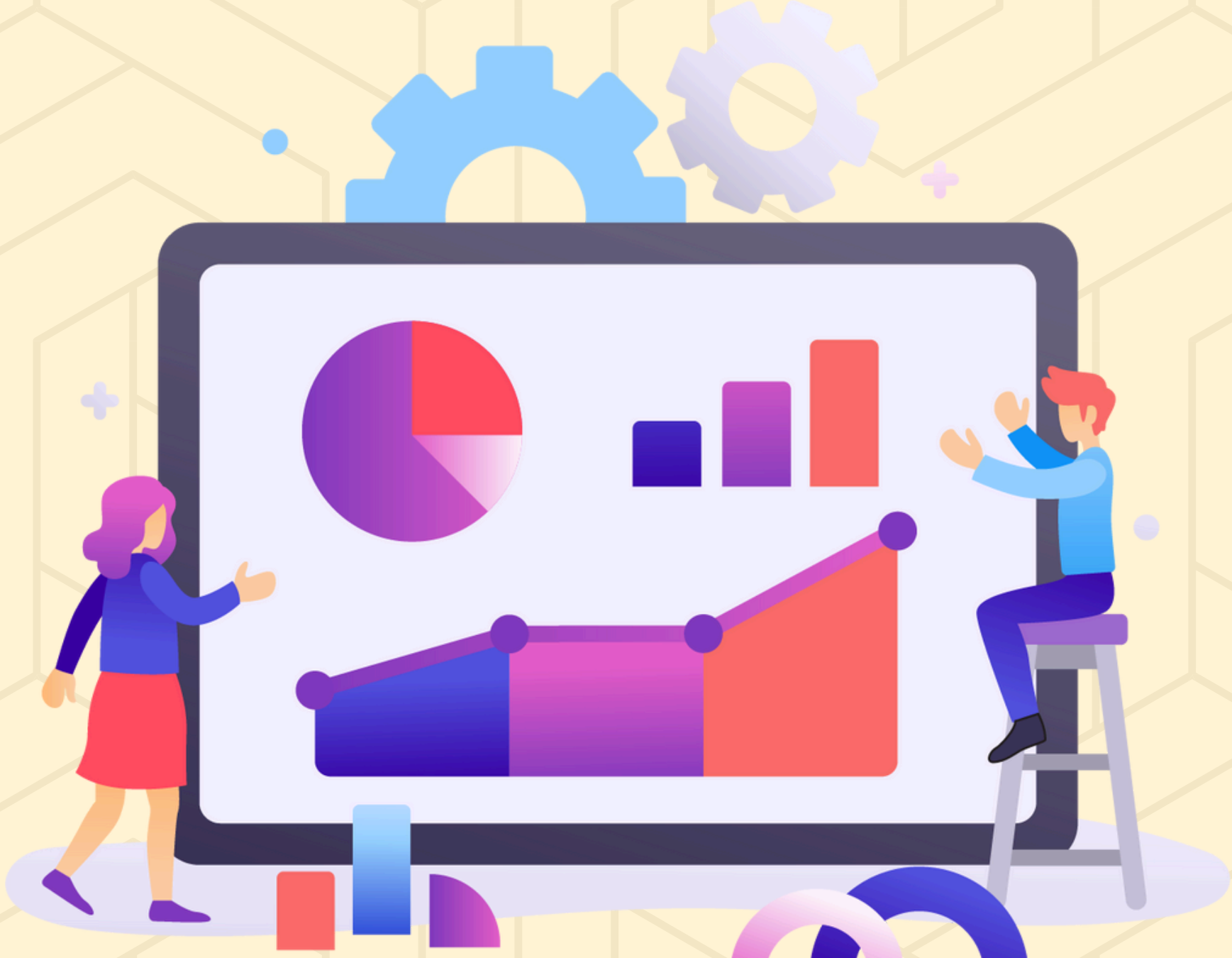
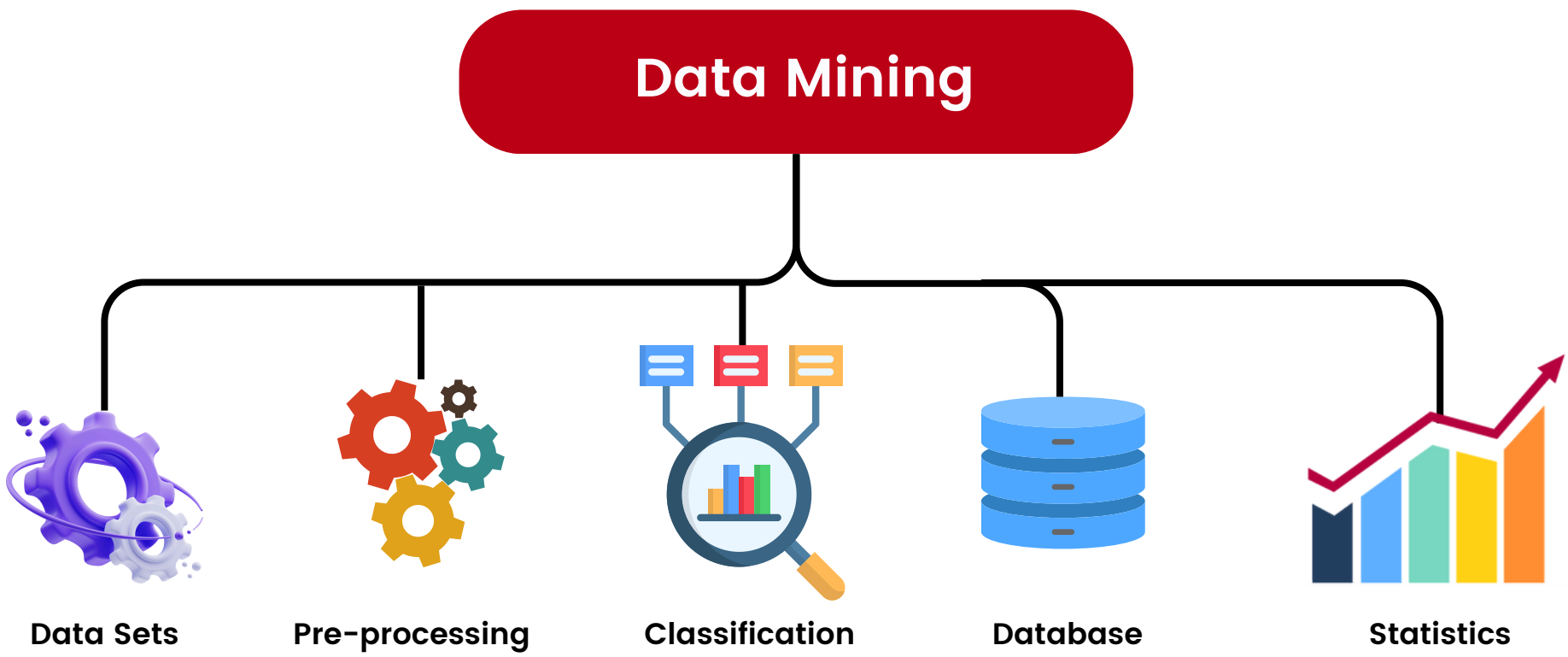
TOP DATA SCIENCE TOPICS YOU NEED TO KNOW



THE RAVIT SHOW

Don't Forget to Save For Later

Data Mining Essentials



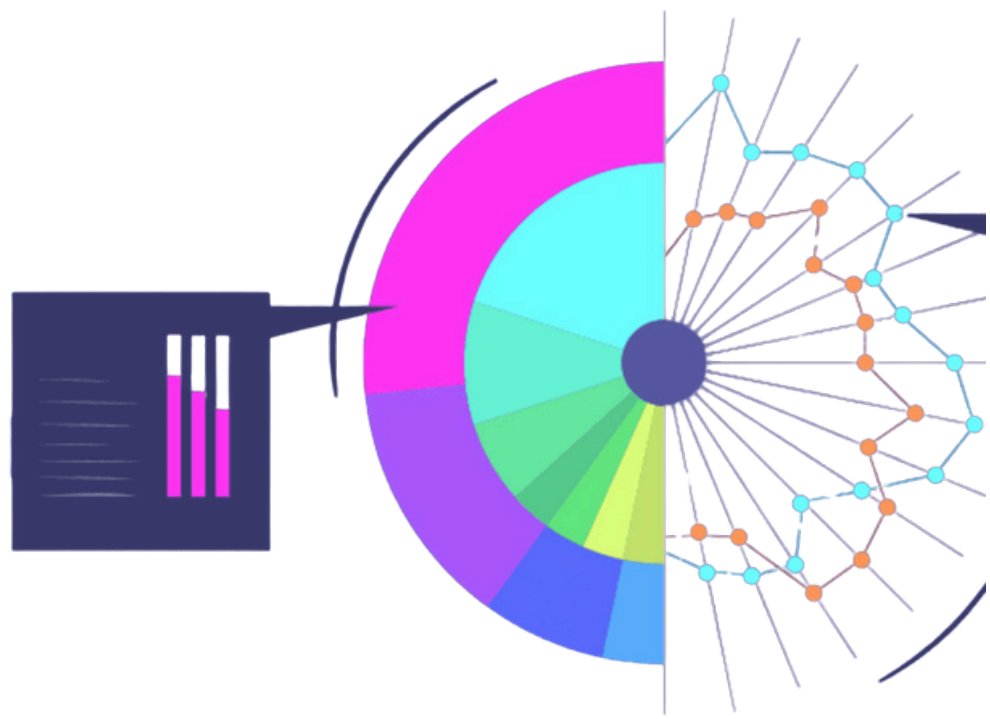
- **What is it?** An iterative process for discovering patterns in large datasets.
- **Objectives:** Identify patterns, establish trends, and solve problems.
- **Stages:** Problem definition, data exploration, preparation, modeling, evaluation, and deployment.
- **Core Terms:** Classification, prediction, association rules, data reduction, exploration, supervised/unsupervised learning, sampling, and model building.

Swipe →

THE RAVIT SHOW

Don't Forget to Save For Later

Data Visualization



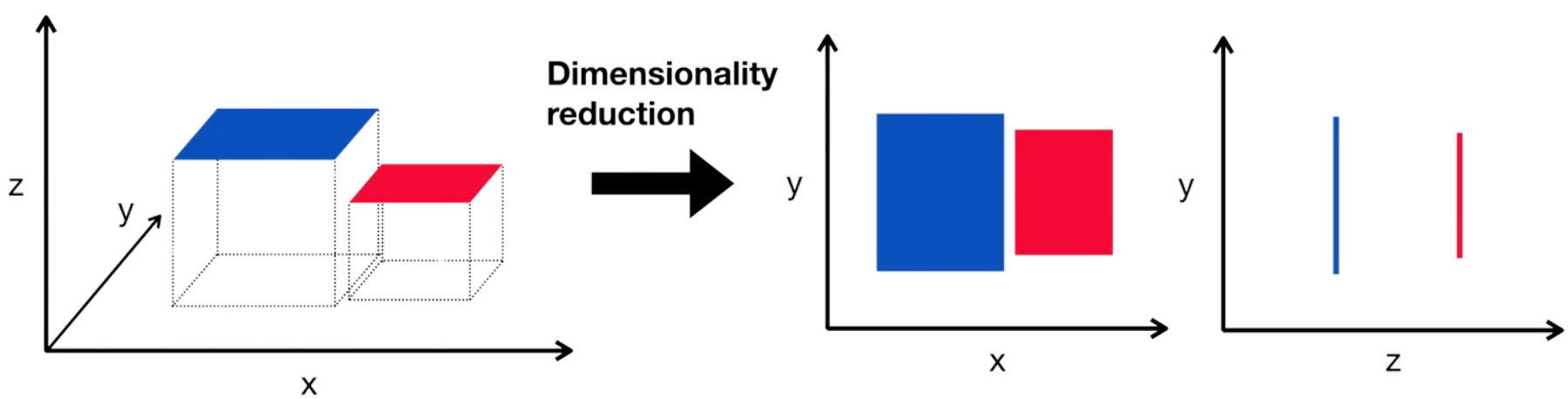
- **What is it?** Presenting data in graphical formats to reveal patterns and trends.
- **Basic Graphs:** Line graphs, bar graphs, scatter plots, histograms, box plots, heatmaps.
- **Advanced Techniques:** Multidimensional variables using color, size, shape, and animations.
- **Manipulation:** Be skilled in rescaling, zooming, filtering, and aggregating data.
- **Specialized Visualizations:** Master map charts and tree maps for advanced insights.

Swipe →

THE RAVIT SHOW

Don't Forget to Save For Later

Dimension Reduction Techniques

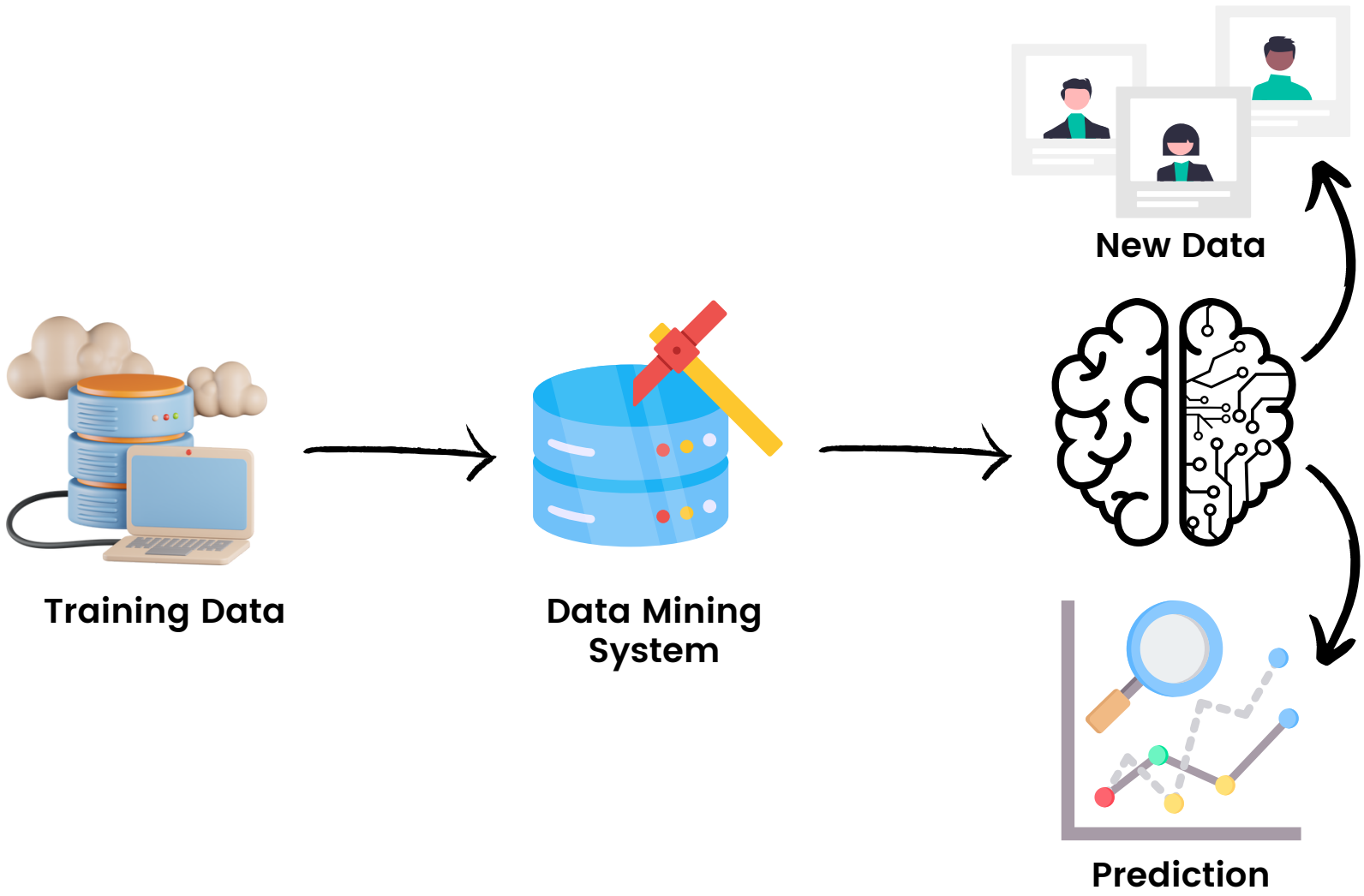


- **What is it?** Reducing dataset dimensions while preserving information.
- **Purpose:** Decrease the number of random variables in data.
- **Popular Dimension Reduction Methods:** Missing Values, Low Variance, Decision Trees, Random Forest, High Correlation, Factor Analysis, Principal Component Analysis, Backward Feature Elimination.

Swipe →

Don't Forget to
Save For Later

Classification



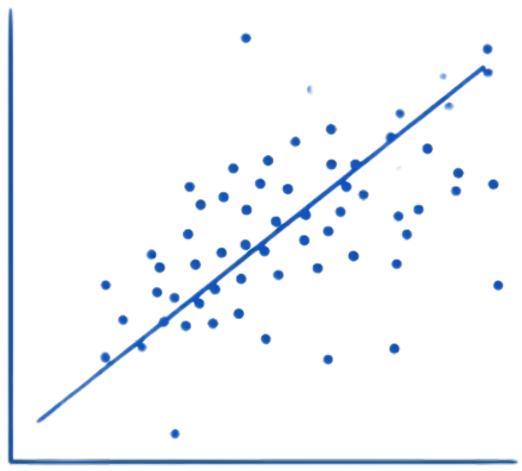
- **What is it?** Presenting data in graphical formats to reveal patterns and trends.
- **Basic Graphs:** Line, bar, scatter plots, histograms, box plots, heatmaps.
- **Advanced Techniques:** Multidimensional variables using color, size, shape, and animations.
- **Manipulation:** Rescaling, zooming, filtering, aggregating data.

Swipe

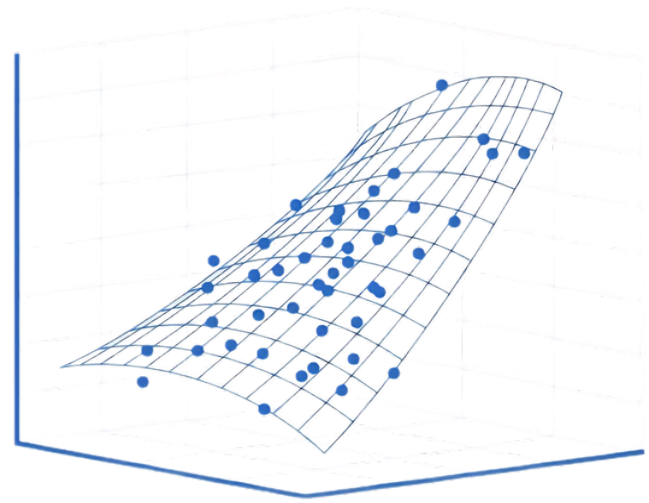
Don't Forget to
Save For Later

Simple and Multiple Linear Regression

Simple Linear Regression



Multiple Linear Regression

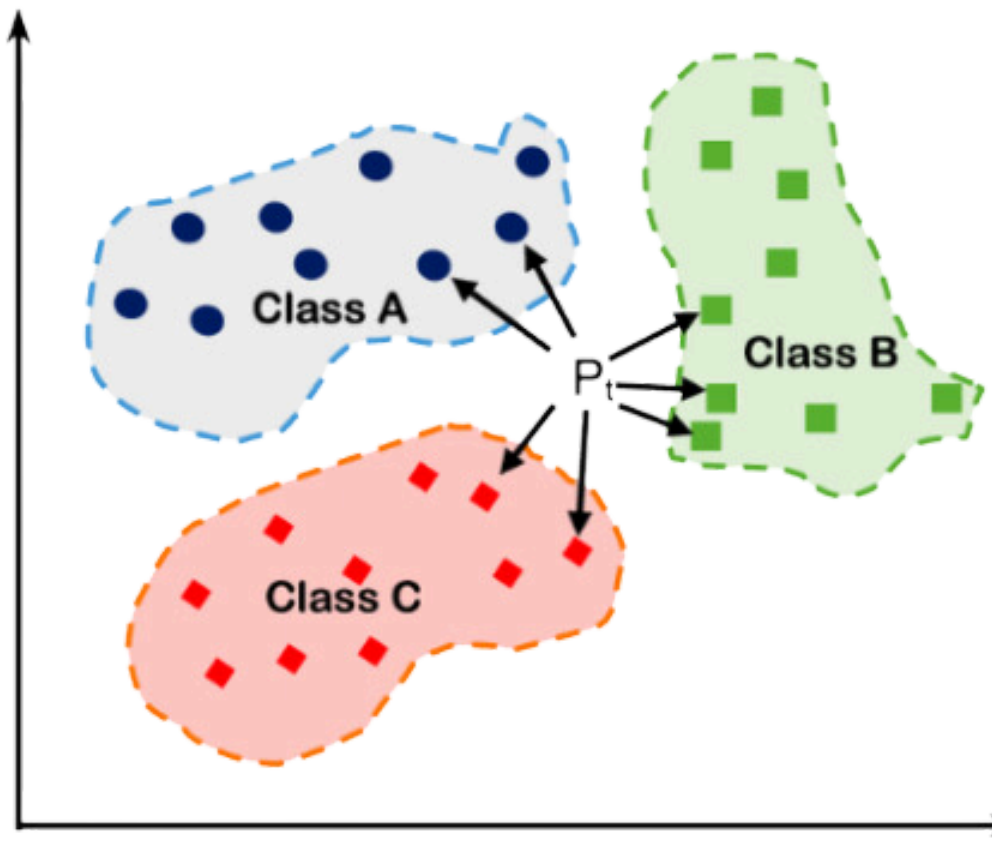


- **What is it?** Basic statistical models for studying relationships between variables.
- **Purpose:** Predicting and forecasting Y values based on X values.
- **Types:** Simple and multiple linear regression.
- **Key Points:** Correlation coefficient, regression line, residual plot, linear regression equation.

Swipe

Don't Forget to
Save For Later

Naive Bayes

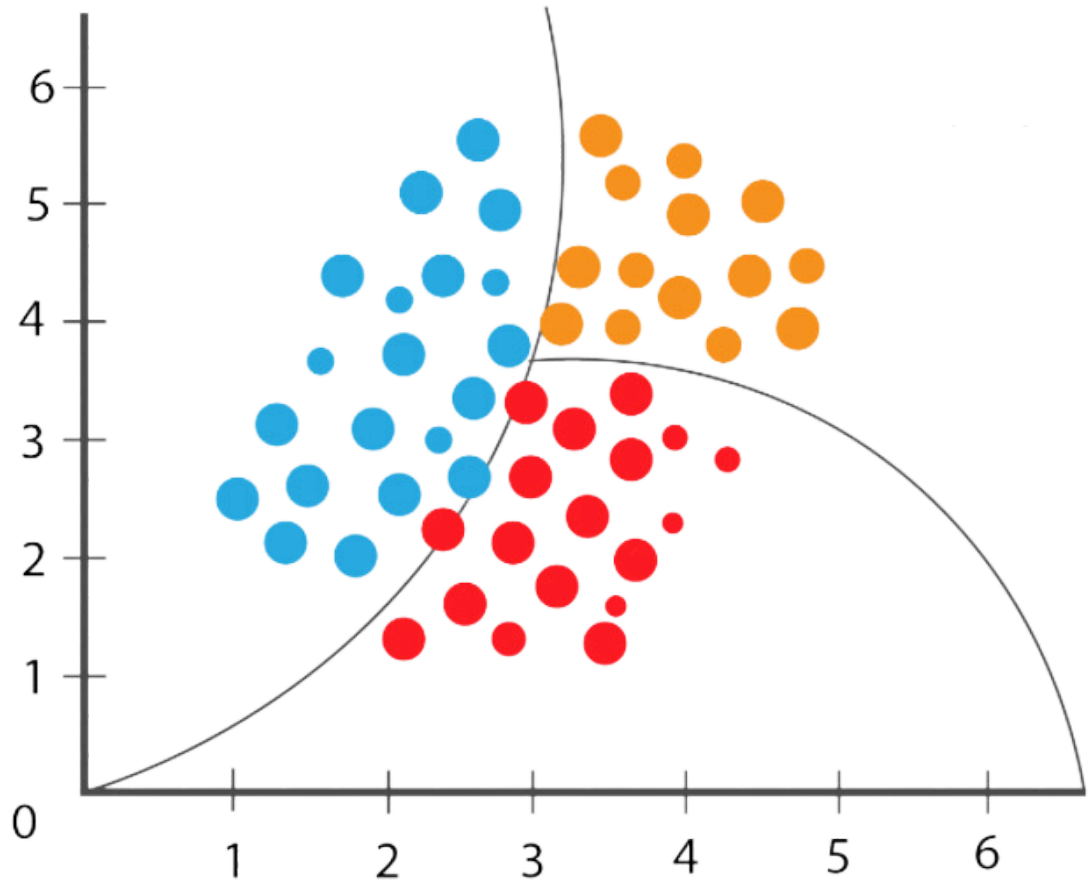


- **What is it?** Data classification algorithm based on proximity to group members.
- **Purpose:** Regression and classification without explicit parameter assumptions.
- **Key Skills:** Determining neighbors, classification rules, choosing k.
- **Applications:** Text mining, anomaly detection.

Swipe

Don't Forget to
Save For Later

K-Nearest Neighbor (K-NN)

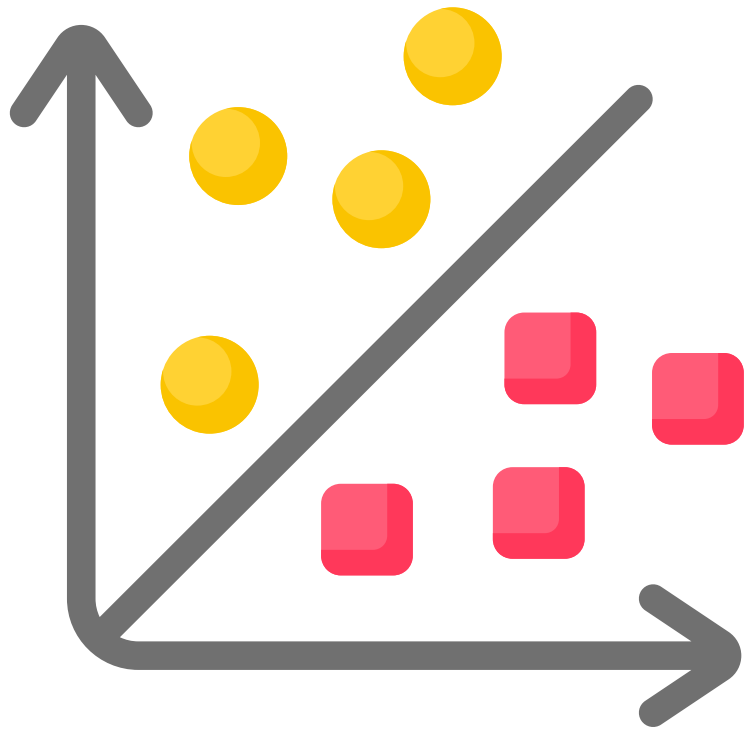


- **What is it?** Collection of classification algorithms based on Bayes Theorem.
- **Applications:** Spam detection, document classification.
- **Variations:** Multinomial Naive Bayes, Bernoulli Naive Bayes, Binarized Multinomial Naive Bayes.

Swipe

Don't Forget to Save For Later

Classification and Regression Trees (CART)



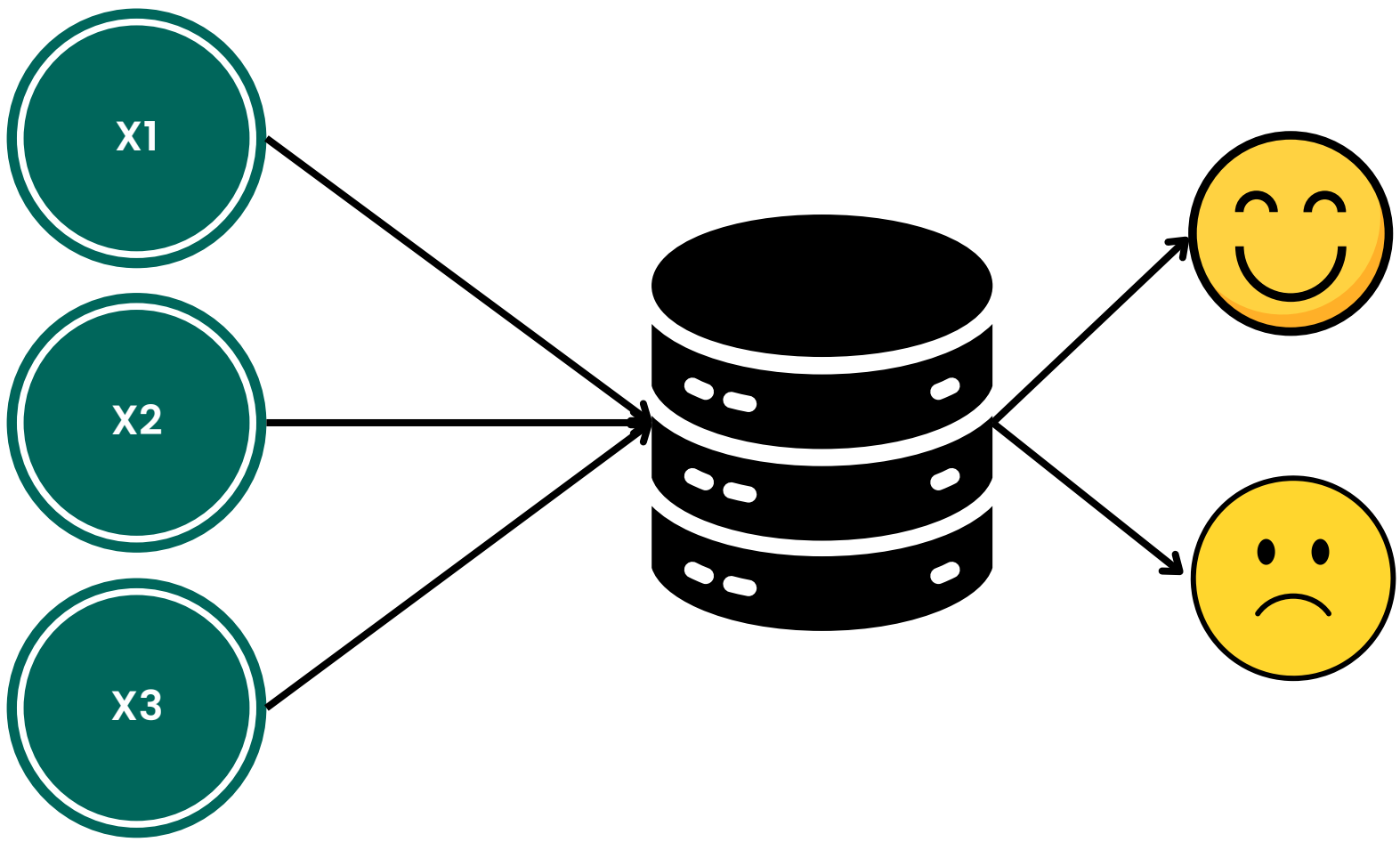
- **What is it?** Popular predictive modeling approach building classification or regression models in tree form.
- **Applications:** Data mining, statistics, machine learning.
- **Key Terms:** CART methodology, classification trees, regression trees, interactive dichotomizer, C4.5, C5.5, decision stump, conditional decision tree, M5.

Swipe

THE RAVIT SHOW

Don't Forget to Save For Later

Logistic Regression



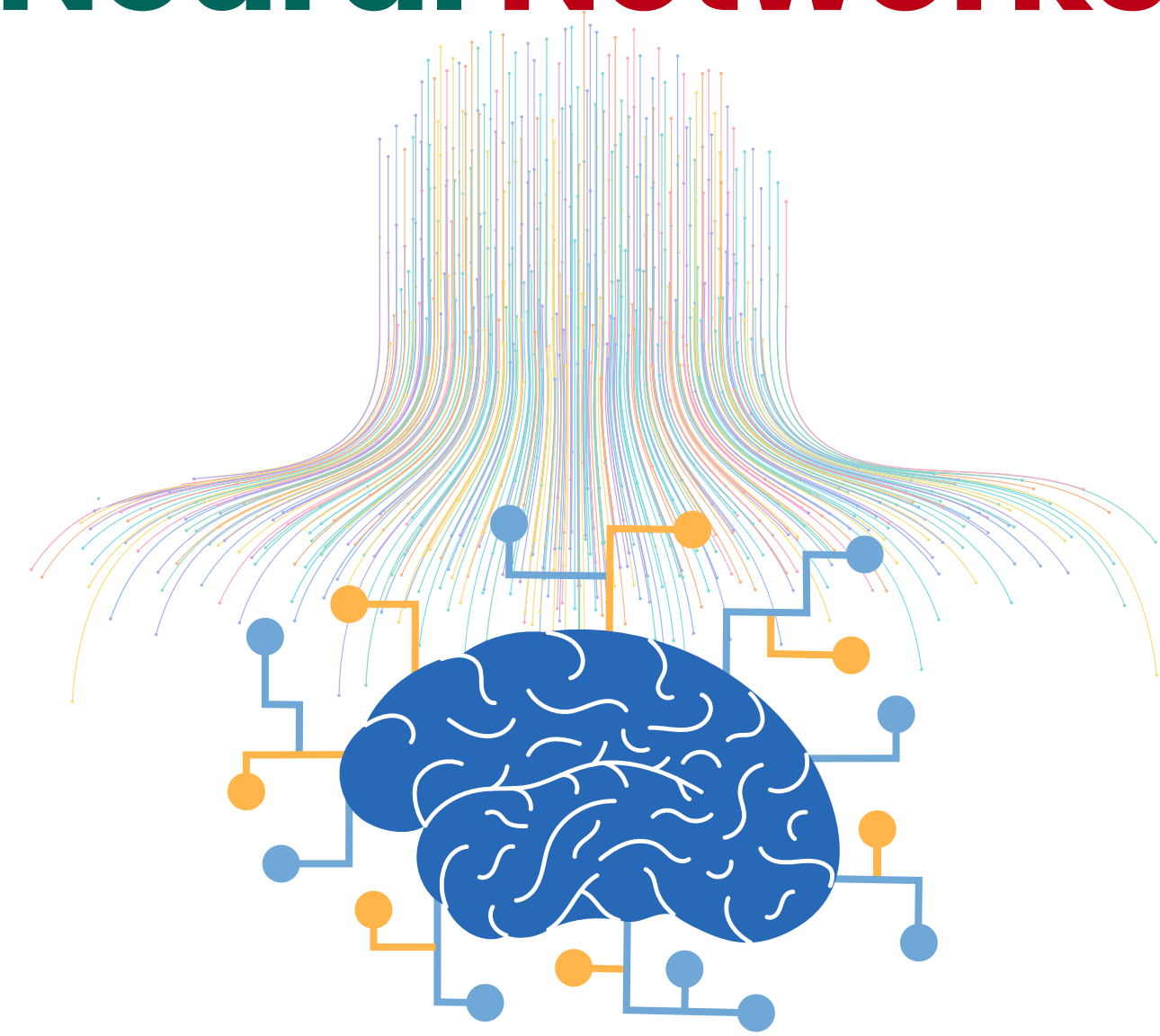
- **What is it?** Analyzes relationship between dependent and independent variables with a binary outcome.
- **Applications:** Used for binary classification tasks.
- **Key Terms:** Sigmoid function, S-shaped curve, multiple logistic regression, categorical and continuous predictors.

Swipe

THE RAVIT SHOW

Don't Forget to Save For Later

Neural Networks



- **What is it?** Mimic human brain neurons to learn data patterns for classification, regression, prediction, etc.
- **Applications:** Deep learning for signal processing, pattern recognition.
- **Key Terms:** Concept and structure of Neural Networks, perceptron, Back-propagation, Hopfield Network.

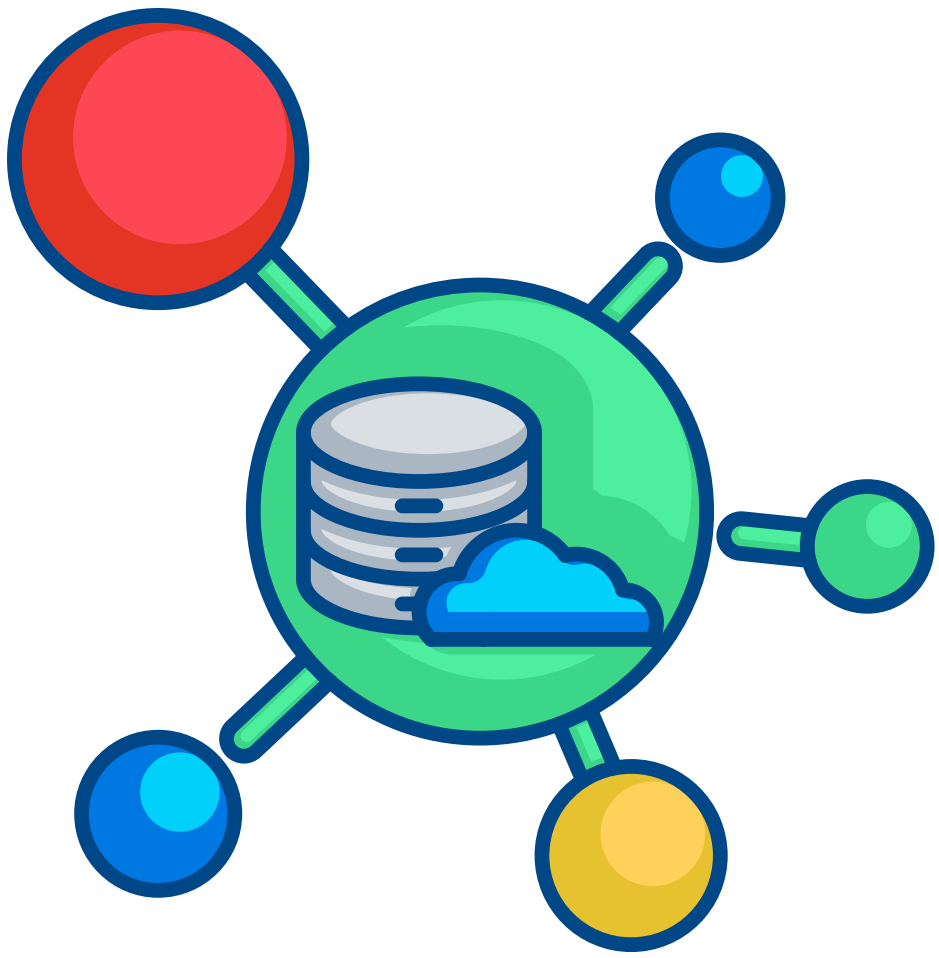
Swipe

THE RAVIT SHOW

Don't Forget to Save For Later

Advanced Topics

- Data engineering – Hadoop, MapReduce, Pregel
- Regression-based forecasting
- Time stamps and financial modeling
- Discriminant analysis
- Association rules
- Cluster analysis
- Smoothing methods
- Fraud detection
- GIS and spatial data
- Time series



Swipe

Don't Forget to
Save For Later

Was it useful?

Let me know in the comments

THE RAVIT SHOW

@theravitshow