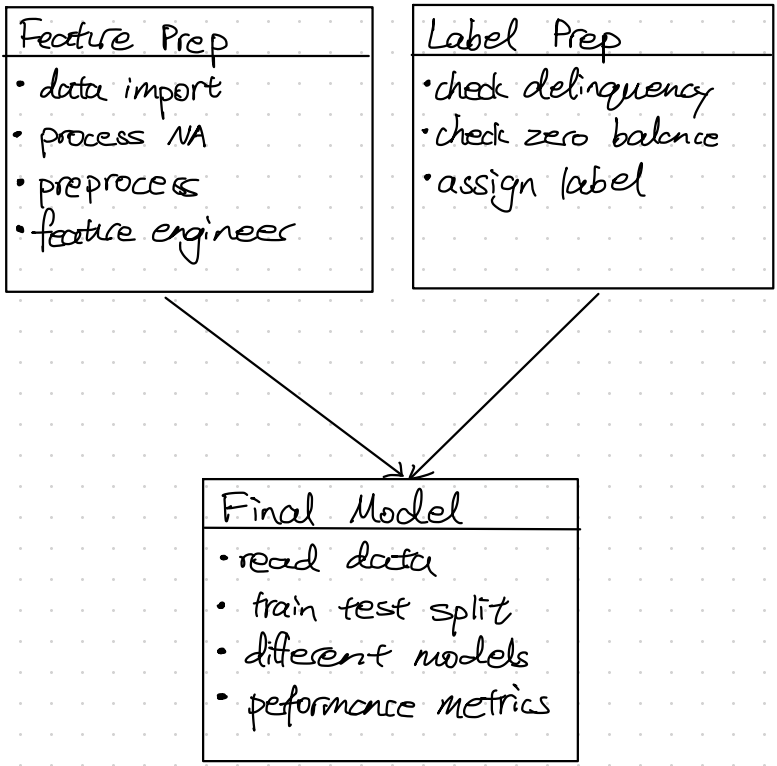
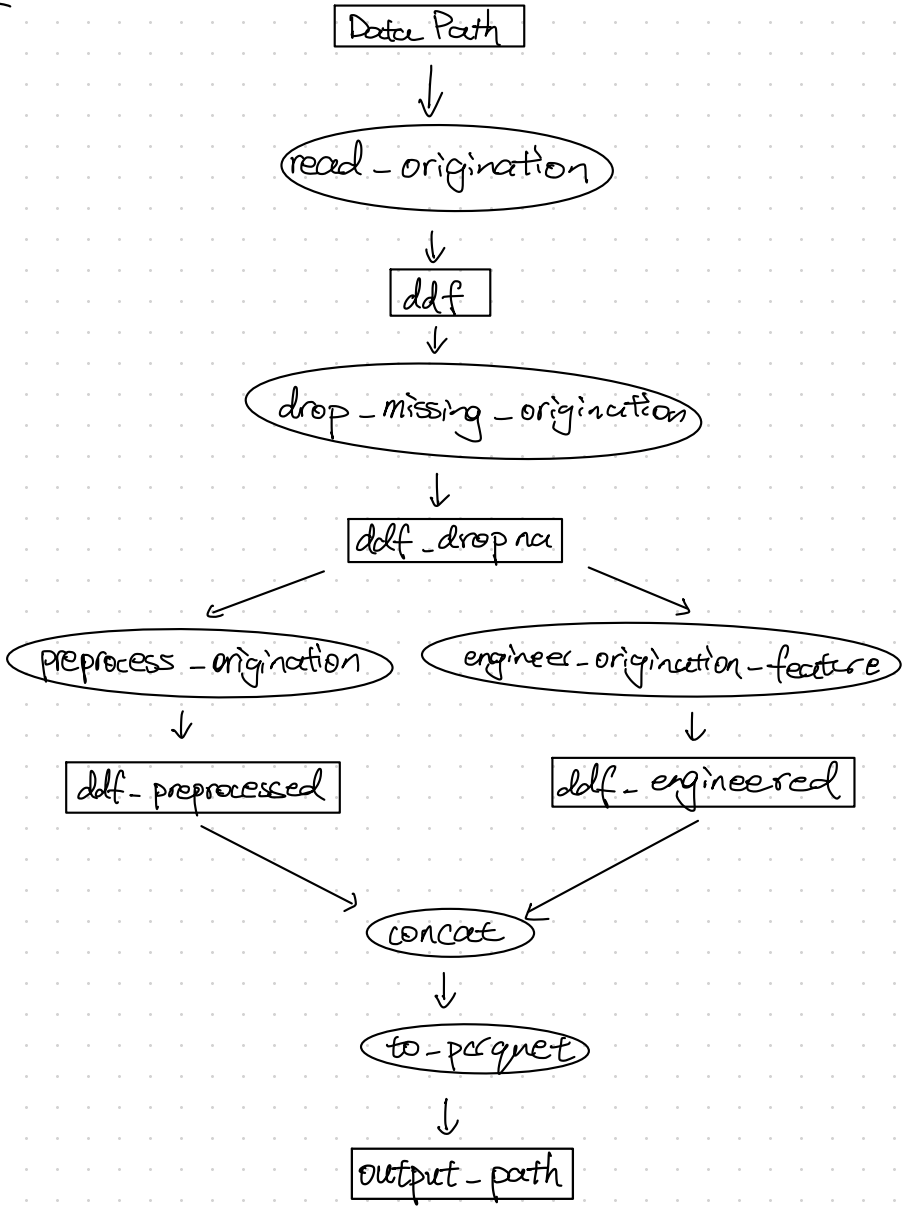
Data Flow Diagram



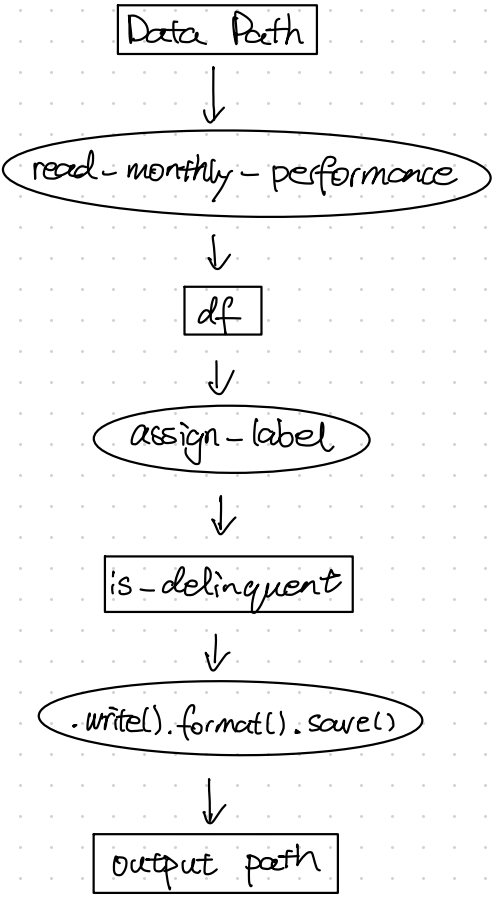
Feature Prep — in parallel w/ Label Prep

Dask

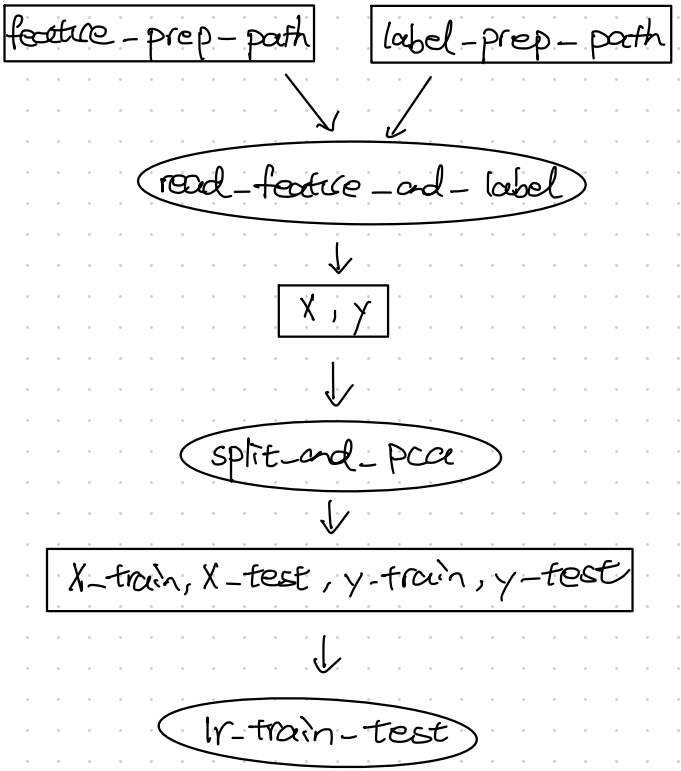


Label Prep — in parallel w/ Feature Prep

Spark



Final Model


feature_prep_path
label_prep_path
read_feature_and_label
X, y
split_and_pca
X_train, X_test, y_train, y_test
lr_train_test