

**DATA Is The New  
Oil**

# Data Mining

It is process of analysing data from different perspective  
summarizing it into useful information  
information that can be used to increase revenue cut cause or both data  
mining helps analyse recognise different information fact relationship  
trends patterns exceptions anomalies that might otherwise School

## **Major data mining tasks are :**

classification,

predicting an item class

class string descriptive finding groups and items deviations

detection predictive finding change for casting predicting a

parameter value description describing a group link analysis

finding relationships and associations

# **Major industries using data mining**

Retail

Finance

Education

Healthcare

Agriculture

Manufacturing

Transportation

Aerospace

**Since you're diving into data mining, Orange is a fantastic choice because it uses a "visual programming" style—no heavy coding required! It's like building with LEGO blocks for data.**

# Orange

Orange is a component based to visual programming software for data mining machine learning and analysis supports communication between data scientist and domain experts

Open source component based on programming data visualisation platform independent software allows plastering and classification data mining progressible programming and python scripting

## **What is Orange?**

**Definition:** An open-source toolkit for data visualization, machine learning, and data mining.

**The "Visual" Edge:** Instead of writing lines of code, you create "Workflows" by dragging and dropping components.

**Platform:** It's Python-based but features a user-friendly Graphical User Interface (GUI).

**Who uses it?** Students, researchers, and data scientists who want to prototype models quickly.

# Key Concepts: Widgets & Canvas

**The Canvas:** This is your "drawing board" where you build your data analysis process.

**Widgets:** These are the individual "tools" or blocks. Each widget performs a specific task:

**Data Reading:** Loading your files (Excel, CSV).

**Preprocessing:** Cleaning data or handling missing values.

**Visualization:** Creating charts and plots.

**Modeling:** Training Machine Learning algorithms.

**Communication:** Widgets "talk" to each other through lines (communication channels) that carry data from one step to the next.

# The 6 Core Widget Categories

Orange organizes its tools into functional groups:

Data: File, Data Table, Select Columns.

Visualize: Scatter Plot, Box Plot, Distributions, Heatmaps.

Model: Decision Trees, Random Forest, k-NN, Neural Networks.

Evaluate: Test and Score, Confusion Matrix, ROC Analysis.

Transform: Discretization, Normalization, Feature Selection.

Unsupervised: Hierarchical Clustering, k-Means, PCA (Principal Component Analysis).

# Advantages of using Orange

**Interactive Exploration:** If you select a data point in a scatter plot, it instantly updates in your linked data table.

**Educational Tool:** Perfect for learning how algorithms work without getting stuck on syntax errors.

**Add-ons:** You can install "Add-ons" for specialized tasks like:

**Text Mining:** Analyzing Tweets or documents.

**Image Analytics:** Recognizing objects in pictures.

**Bioinformatics:** Analyzing gene expressions.

**Python Integration:** Advanced users can write their own Python scripts directly inside Orange.

# Step-by-Step Workflow Example

**To build a simple predictive model, you follow these steps:**

**Load Data:** Drag the File widget and select your dataset.

**View Data:** Connect it to a Data Table to see the rows and columns.







**Visualize:** Connect the File widget to a Scatter Plot to find patterns.

**Train Model:** Connect the data to a Tree widget to create a Decision Tree.

**Evaluate:** Use the Test and Score widget to see how accurate your model is.



Filter...

-  Data
-  Transform
-  Visualize
-  Model
-  Evaluate
-  Unsupervised

Select a widget to show its description.

See [workflow examples](#), [YouTube tutorials](#), or open the [welcome screen](#).


















Filter...

Data

			
File	CSV File Import	Datasets	SQL Table
			
Data Table	Paint Data	Data Info	Rank
			
Edit Domain	Color	Column Statistics	Save Data

-  Transform
-  Visualize
-  Model
-  Evaluate
-  Unsupervised

Select a widget to show its description.

See [workflow examples](#), [YouTube tutorials](#), or open the [welcome screen](#).




























Filter...

Transform

 Data Sampler	 Select Columns	 Select Rows	 Transpose
 Split	 Merge Data	 Concatenate	 Select by Data Index
 Unique	 Aggregate Columns	 Group by	 Pivot Table
 Apply Domain	 Preprocess	 Impute	 Continuize
 Discretize	 Randomize	 Purge Domain	 Melt

Select a widget to show its description.

See [workflow examples](#), [YouTube tutorials](#), or open the [welcome screen](#).




























Filter...

Visualize

 Tree Viewer	 Box Plot	 Violin Plot	 Distributions
 Scatter Plot	 Line Plot	 Bar Plot	 Sieve Diagram
 Mosaic Display	 FreeViz	 Linear Projection	 Radviz
 Heat Map	 Venn Diagram	 Silhouette Plot	 Pythagorean Tree
 Pythagorean Forest	 CN2 Rule Viewer	 Nomogram	 Scoring Sheet Viewer





















Select a widget to show its description.

See [workflow examples](#), [YouTube tutorials](#), or open the [welcome screen](#).

Information, Home, Text, Eraser, Pause, Help icons

Filter...

Model

 Constant	 CN2 Rule Induction	 Calibrated Learner	 kNN
 Tree	 Random Forest	 Gradient Boosting	 SVM
 Linear Regression	 Logistic Regression	 Naive Bayes	 Scoring Sheet
 AdaBoost	 PLS	 Curve Fit	 Neural Network
 Stochastic Gradient De...	 Stacking	 Save Model	 Load Model

Select a widget to show its description.

See [workflow examples](#), [YouTube tutorials](#), or open the [welcome screen](#).

Information, Home, Text, Edit, Pause, Help icons

Filter...

AdaBoost	PLS	Curve Fit	Neural Network
Stochastic Gradient De...	Stacking	Save Model	Load Model
<b>Evaluate</b>			
Test and Score	Predictions	Feature as Predictor	Confusion Matrix
ROC Analysis	Performance Curve	Calibration Plot	Permutation Plot
Parameter Fitter			
<b>Unsupervised</b>			


















Select a widget to show its description.

See [workflow examples](#), [YouTube tutorials](#), or open the [welcome screen](#).

Information, Home, Text, Eraser, Pause, Help icons

Filter...

### Unsupervised

 Distance File	 Distance Matrix	 t-SNE	 Correlations
 Distance Map	 Hierarchical Clustering	 k-Means	 Louvain Clustering
 DBSCAN	 Manifold Learning	 Outliers	 PCA
 Neighbors	 Correspondence Analysis	 Distances	 Distance Transformation
 MDS	 Save Distance Matrix	 Self-Organizing Map	

Select a widget to show its description.

See [workflow examples](#), [YouTube tutorials](#), or open the [welcome screen](#).

Navigation icons: Home, Grid, Text, Eraser, Pause, Help

## **Next Step**

**I would like to explain how to perform a specific task in Orange, like how to load an Excel file or how to create a Scatter Plot?**

# Conclusion

Orange bridges the gap between complex data science and intuitive design.

It is a powerful, free, and extensible tool for anyone looking to uncover hidden patterns in data.

Final Thought: "Data mining shouldn't be a chore; Orange makes it fruitful and fun!"

**Thanks**