Ensemble Methods, Random Forests, and Boosting
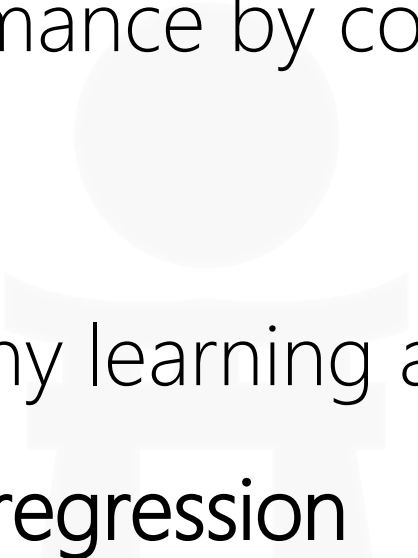
Bagging, Random Forest, Boosting, and AdaBoost
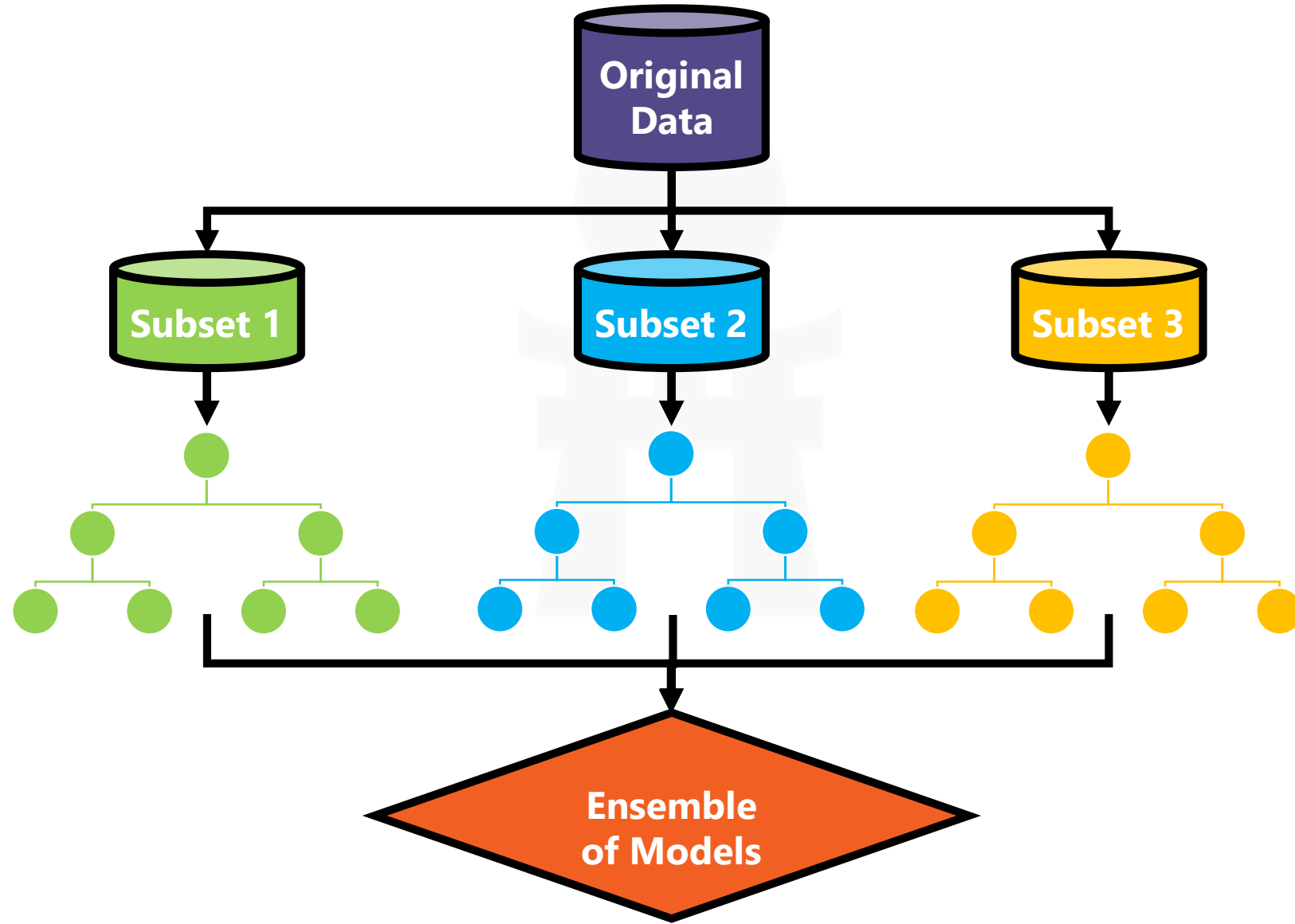
Agenda

- Overview and rationale

- Why ensemble?

  - Binomial Distribution

- Ensemble models

  - Bagging

  - Random Forests

  - Boosting

  - AdaBoost

Ensemble Methods

- Improve model performance by combining multiple models

- Ensembles can be of any learning algorithm, including both **classification** and **regression**

Ensemble of Decision Tree Models

Ensemble of Decision Tree Models

New Data

- A new data point (future data)
- Like a new Titanic passenger

Prediction 1

Prediction 2

Prediction 3

- Combine a bunch of independent models together
- Can be any type of algorithms

Vote

- Vote for classifiers
- Average for regression

Prediction of the Ensemble