

Project plan

Summary

- Acelerador para problemas de ML supervisado regresión y clasificación.
- Funcionalidades: EDA, detección valores atípicos, imputación nulos, generación de variables,
 - selección de variables, optimización de hiper parámetros,
 - evaluación, explicabilidad,
 - generación automática de informes sobre el flujo completo
- Estado del arte: OPTUNA, FEATURE ENGINE, SHAP VALUES
- Aplicaciones:
 - POC rápidas y de alta calidad para propuestas
 - Fácil productivización
 - Ejemplo MERCURY BBVA
 - Futuras formaciones internas
 - Proyecto end-to-end, modular y escalable
 - Proyecto formativo de larga duración
 - Herramienta para participar en hackathons ágilmente. Plataformas:
 - Kaggle
 - DrivenData
 - Analytics Vidhya
 - CodaLab
- Casos uso:
 - Estimación criticidad tickets (BAYER)
 - Estimación quejas en supermercado (MERCADONA)
 - Estimación probabilidad de fuga cliente
 - Estimación concesión de préstamo
 - Estimación batería restante de una dispositivo (SECURITAS)

	Week 1-2 3-10/04/23	Week 3-4 17-24/04/23	Week 5-6 1-8/05/23	Week 7-8 15-22/05/23	Week 9-10 29-05/23	Week 11-12 12-19/06/23	Week 13-14 26-03/06/23	Week 15-16 10/07/23			
Task 1											
Task 2											
Task 3											
Task 4											

Backlog:

- **autoEDA**: investigar
 - buscar dataset completo (num + categóricas)
 - ejecutar informes de las diferentes librerías autoeda
 - comparar informes
- **mkdocs** (estilo, estructura, plugins)
 - enlaces a sklearn y feature engine
- **paquetizar** (docker - requirements, wheel [poetry])
- **tests unitarios**
- **GitLab**
 - CI/CD
 - Bonito README
- **Desarrollo módulos**
 - check_input_format method for params_dict
 - feature_engineering_dict
 - estimators_dict

- cv_technique_dict
- dataset para cat/date encoders
- feature engineering
 - add iterative imputer sklearn

Tasks' description:

- **Task 3:**
 - **Tania:** módulo escalado
 - **Arnau:** apoyo, módulo modeling
 - **Pablo:** dependencias con poetry
 - **Eduardo:** módulo autoeda
 - **Sergio:** completar módulo imputers - isnull, mean/median, iterative imputer sklearn
- **Task 2:**
 - **Jose:** revisar estructura transformers, completar módulo imputers
 - **Tania:** gitlab conn, módulo shap values generalizado
 - **Arnau:** apoyo, estructura módulo de modelado
 - **Pablo:** dependencias con poetry
 - **Eduardo:** gitlab conn, vscode config, venv, trastear con el repo, módulo autoeda
 - **Sergio:** gitlab conn, vscode config, venv, trastear con el repo, módulo outliers
- **Task 1:**
 - **Jose:** añadir función de agregaciones en el módulo FE
 - **Tania:** añadir función shap values en el módulo Evaluation
 - **Arnau:** repo, venv, calibración de probabilidades, estructura transformers, mkdocs
 - **Pablo:** CI/CD en GitLab (versionado, flake8, unitarios, etc)