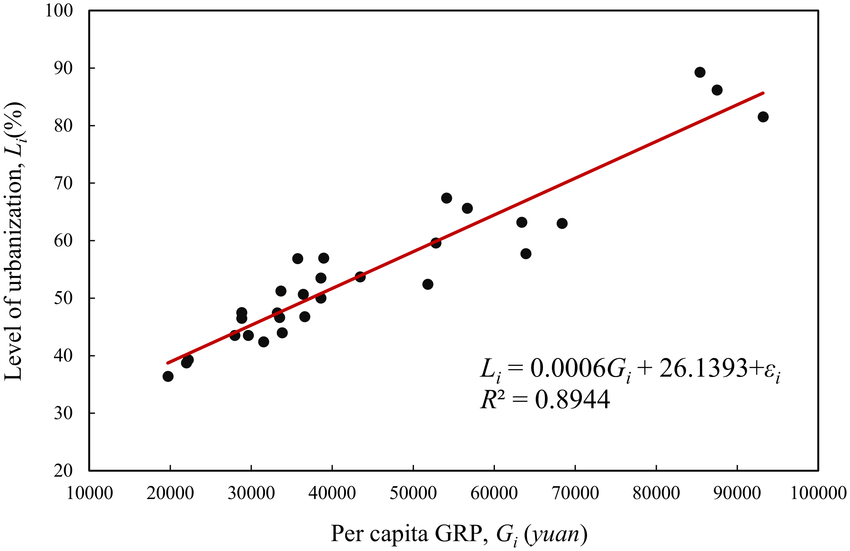
Machine learning algorithms

* **Linear regression**

Linear regression is a simple algorithm used for predicting a continuous value based on the relationship between independent and dependent variables. It assumes that the relationship between the input and output is linear, meaning a straight-line relationship. This method is ideal for forecasting numerical data.

**Example**: Predicting the price of a house based on features like size and location.

**Use Case**: Predicting sales based on advertising spend or forecasting the temperature based on past trends.

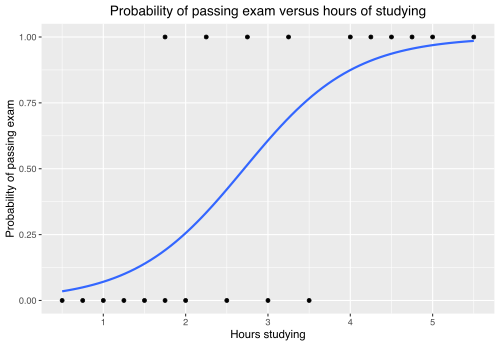


### **Logistic Regression**

Logistic regression is used for binary classification tasks, where the output is a probability between 0 and 1, representing one of two classes. It's used to classify data into categories like "yes/no" or "true/false." Despite its name, it is primarily used for classification, not regression

**Example**: Classifying emails as spam or not spam based on the content.

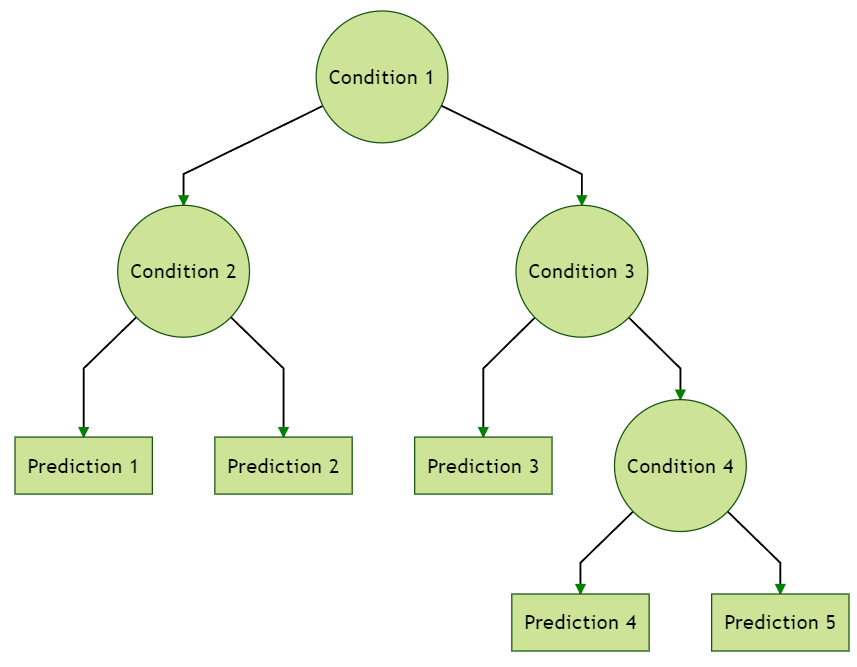
**Use Case**: Predicting whether a customer will churn, or diagnosing whether a patient has a specific disease based on medical data.



### **Decision Trees**

A decision tree is a flowchart-like model used for classification and regression tasks. The tree splits the data based on feature values, making decisions at each node until it reaches a conclusion (leaf node). Decision trees are easy to interpret and visualize.

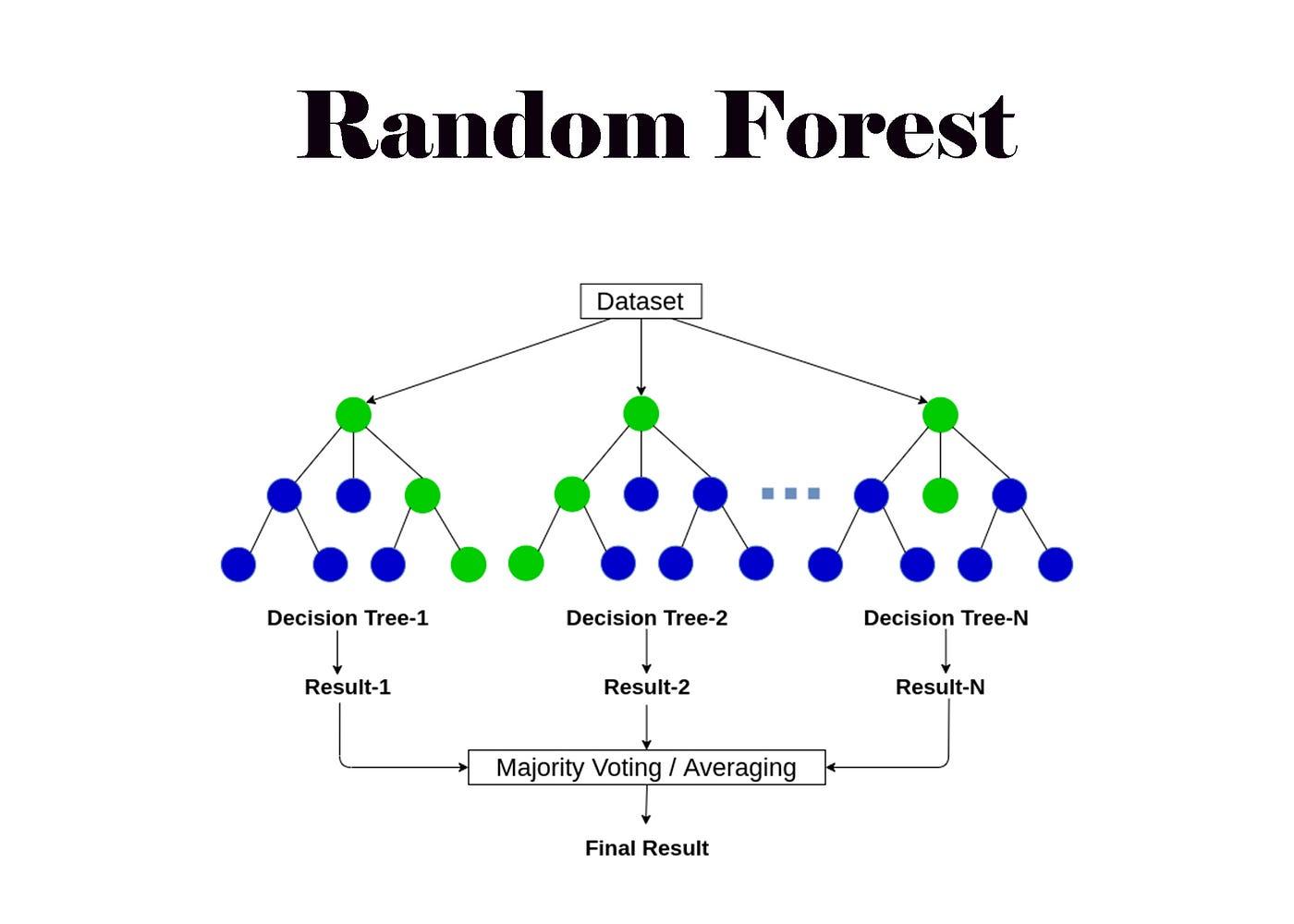
**Example**: Deciding whether to approve a loan based on factors like income, age, and credit score.  
**Use Case**: Predicting customer churn, classifying animals by features, or diagnosing medical conditions based on patient data.



### **Random Forest**

Random Forest is an ensemble method that uses multiple decision trees to improve prediction accuracy. Each tree is trained on a random subset of the data and makes independent predictions, which are aggregated (voted for classification or averaged for regression). This helps reduce the risk of overfitting.  
 **Example**: Predicting customer satisfaction based on behavior and demographics.

**Use Case**: Fraud detection, image classification, or recommendation systems (like predicting which movie a user will like).

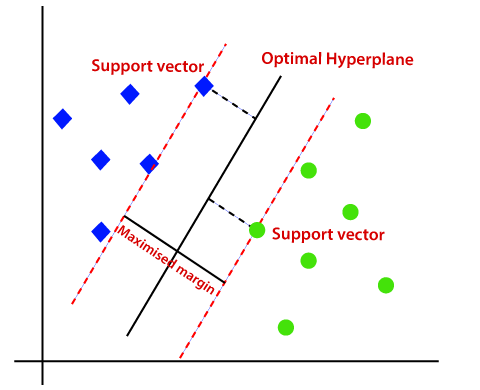


* **Support Vector Machines (SVM)**

Support Vector Machines are classification algorithms that try to find the best boundary (hyperplane) that separates data into different classes. SVM works well even when data is not linearly separable by using a kernel trick to map data into higher-dimensional space.

**Example**: Classifying images of cats and dogs using pixel values.

**Use Case**: Handwritten digit recognition, classifying cancer cells, or detecting fraud in credit card transactions.

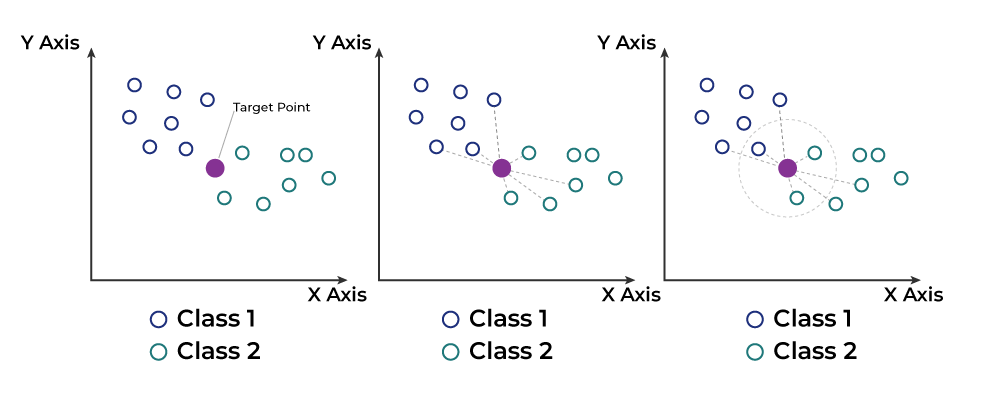


* **K-Nearest Neighbors (KNN)**

K-Nearest Neighbors is a simple algorithm that classifies new data based on the majority class of its nearest neighbors. It doesn’t require training, and its performance depends on the number of neighbors (k) chosen.

**Example:** Classifying a new flower as a specific species based on its features (like petal length and width).

**Use Case:** Recommending products based on user behavior, classifying handwritten digits, or predicting house prices based on nearby properties.

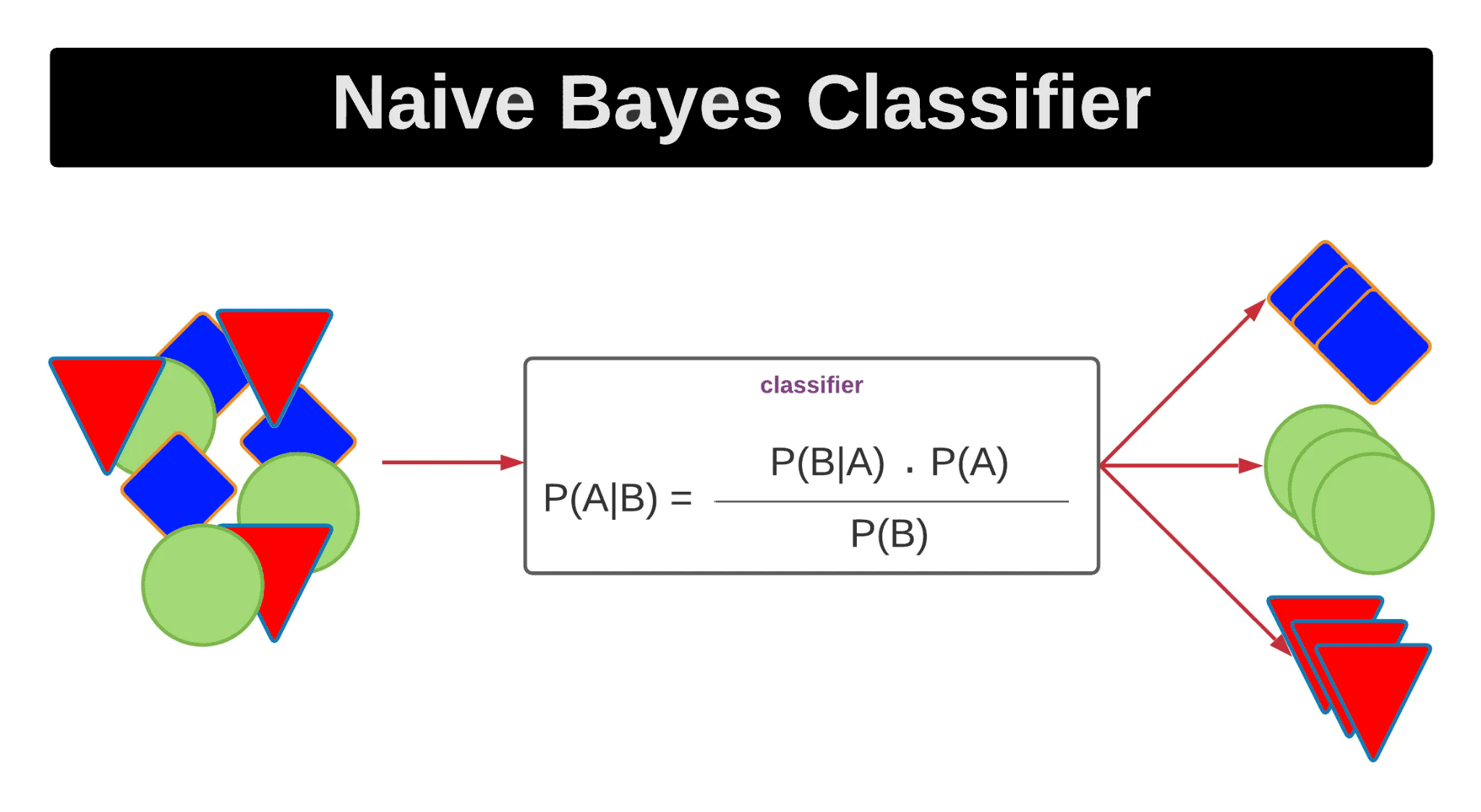


* **Naive Bayes**

Naive Bayes is a probabilistic classifier based on Bayes' Theorem. It assumes that the features are independent of each other, which makes it "naive." Despite the simplification, it performs well in many real-world tasks, especially when the independence assumption roughly holds.

**Example:** Classifying emails as spam or not spam based on word frequency.

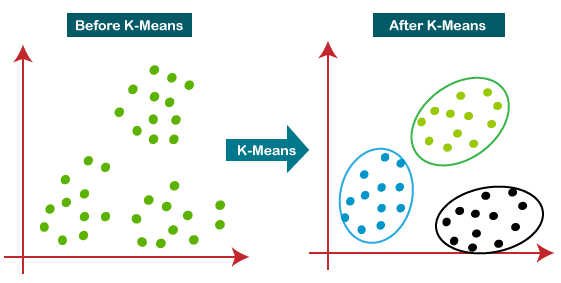
**Use Case:** Sentiment analysis, text classification, or medical diagnosis (e.g., predicting if a patient will develop a certain condition).

****

* **K-Means Clustering**

K-Means is an unsupervised learning algorithm used to group data into clusters based on similarity. It assigns each data point to the cluster whose center (centroid) is closest. K-Means is useful when you want to find natural groupings in the data.

**Example:** Grouping customers into segments based on purchasing behavior (e.g., high-spending vs. low-spending customers).

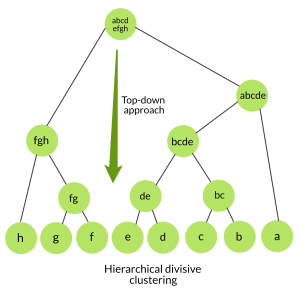
**Use Case:** Customer segmentation, grouping news articles by topic, or organizing images by similarity.****

### **Hierarchical Clustering**

Hierarchical clustering builds a tree-like structure (dendrogram) that shows how data points are grouped at different levels of similarity. It’s useful when you want to understand the relationships between clusters in a hierarchy.

**Example:** Grouping species based on genetic similarity or clustering similar articles in a dataset.

**Use Case:** Gene expression analysis, customer segmentation for targeted marketing, or organizing documents into themes.

****