



ML model pipeline

▼ Difficulty	
☰ Skills	

Car loan default prediction:

Find default car loan clients

Task description

Brief

A lending company is specialized in providing car loans to retailer clients. To protect their portfolio and facilitate loan process decisions, they would like to train a machine learning model to predict if this applicant is going to default.

Dataset

To train the default prediction model, a dataset with demographical and loan history data of 110000 car loan clients is provided, along with label `loan_default`.

Access the link to download `car_loan_trainset.csv` and `column_definition.csv`.

links:

https://october-data-exercises.s3.eu-west-1.amazonaws.com/datasets/car_loan_trainset.csv

https://october-data-exercises.s3.eu-west-1.amazonaws.com/datasets/column_definition.xlsx

Evaluation

Build a pipeline to:

1. Ingest this data into a local Postgres db
2. Train a simple model (as a Data engineer, please do not spend much time on training - you can create a simple logistic regression model) and save it as `.pkl` locally
3. Run predictions on the entire dataset and store the them into another Postgres table