRATTINI**  
\$89.899  
EXTRALARGA  
TALLA ÚNICA  
PROBES

**BOLSO ACC. AC**  
\$89.899  
EXTRALARGA  
TALLA ÚNICA  
PROBES

**bags 86%**

**bags 84%**

**bags 93%**

Makeup



**Kit Alimentos**

- 1 Aceite Olisoya 420ml
- 2 Arroz Diana 500g
- 1 Harina Doña Anpa 500g
- 1 Spagueti San Remo 250g
- 1 Leche en Polvo el Rodeo 380g
- 1 Sal Alta Pureza Refital 500g
- 1 Azúcar Kropala 500g
- 1 Atún en Aceite La Soberana 170g
- 1 Sardina La Soberana 425g
- 1 Maíz Pira - Crispeta El Trece 460g
- 1 Fíjoi Cargamento Rojo 460g
- 2 Lenteja El Trece 460g
- 5 Sobres Suntea

**\$49.635**



**Kit Aseo Personal**

- 1 Health & Shoulders 180g
- 1 Papel Higiénico Scott 4 Rollos
- 1 Barra de Jabón Protex 130g
- 1 Nosotras diarios MultiEstilos
- 1 Lady Speed Stick Clinical 30ml
- 1 Speed Stick Clinical 30ml
- 1 Crema Dental Colgate 60ml
- 1 Cepillo Dental Colgate

**\$29.046**



**Kit Aseo Hogar**

- 1 Limpido Multisusos 460ml
- 1 Limpiaojos Lavanda Cliner 960ml
- 1 Lavajozza Líquido Cliner 500ml
- 1 Desengrasante Clínico Cliner 500ml
- 1 Barra Jabón Rey 300g
- 1 Asion Limón 150g
- 1 Servilletas Cortadas Favorita 150und
- 1 Ariel Multisusos 225g

**\$14.575**



Ahora podrás encontrar todo para tu hogar en un solo lugar





**Ariel Multisusos**  
225 g  
~~\$2.690~~  
**\$2.420**



**Ariel Multisusos**  
125 g  
~~\$1.490~~  
**\$1.260**



**Fab**  
450 g  
~~\$4.180~~  
**\$3.890**



**Fab**  
100 g  
~~\$1.250~~  
**\$1.150**



**Dersa Ropa**  
500 g  
~~\$3.750~~  
**\$3.125**



**3D Multisusos**  
500 g  
~~\$3.750~~  
**\$3.125**



Ahora podrás encontrar todo para tu hogar en un solo lugar



Hogar