

# Definición de tema para Trabajo Final EEA

## Alumnos:

Gil, Renso

Morinigo, Leandro

- ¿Cuál es el **tema** (título) que se busca explicar?

Los modelos de machine learning, si bien suelen ser una gran herramienta para realizar predicciones y encontrar patrones en los datos, suelen funcionar como cajas negras, de la que suele ser difícil obtener explicabilidad que sea interpretable en el campo del fenómeno modelado. Nos pareció relevante investigar esta temática en base al trade-off entre explicabilidad y capacidad de predicción de los diferentes modelos, tratado en la materia. Para eso nos proponemos investigar y probar algoritmos que ayudarían a darle interpretabilidad a la predicción realizada por los modelos de machine learning, en función de sus variables de entrada. Específicamente planeamos en principios centrarnos en los algoritmos de SHAP (Shapley Additive Explanation) y Local interpretable model-agnostic explanations (LIME), para evaluar posibilidades, limitaciones y sus desempeños en esa tarea.

- ¿Qué **elementos** son necesarios: datasets, librerías, etc?

Aún no tenemos completamente definido el/los datasets sobre los que se trabajará y realizarán las pruebas, algunos candidatos son:

**Estimaciones agrícolas:** Serie completa de cultivos por campaña, provincia y partido/departamento de Argentina. Contiene datos de: superficie sembrada, cosechada, producción y rendimiento

link: <https://www.datos.gob.ar/dataset/agroindustria-agricultura---estimaciones-agricolas>

**SUBE - Cantidad de transacciones (usos) por fecha:** Serie dia\_transporte, nombre\_empresa, linea, amba, tipo\_transporte, jurisdiccion, provincia, municipio, cantidad, dato\_preliminar

link: <https://www.datos.gob.ar/dataset/transporte-sube---cantidad-transacciones-usos-por-fecha/archivo/transporte>

**Income Dataset:** age, workclass, fnlwgt, education, educational-num, marital-status, occupation, relationship, race, gender, capital-gain, capital-loss, hours-per-week, native-country, income\_>50K

link: <https://www.kaggle.com/code/mammadabbasli/customer-income-prediction>

**Properati Dataset:** link: <https://www.properati.com.ar/data>

- ¿Cuáles son **trabajos/investigaciones previas** sobre este tema?

- **SHAP and LIME: An Evaluation of Discriminative Power in Credit Risk**, Alex Gramegna and Paolo Giudici (Department of Economics and Management, University of Pavia, Pavia, Italy)

link: [\(PDF\) SHAP and LIME: An Evaluation of Discriminative Power in Credit Risk](#)

- **Do Not Trust Additive Explanations**, Alicja Gosiewska and Przemyslaw Biecek (Faculty of Mathematics and Information Science Warsaw University of Technology)

link: <https://arxiv.org/pdf/1903.11420v3.pdf>

- **LIME vs. SHAP: Which is Better for Explaining Machine Learning Models?**, Web Article.

link: [LIME vs. SHAP: Which is Better for Explaining Machine Learning Models? | by Dario Radečić | Towards Data Science](#)

- **ML Interpretability: LIME and SHAP in prose and code**, Web Article.

link: [ML Interpretability: LIME and SHAP in prose and code - Cloudera Blog](#)

- **LightGBM model explained by shap**, Web Article.

link: <https://www.kaggle.com/code/cast42/lightgbm-model-explained-by-shap/notebook>

- **Explainable AI tabular data: LightGBM with LIME**, Web Article.

link: <https://www.kaggle.com/code/teesoong/explainable-ai-tabular-data-lightgbm-with-lime>

- ¿Qué **objetivo** busca el video? ¿Qué se puede esperar que haya aprendido una persona que vea el video?

El objetivo es que la persona que lo vea, conozca del problema de interpretabilidad en los modelos de machine learning y algunas opciones para tratar de solucionar esa dificultad, para finalmente conocer más en detalle al menos dos algoritmos usados para ello, shap y lime, y sobre todo puede hacerse una idea de su funcionamiento, alcance y limitaciones.