



UNC

FAMAF



CCAD

Centro de  
Computación  
de Alto  
DesempeñoCórdoba  
Technology  
Clustermercado  
libreDiplomatura en Ciencia de Datos, Aprendizaje Automático y sus Aplicaciones 2019

---

# Estimación de peso y dimensiones de los envíos de Mercado Libre

## Materia: Introducción al aprendizaje supervisado

### Análisis del dataset.

### Comunicación de resultados y conclusiones

A partir de lo visto en la teoría de la materia y del segundo laboratorio, diagramar una comunicación en formato textual o interactivo describiendo la solución de las actividades propuestas a continuación. Al final de las mismas se proveen actividades opcionales (no obligatorias) que pueden resultar de interés.

#### Actividades Propuestas:

1. Definir el target. En nuestro caso, proponemos usar la columna `SHP\_WEIGHT` que representa el peso del item en gramos.
2. Hacer un split de train/test del dataset limpio. Se recomienda usar [train test split](#) con 80% para training y 20% para test.
3. Elegir entre 2 modelos distintos vistos en clases (Sugerencia: [LinearRegression](#) y [kNN](#)) y entrenar los modelos.
4. Evaluar y reportar algunas métricas (Sugerimos: Mean absolute error, Median absolute error, Mean Squared Error y Root Mean squared error). Reportar como mínimo dos métricas y entender la diferencia entre ellas. Por ejemplo, responder qué diferencia tiene evaluar usando media vs mediana?
5. Splitear el training set en train y validación (Sugerencia 80-20). Re-entrenar usando el conjunto de train y evaluar en el conjunto de validación con distintos hiper-parámetros los distintos modelos elegidos. Reportar las métricas de las distintas pruebas y describir cual es la mejor elección de hiper-parámetros. Reportar las métricas del mejor modelo sobre el conjunto de test y compararlo con los modelos entrenados en el punto 3.



UNC

FAMAF



CCAD

Centro de  
Computación  
de Alto  
DesempeñoCórdoba  
Technology  
Clustermercado  
libre

## Diplomatura en Ciencia de Datos, Aprendizaje Automático y sus Aplicaciones 2019

---

La comunicación debe estar apuntada a un público técnico pero sin conocimiento del tema particular, como por ejemplo, sus compañeros de clase o stakeholders del proyecto. Idealmente, además del documento se debería generar una presentación corta para stakeholders explicando el análisis realizado sobre los datos y las conclusiones obtenidas de tal análisis.

Se evaluarán los siguientes aspectos:

- El informe debe contener un mensaje claro y presentado de forma concisa.
- Los gráficos deben aplicar los conceptos de percepción visual vistos en clase.
- Se debe describir o estimar la significancia estadística de su trabajo.

## Entrega

La entrega debe realizarse antes del **09-09**. Definan un repositorio en el cual podamos ver el notebook con el que hayan procesado los datos y manden el link al mismo. Pueden utilizar cualquier herramienta para generar el informe y la presentación (sugerimos Google Drive) y pasen el link a dónde podamos verlos.

## Presentación

A charlar. Estaría muy bueno tomarnos entre 30 y 60 minutos, aunque sea por videollamada, para que puedan exponer los resultados y charlamos sobre ellos. El objetivo de esto es hacer más fluída la devolución sobre el laboratorio.