

# The Modelling Process

St. Joseph's University, Bengaluru

The Modelling Process

# Engineering features and selecting a model

Feature engineering is the process of selecting and transforming the input data to create meaningful and informative features that can improve the performance of machine learning models.

- ▶ Better representation of data
- ▶ Improved model accuracy
- ▶ Reduction of overfitting
- ▶ Handling missing data
- ▶ Interpretability

# Training the model

In this phase you present to your model data from which it can learn.

# Validating a model

Common error measures for machine learning are:

- ▶ classification error rate
- ▶ mean squared error

Strategies for model validation include:

- ▶ Dividing your data into a training set with  $X$  as a holdout data set.
- ▶ K-folds cross validation
- ▶ Leave-1 out

# Predicting new observations

The process of applying your model to new data is called model scoring. It involves two steps:

1. Prepare a data set that has features exactly as defined by your model.
2. Apply the model on this new data set, and this results in a prediction.