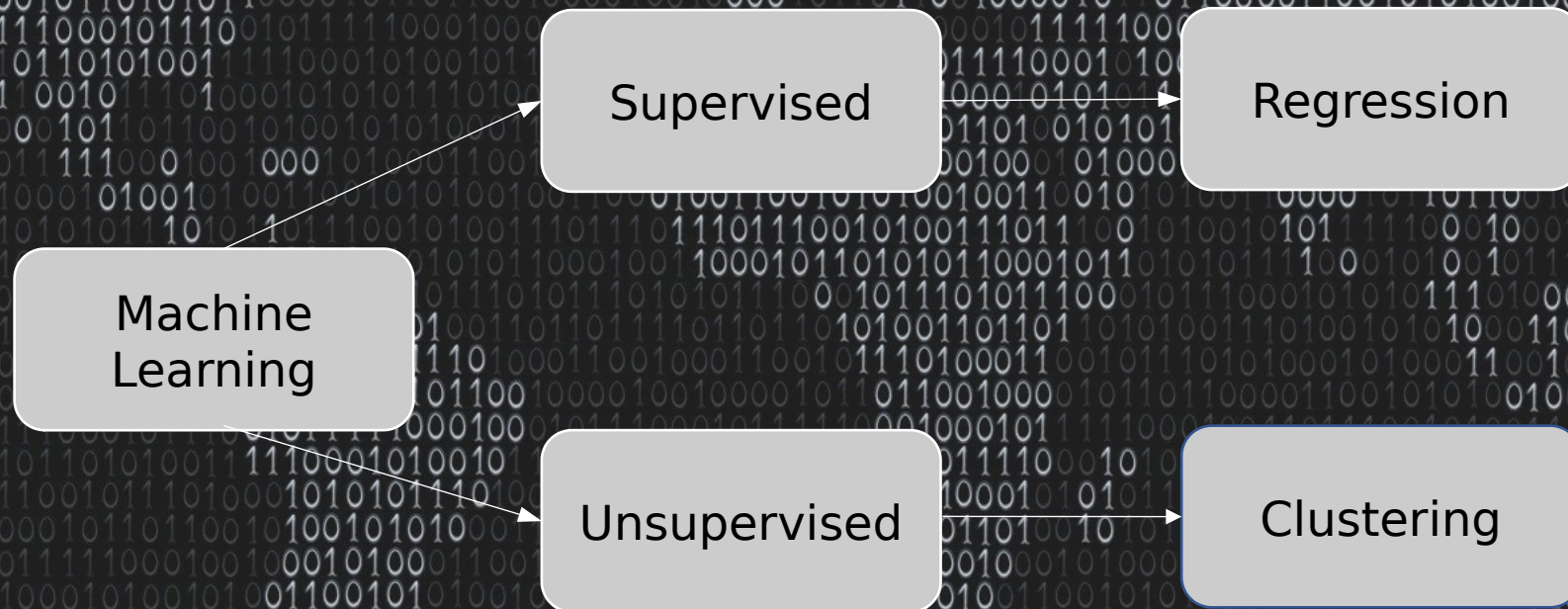


Machine Learning - Modeling

Models

- Different types of algorithms that facilitate implementing a Machine Learning based solution for a given problem
- Choosing a model depends on the use-case in question
- High level categorization of ML models:



Supervised Learning

- Learning with a teacher

- Learning a function that maps an input to an output based on example input-output pairs

Example:

Age	Height (cm)
5	119.38
7	132.08
11	154.94
18	180.34
8	??

Supervised
Learning Model

Predict the height
of 8 years old child

Age	Height (cm)
5	119.38
7	132.08
11	154.94
18	180.34
8	140.98

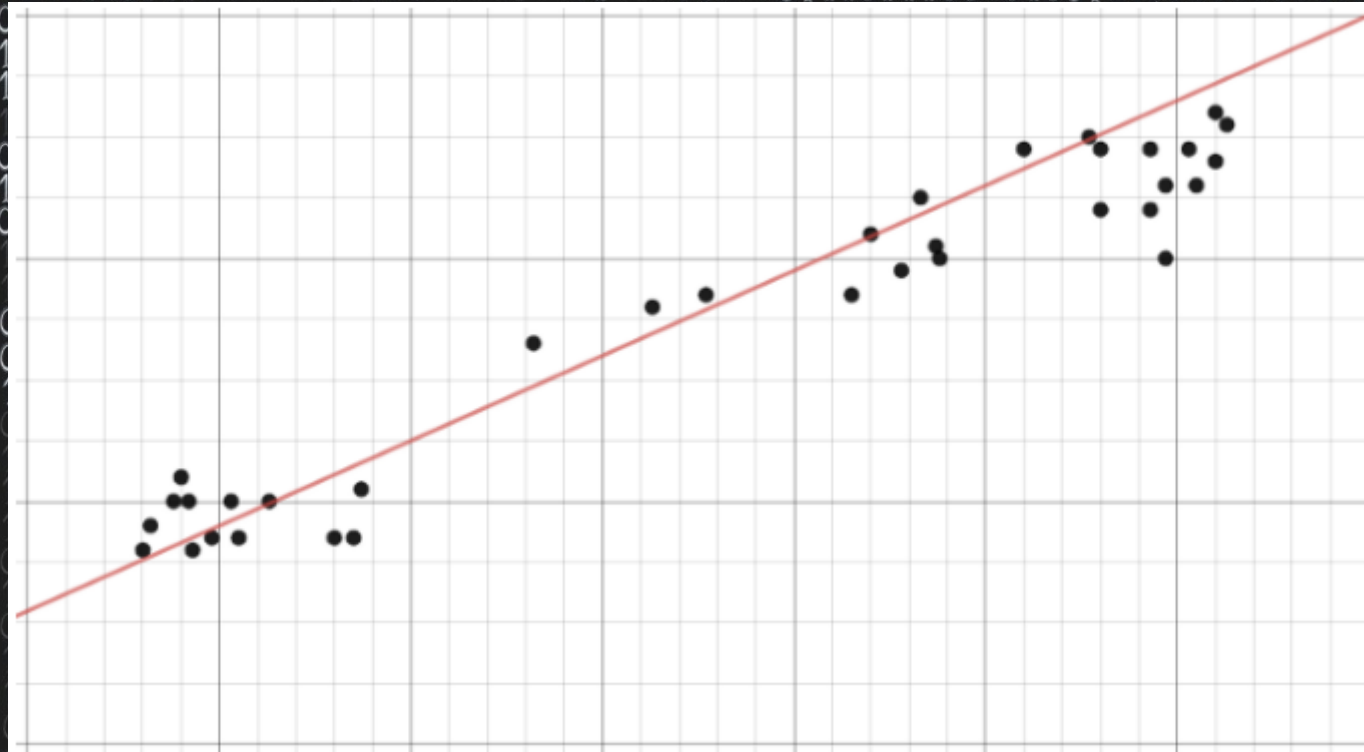
- There are two sub-categories under Supervised Learning – classification and regression

Regression

- The prediction or outcome is a continuous range
- Some of the most common regression models are:
 - Linear Regression
 - Decision Trees
 - Random Forest
 - Neural Network

Linear Regression

- Identify the geometrical line that best fits the distribution of observed values

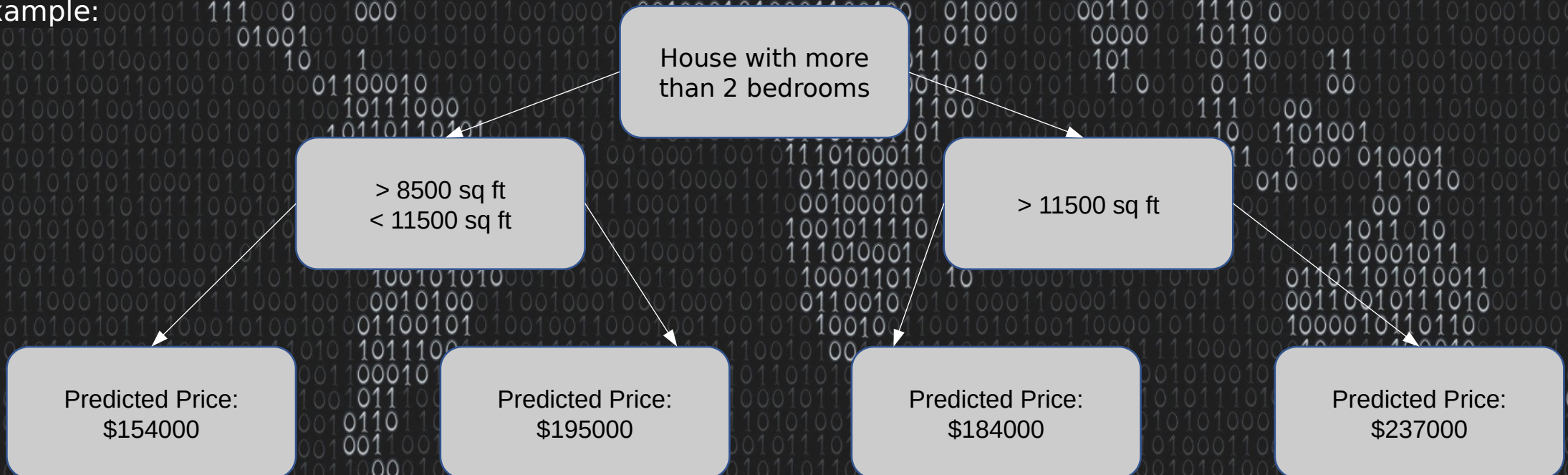


- Extensions of linear regression include – multiple linear regression, polynomial regression

Decision Tree

- It is one of the most popular models that is used in operations research, predictive analytics, strategic planning and machine learning
- It's a flowchart with blocks called nodes at each level posing a decision/choice to be made to move to the next level
- Generally, more the number of levels or depth of the decision tree, the more accurate is the model
- The last node of the decision tree is called a 'leaf' – which indicates the final outcome (category)

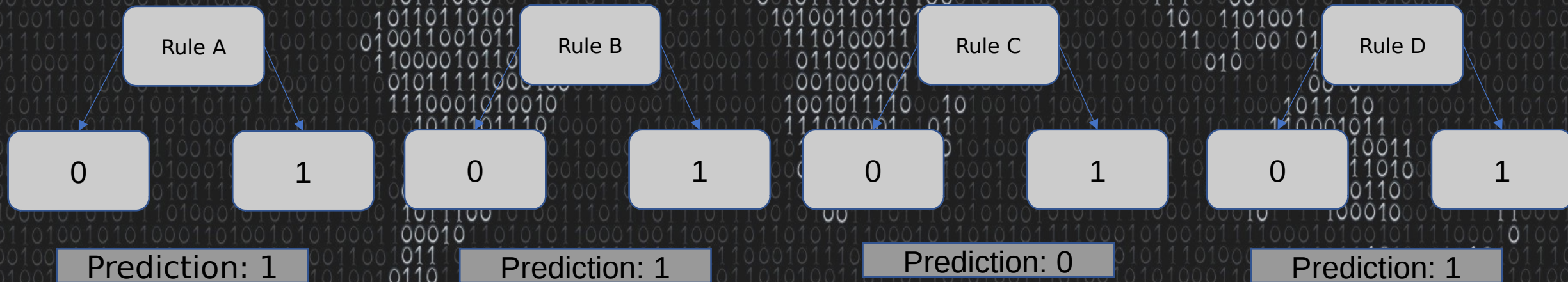
Example:



Random Forest

- Random forests are an ensemble learning technique based on decision trees
- At a high level, it's a collection of multiple decision trees generated using subsets of source datasets (bootstrapped datasets)
- The 'Mode' value of all the predictions from each individual decision tree gives the final prediction
- It relies on the 'majority wins' approach and eliminates the risk of error caused when prediction is made only from a single tree

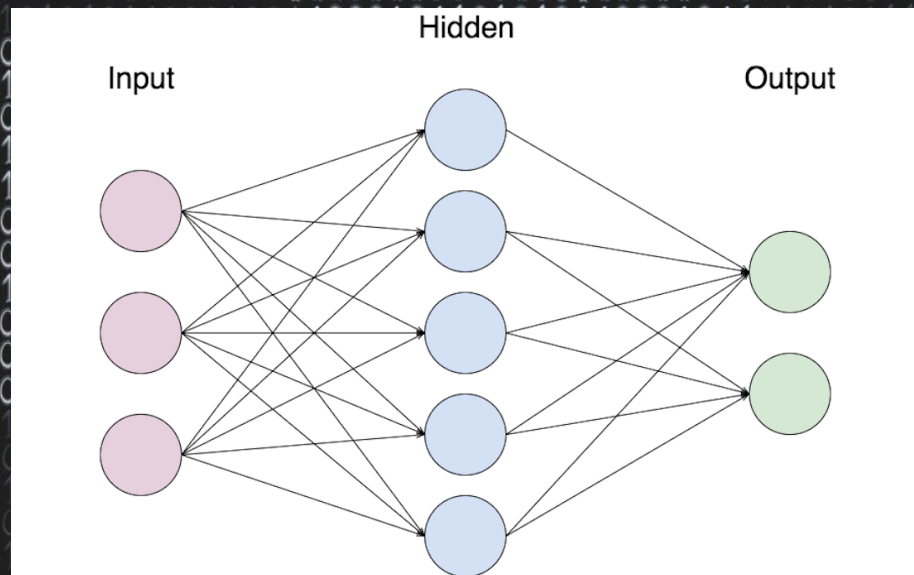
Example:



- Final Prediction = Mode Value = 1

Neural Network

- Neural Network or Artificial Neural Network is a collection of connected units or nodes called neurons (artificial neurons)
- There are three components to it – Input layer, Hidden layer and Output layer
- In simple terms, the hidden layer is a group of mathematical functions that can take one or many values from the input layer, process them and generate one or many layers of output layer
- The links connecting the layers can have a 'weight' which can be altered based on the observed behavior to improve on the output



Unsupervised Learning

- Learning without a teacher

- Unsupervised learning algorithm draws inferences and find patterns from input data without the help of any labelled outcomes

- One of the core methods that is used under unsupervised learning is 'Clustering'

Clustering

- An unsupervised technique where the data is grouped or clustered based on certain mathematical characteristics – vicinity in the two-dimensional plane etc.

- Primary used cases include – customer segmentation, fraud detection, document classification etc

- Examples of Clustering Algorithms – k-means clustering, hierarchical clustering, mean shift clustering etc

