Estimación de peso y dimensiones de los envíos de Mercado Libre

Materia: Aprendizaje 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. Splitear el dataset en train/test (80-20). Recordar la utilidad train_test_split de scikit-learn y utilizar los parámetros `random_state` y `stratify` y explicar su función. El target en este práctico será múltiple: *SHP_WEIGHT*, *SHP_LENGTH*, *SHP_HEIGHT* y *SHP_WIDTH*. Esto significa que los modelos deberán predecir 4 valores en simultáneo en vez de 1.

2. Entrenar y evaluar con al menos 3 nuevos modelos (Sugerencias: SVR, RandomForestRegressor, GradientBoostingRegressor, etc.) **Obligatorio:** Probar con una red neuronal. Puede ser de scikit-learn o de alguna otra librería que deseen como keras, pytorch, etc.). Junto con las métricas debe entregarse una breve descripción de cómo funciona cada modelo. **Importante:** Para evaluar, por ejemplo mean_absolute_error provee un parámetro multioutput que debería tomar el valor `*raw_values*` para reportar métricas para cada dimension de output por separado.

3. Para estos nuevos modelos tunear hiper-parámetros. Para las evaluaciones utilizar la técnica de k-fold cross-validation (ver cross-validation) y explicar los resultados.

4. Elegir el mejor modelo entrenado hasta el momento según f1-score. Comparar las métricas de este modelo vs. las métricas de evaluar 4 modelos por separado, un modelo para cada uno de los targets. Interpretar los resultados.

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 **14-10**. 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.