

Proyecto Supermercados:

Objetivo: Hacer un workflow tomando como insumo los datos de ventas de supermercados de [www.datos.gov.ar](http://www.datos.gov.ar) y predecir las ventas del mes siguiente utilizando Prophet.

Los datos.

El sitio [datos.gov.ar](http://datos.gov.ar) proporciona la serie de ventas en supermercados a precios constantes (es decir se remueve el efecto de la inflación) desde enero de 2017. Esta serie es mensual. Al removerse el efecto de movimientos de precios en los artículos los movimientos en esta serie corresponden a variaciones en las cantidades compradas, por lo cual resulta un indicador de cantidades consumidas.

Es por eso que sería de utilidad predecir esta serie para el mes posterior al publicado. Buscando modelos de predicción me encontré con Prophet, un modelo de predicción desarrollado por Facebook.

Por un lado, desarrollé el modelo en forma estática, es decir tomé los datos hasta Julio de 2022, los guardé en un .csv, analicé la performance de Prophet y efectué análisis de cross validation. Este análisis está en la carpeta `jupyter/notebooks/desarrollo`

Una vez que el modelo estuvo aceptable, lo pasé a la carpeta `jupyter/notebooks/deploy`, es decir ese será el modelo de predicción en producción que implementaremos en airflow.

Enfoque del problema con los containers:

Necesitaremos los siguientes containers:

Jupyter: para hacer predicción con Prophet necesitaba esta librería, por lo cual necesité buildear una imagen de Jupyter que contemplara la inclusión de varias librerías (Prophet, plotly y scikit-learn).

Airflow: de igual manera, armé una imagen para que se pudiera cargar las librerías mencionadas.

Ambas imágenes las armé con los comandos en la consola Ubuntu parados en el directorio clonado del repositorio `tpDOCKERdf`: `./dockers.sh build_airflow` y `./dockers.sh build_jupyter`

Finalmente, levantamos el ambiente con `dockers.sh`.

parados en el directorio clonado ingresamos `./dockers.sh start_all`

Airflow

La idea sería que a través de airflow se automatice el proceso de toma de datos y proceso de predicción de manera mensual, ya que los datos se actualizan mensualmente.

Para esto, se creó el DAG `predsuper_dag`

Al ejecutar este DAG, se tomarán los datos de la API de [datos.gov.ar](http://datos.gov.ar), se efectuará el pronóstico con prophet y luego se guardará en la carpeta `airflow/dags/` un csv con tres columnas: fecha, serie original y serie pronosticada (con el pronóstico al mes siguiente).

para stopear el ambiente tenemos que ingresar `./dockers.sh stop_all`