



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: Análisis y Visualización de datos

Análisis del dataset.

Comunicación de resultados y conclusiones

A partir de lo visto en la teoría de la materia y del primer laboratorio, diagramar una comunicación en formato textual o interactivo describiendo un aspecto en particular del dataset provisto para esta problemática. Ejemplo:

- Variables más correlacionadas y posibles interpretaciones. Analizar la independencia de variables utilizando el test de chi-cuadrado para variables categóricas.
- Distribución de los ejemplos con respecto a una clase
- Análisis de outliers

Algunas consultas concretas que deben ser respondidas:

- ¿Qué criterio les parece mejor para limpiar el dataset de datos faltantes?
- ¿Cuál es la distribución de los datos en cada una de las dimensiones informadas?
- ¿Qué categorías/dominios presentan más homogeneidad en su peso y dimensiones? ¿Cuáles presentan mayor heterogeneidad?
- ¿Existen outliers en los datos? ¿Cuál es la mejor forma de tratarlos? Si existieran, ¿a qué podrían deberse?

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.



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

Entrega

La entrega debe realizarse antes del **10-05**. 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 30' minutos, aunque sea por teleconf, para que puedan exponer los resultados y charlamos un rato sobre ellos. El objetivo de esto es hacer más fluída la devolución sobre el laboratorio.