This document outlines the structure of the project files and provides instructions on how to run the model for the First Payment Default (FPD) prediction task.

|  |  |
| --- | --- |
| **FILE** | **DESCRIPTION** |
| LucidIntel - FPD Model Workflow.ipynb | The main Jupyter Notebook containing all steps: EDA, preprocessing, feature engineering, model training, evaluation |
| final\_model\_pipeline.pkl | The serialized pipeline pickle object including preprocessing steps and the trained LightGBM model. Used for making predictions on new data. |
| features\_summary.csv | A summary of the final set of features used in the model, including short descriptions and data types. |
| model\_utils.py | Utility script containing functions for transforming the data |
| requirements.txt | A list of Python libraries required to run the notebook and supporting scripts |
| input.json | Input file containing parameters used by the python code |
| TE.pkl | TargetEncoder object used to encode fields in data (used by python code) |
| FPD\_Model\_Report.pdf | Report outlining end-to-end process of model development |
| Interview\_data\_set 3.csv | Dataset used for this problem |
| tree.png | Sample image denoting the final model tree |
| model\_inference.ipynb | Jupyter notebook to run the pipeline and get predictions on dataset |

**HOW TO RUN THE MODEL ON NEW DATA**

* Set Up Your Environment. Ensure Python 3.8+ and dependencies are installed (first cell of notebook will install all libraries. Comment out that code if already installed.
* Enter path (or filename if in same folder) of the prediction dataset in “file\_name” field of input.json
* Run Inference notebook – model\_inference.ipynb