



UNC

FAMAF



CCAD

Centro de  
Computación  
de Alto  
DesempeñoCórdoba  
Technology  
Clustermercado  
libre

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

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

## Materia: Aprendizaje No supervisado

### Análisis del dataset.

### Comunicación de resultados y conclusiones

A partir de lo visto en la teoría de la materia y del cuarto 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. Aplicar [PCA](#) (Análisis de componentes principales) sobre el conjunto de features para reducir su dimensionalidad. Probar con distintos valores del parámetro que determina la cantidad de componentes finales (`n_components`), por ejemplo, 2, 5 y 10. Para cada versión resultante agregar el target y entrenar los 3 mejores modelos encontrados en el práctico anterior y reportar métricas.

Nota: Recordar que para PCA es importante que los features se encuentren normalizados, por lo que recomendamos usar [StandardScaler](#).

2. Aplicar [K-Means](#) para generar clustering sobre los siguientes features propuestos: `SHP_WEIGHT` (Peso físico del ítem) y `SHP_LENGTH` (Largo del ítem) (Probar distintos K: por ejemplo: 5, 10, 15).  
Luego graficar con distintos colores los distintos dominios (variable `DOMAIN_ID`) dentro de cada cluster.

Analizar:

- a. Cuales son los dominios más frecuentes de cada cluster
- b. Hay dominios que se encuentran en más de un cluster? Cuales?

3. [Opcional] Aplicar otro algoritmo de clustering, como [mixtura de gaussianas](#)



UNC

FAMAF



CCAD

Centro de  
Computación  
de Alto  
Desempeño



Córdoba  
Technology  
Cluster



mercado  
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 **28-10**. 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.