## TidyModels

* **Packages:** 
  + Recipes: Do preprocessing of predictors or create new predictors
  + parsnip: 30 machine learning models to choose from and their respective engines
  + tune: Hyperparameter tuning for tidymodels packages
  + rsample: Random sampling. Resampling of data
  + workflows: Bundle pre-processing, modelling, post-processing
  + yardstick: Estimate how well models are working
* Initial\_Split(): Partition raw data into train and test. Train and test retain some proportions
* rsample: Apply resampling methods to training data
  + analysis(): extract resampled “training data”
  + assessment(): extract resampled “test data”
* Recipes: Do preprocessing of predictors or create new predictors
  + Define 2 roles: Outcome ~ Predictors
  + Update\_role: Specify what variables to exclude from the modelling but retained in the dataset (eg. for id)
  + Step\_: Data feature engineering from current data
  + Recipe selections: all\_of(), one\_of(). Apply steps to multiple variables at one time.
* Specify and train models with different engines using **Parnsnip Package**
  + Defines all the machine learning models with a simple interface. Used to add to workflow later on. Create model objects.
  + Set\_engine(): Define method for training model
  + Set\_mode(): Regression/Classification
* Workflow(): Pairs a model and recipe together. Bundle pre-processing, modelling, post-processing
  + Add\_model()
  + Add\_recipe()
  + Add\_formula(): If no recipe was defined
  + Fit: Train the model on training data
    - If you don’t use recipe, have to define model formula here
  + Pull\_workflow\_fit() %>% tidy(): Get model coefficients
  + Tidy(lm\_fit): Return description of fitted model parameters estimates. Nicer view than summary()
  + Predict(lm\_fit, new\_data = new\_points): Applies recipe to new data, passes them to fitted model.
  + Piping functions from yardstick package to measure performance of models
  + Fit\_resamples(): Use rsample object previously defined for resampling to fit into model
    - Collect\_metrics():
* Tune(): Model tuning with tidymodels uses the specification of the model to declare what parts of the model should be tuned.

Ensemble: Using multiple machine learning algorithms to obtain better predictive performance than a single algo

Resampling: Simulate how well model will perform on new data since test set should only be used as the final check for model’s performance. Create datasets from training data (split training data) for the purpose of calculating performance metrics without predicting the training set directly as a whole. Models from training are not kept. Cross-validation, Bootstrap.

* Train Set
  + Analysis Set
  + Assessment Set
  + Each set is called a fold
* Test Set

Diagram

Description automatically generated

## Process

Objective: Predict 1/0 buy sell signal from a bunch of datapoints. Supervised machine learning.

1. Getting Data (Price, Indicators, Alternative Data, Stock Index)
2. Feature Engineering. Column bind all the data into one big df
3. Predictive Modelling
   1. Define recipe: Datapoints
   2. Define model: Machine learning model and its parameters
   3. Define workflow: Add recipe and model into workflow
   4. Set cross validation
   5. Run model in parallel. Running multiple models, tuning something.
4. Making Predictions
   1. Choose best model
   2. Fit resamples?
5. Evaluate Model
   1. Visualising results. Compared to benchmark strategy
   2. Descriptive Stats
      1. Confusion Matrix, Log loss
      2. Finance Stats: Returns, SD, Sharpe, Drawdown, Sortino

## News

Reasons for surge in crypto

Cryptos available: Bitcoin, Ether, Ripple

Crossing all time high of 1 bitcoin = US$19,000

Square

Paypal: Allow users to buy and sell crypto

Robinhood

Easier to buy and transact with, taking down barrier of entry.

High profile money manager recommend it.

Blackrock: Bitcoin here to stay because of millenials’ interest in digital payments

Only 21m bitcoins will be produced. Scarcity, hence value it similar to gold.

Inflation hedge, weakening US dollar.

Elliott wave theory: Detects herding mentality and emotions driving the price

~800-900 bitcoins added daily

## Potential Data

bitcointreasuries.org

Github activity. API

News. NLP.