

Trabajo Práctico 2 - Machine Learning

Para el TP2, los alumnos deberán participar en la competencia de kaggle [Predict Future Sales](https://www.kaggle.com/competitions/competitive-data-science-predict-future-sales). El trabajo deberá realizarse en grupos de 2 personas (pudiendo cambiar los grupos del TP1 si lo desean).

Dataset a Utilizar

El dataset a usar está disponible en <https://www.kaggle.com/competitions/competitive-data-science-predict-future-sales/data>.

Los datos están relacionados con las ventas diarias que realiza la empresa de software 1C Company en Rusia, y el objetivo es intentar predecir las ventas que realizarán durante el siguiente mes.

La métrica que se utilizará para la evaluación es **RMSE**.

Análisis Exploratorio

Se pide realizar el análisis de los datos, incluyendo (al menos) 6 visualizaciones interesantes que ayuden a explicar el target. En la entrega deben incluirse al menos un plot de cada tipo:

- Bar plot
- Violin plot
- Box plot
- Heatmap

Baseline

Vamos a construir un modelo muy sencillo para utilizar como baseline y tener referencia de cómo funciona un modelo básico y que tomar como punto de partida para mejorar. En este caso, les pediremos que implementen un regresor lineal y evalúen su funcionamiento.

Modelos

Entrenar (al menos) 2 modelos distintos con búsqueda de hiperparametros (¿cómo conviene elegir los datos de validación respecto de los de train?). Incluir el análisis de los features.

¿Cuáles resultaron más importantes? ¿Cuáles decidieron utilizar en los modelos?

Los modelos deben cumplir las siguientes condiciones:

- Deben utilizar RMSE como métrica de validación (al igual que la competencia).

- Deben medirse en validación y presentar el resultado (no solo contra la competencia).
- Deben ser reproducibles (correr el notebook varias veces no afecta al resultado).
- ¿Cuál es el mejor score en la competencia? (guardar el csv con predicciones para entregarlo después)
- Deben realizar análisis de feature importance.
- Deben realizar feature engineering, encodeando features y generando nuevas features en base a las existentes.

Entrega

La entrega constará de un informe, en donde se detalle toda la información relacionada a los modelos como se describe en el punto anterior. El mismo debe incluir también un link a los notebooks / repositorio utilizado con el código.

Por último, también deberá incluirse un link a un video de no más de 4 minutos en el que deben participar todos los integrantes del grupo, explicando los puntos más importantes o interesantes del trabajo presentado.

Luego de la entrega, coordinaremos para realizar también una breve reunión de cierre con cada grupo.