

Dataiku

Plan and workflow

Tidy data and exploratory data analysis

Space delete
Null value
delete column
Imputer

Histogram (numerical)
Bar plot (non numerical)

Pair plot etc

Preprocessing and simple model score

Scaling – numerical value
Encoding – non numerical value
Preprocessing pipeline

Simple model

Get prediction
and evaluating
scores

Logistic regression model

Splitting learn data to
validating and training data

Logistic regression binary
classification model

Finding best model by cross
validation
Gridsearch / randomsearchg

Get prediction
and evaluating
scores

accuracy vs
precision vs
recall
tradeoff

Cloud, api and web page

Making docker file
Container
Or in vm

Running api
+ fatapi template or
Github page

Making presentaion