

Estructura

Documentación para mkdocs

- [Material for mkdocs](#)
- [Build mkdocs project](#)
- [mkdocstring usage](#)
- <https://www.bbvaaiactory.com/mercury/>

Documentación para OPTUNA

- [Optuna study_params](#)

Auto EDA ([review](#))

- [Pandas Profiling](#)
- [D-Tale](#)
- [Sweetviz](#)
- [Autoviz](#)
- [Titanic dataset](#)

src:

- **eda**
 - distribuciones
 - relaciones univariantes/multivariantes
 - estudio categorías
 - análisis nulos
 - análisis duplicados
 - tests estadísticos (normalidad, igualdad varianzas, etc)
- **feature engineering**
 - **tratamiento outliers**
 - **tree-based models:** -
 - **non-tree-based models:** percentil / log / exp
 - **tratamiento nulos**
 - **tree-based-models:** -999, -1, isnull binary feature, XGBoost can handle
 - **non-tree-based-models:** mean, median, isnull binary feature
 - **time series:** knn, interpolation
 - reconstruir valor
 - encoders
 - **categorical encoders**
 - **tree-based models:** label enc, frequency enc

- non-tree-based models: one-hot enc, interactions
 - datetime encoders (feature-engine)
 - periodicity features (cyclic transform)
 - time since event (holidays)
 - coordinates encoders:
 - distance to key places
 - distance to centers of clusters
 - aggregated statistics
 - **agregaciones**
- escalado
 - tree-based models: -
 - non-tree-based models: MinMax / Standard
- feature selection
 - alta correlación
 - RFE, PCA, KBest
- **modeling**
 - train-test split
 - técnicas reequilibrado (pesos, oversampling, undersampling)
 - model selection (lazy predict ?)
 - **model dispatcher**
 - **tree-based-models** (LGBM classifier de momento)
 - **non-tree-based models**: linear, KNN, NNs
 - hyperparameter tuning (OPTUNA)
 - **cross-validation techniques**
 - **metrics**
 - mover punto trabajo en clasificación (multiclase - argmax)
- **evaluation**
 - plots explicativos métricas
 - **calibración probabilidades** (añadir multiclase)
 - comparación benchmark probabilístico
 - **feature importance** (shap values)
- **utils**
 - optuna dashboard
 - generar informes html
 - gestión errores - check_methods

Legend

NOT STARTED

ON TRACK

BLOCKED OR DELAYED

COMPLETE