Hands-on classification model

Tengku Muhammad Hanis Bin Tengku Mokhtar, PhD

**October 23, 2024**

Contents

Tidymodels

- Tidymodels - a collection of packages for modeling and machine learning using a tidyverse principles
  - Tidyverse - an opinionated collection of R packages designed for data science
- Some of the core packages:
  - rsample - data splitting and resampling
  - parsnip - unify various ML models/algorithms from different packages
  - recipes - data pre-processing tools for feature engineering
  - tune - hyperparameter tuning
  - yardstick - assess the performance of ML models

- Other individual and small packages are available as well like nnet, kernlab, etc
- Alternatives to tidymodels:
  - caret - R only
  - mlr3 - R only
  - h2o - available in python and R

Basic workflow in tidymodels
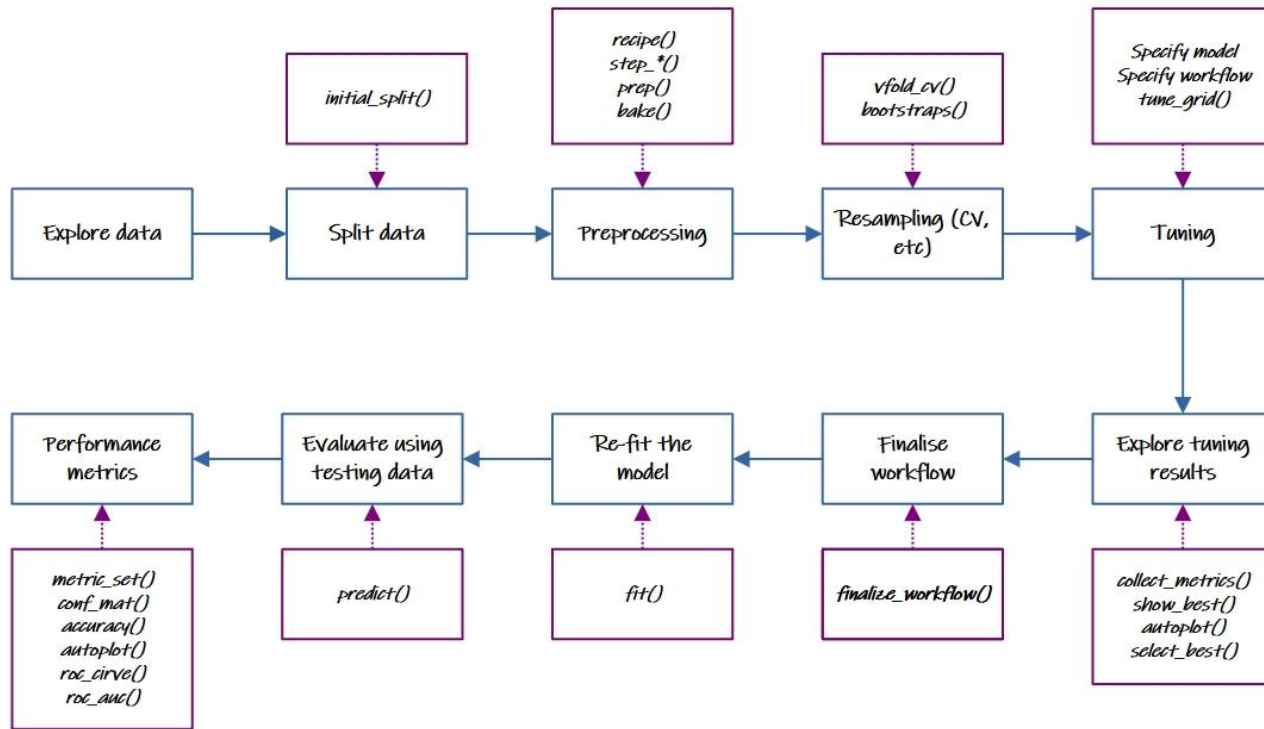
Data1 - Pima Indians Diabetes

- Aims - to predict whether or not a patient has diabetes or not
- Data source is [here](#)
- All patients are females at least 21 years old of Pima Indian heritage
- Variables:
    - Pregnant - number of times pregnant
    - Glucose - plasma glucose concentration (glucose tolerance test)
    - Pressure - diastolic blood pressure (mm Hg)
    - Triceps - triceps skin fold thickness (mm)
    - Insulin - 2-Hour serum insulin (mu U/ml)
    - Mass - Body mass index (weight in kg/(height in m)
    - Pedigree - Diabetes pedigree function
    - Age - Age (years)
    - Diabetes - Class variable (test for diabetes)

Data2 - Predict Students' Dropout and Academic Success

- Aims:
  - To predict whether students will dropout, enrolled, or graduated at the early stage of their studies
  - To support a predicted dropout student
- Data sources from [here](#)
- Some of the 37 variables:
  - Target: dropout, enrolled, graduated (the outcome)
  - Marital status
  - Admission grade
  - Nationality
  - Mother's occupation
  - Father's occupation
  - Etc

Hands-on in R

- Sign up for a free account at Posit Cloud
- Copy link from https://github.com/tengku-hanis/iku_ml

8

- Setting up Posit Cloud:
    - Log in to Posit Cloud
    - Click New Project → New Project from Git Repository → paste the url