# MLPILOT MACHINE LEARNING CANVAS GRP 14 Designed for: FSDS Designed by: GROUP 14 Date: April 9, 2025 Iteration: 2 .

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| PREDICTION TASKWhat is the type of task? Which entity are predictions made on? What are the possible outcomes to predict? When are outcomes observed?   1. This app predicts based on the needs of the user. 2. Based on the conversation with the user, the model decides if this is a prediction problem or a classification problem. | DECISIONSHow are predictions turned into actionable recommendations or decisions for the end-user? (Mention parameters of the process / application for this.) Model predictions will be converted to insights via explainable tools (SHAP).  UI will guide user on how to interpret the accuracy, SHAP value and confusion matrix. | VALUE PROPOSITIONWho is the end beneficiary, and what specific pain points are addressed? How will the ML solution integrate with their workflow, and through which user interfaces?  1. Data Analysts and Domain experts with minimal ML experiences. 2. Building models usually requires code and ML knowledge. 3. Upload CSV - Agent driven chat & Model training – Get downloadable model as an output (.pkl or .onnx). | DATA COLLECTIONHow is the initial set of entities and outcomes sourced (e.g., database extracts, API pulls, manual labeling)? What strategies are in place to update data continuously while controlling cost and maintaining freshness?  1. Data source will be from uploaded CSV. 2. NO external APIS or cloud involved. 3. Users will be responsible to keeping their data. | | DATA SOURCESWhere can we get data on entities and observed outcomes? (Mention internal and external database tables or API methods.)  * Local files only (CSV) – provided by the user * No online data or third-party involvement to maintain privacy. * Online databases to be considered for the future. | |
| IMPACT SIMULATIONWhat are the cost/gain values for (in)correct decisions? Which data is used to simulate pre-deployment impact? What are the criteria for deployment? Are there fairness constraints?  1. Incorrect decisions can be assessed using confusion matrices, SHAP values, and performance metrics. 2. Users can simulate model effectiveness on test sets within the app. 3. Deployment to production is manual and under user control. 4. Optional fairness checks (via slkearn) can be introduced. | MAKING PREDICTIONSAre predictions made in batch or in real time? How frequently? How much time is available for this (including featurization and decisions)? Which computational resources are used?  1. **Predictions are made in Batch mode, after model training.** 2. **Prediction is executed in locally.** |  | BUILDING MODELSHow many models are needed in production? When should they be updated? How much time is available for this (including featurization and analysis)? Which computation resources are used?  1. One model per training session 2. Model: XGBoost, Random forest, Logistic Regression or Transformers for NLP. 3. Hyperparameter tuning via Optuna. 4. Resources: CPU/GPU based on user setup. 5. Time: Adaptable depending on the data set size, shown via Progress bar. | | FEATURESWhat representations are used for entities at prediction time? What aggregations or transformations are applied to raw data sources?  1. Autodetect from uploaded dataset if user not mentioned. 2. Categorical: OneHot or ordinal encoding 3. Text: Tokenized using Hugging face. 4. Missing values are imputed (mean/median/constant based on type). | |
|  | MONITORINGWhich metrics and KPIs are used to track the ML solution’s impact once deployed, both for end-users and for the business? How often should they be reviewed? | 1. Accuracy, precision, recall, F1, RMSE depending on task 2. Visualizations: Confusion matrix, SHAP summary plots. 3. Evaluation happens after model trained and is embedded in the report. 4. LTI is used to track progress through the workflow | |  | |  | |

|  |  |  |
| --- | --- | --- |
|  | Version 1.2. Created by Louis Dorard, Ph.D. Licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/). Please keep this mention and the link to [ownml.co](https://www.ownml.co/) when sharing. | [**OWNML.CO**](https://www.ownml.co/) |

# 