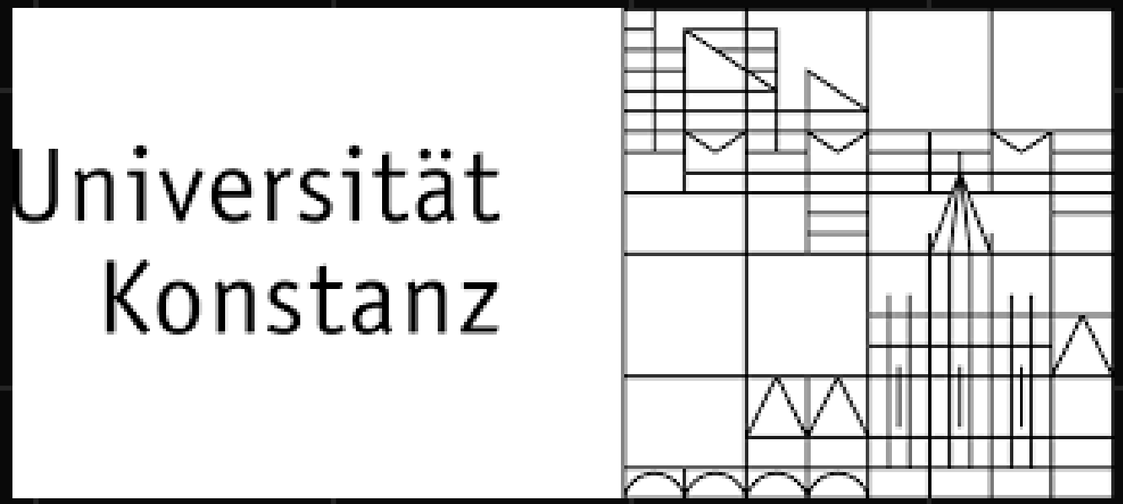
Introduction to KNIME

Overview:

- Open-source data analytics, reporting, and integration platform.
- Allows users to create data flows (workflows), execute selected analysis steps, and visualize results.

History:

- Developed by the University of Konstanz in Germany.
- Widely used in various industries including pharmaceuticals, financial services, and manufacturing.

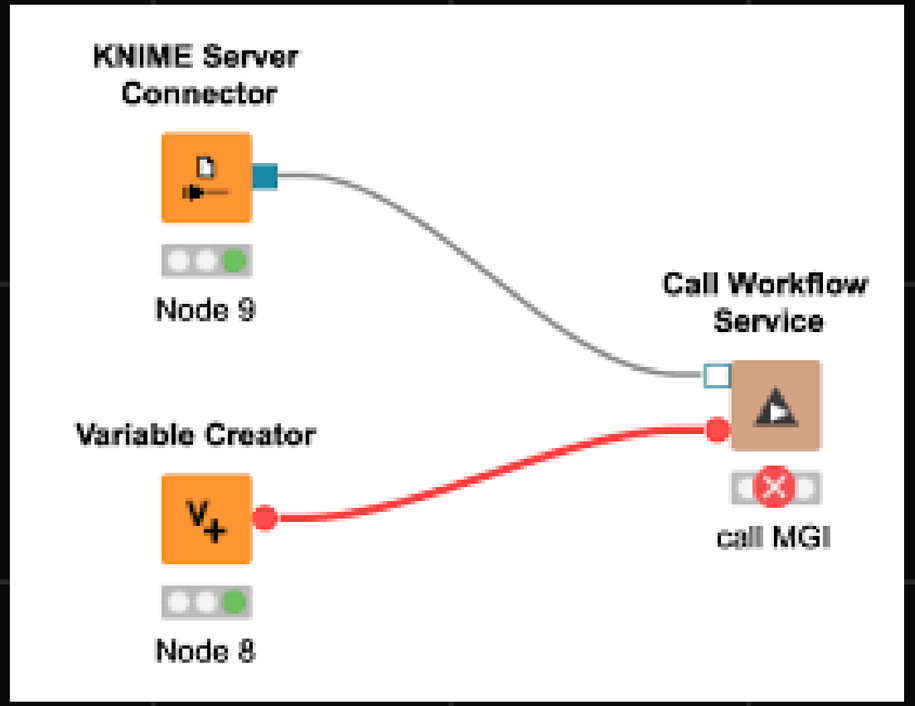


What is KNIME

- Visual programming tool where users can drag and drop nodes to create data pipelines.

Components:

- Nodes: Individual steps in a workflow (e.g., data reader, data transformation).
- Connectors: Link nodes together to form a complete workflow.

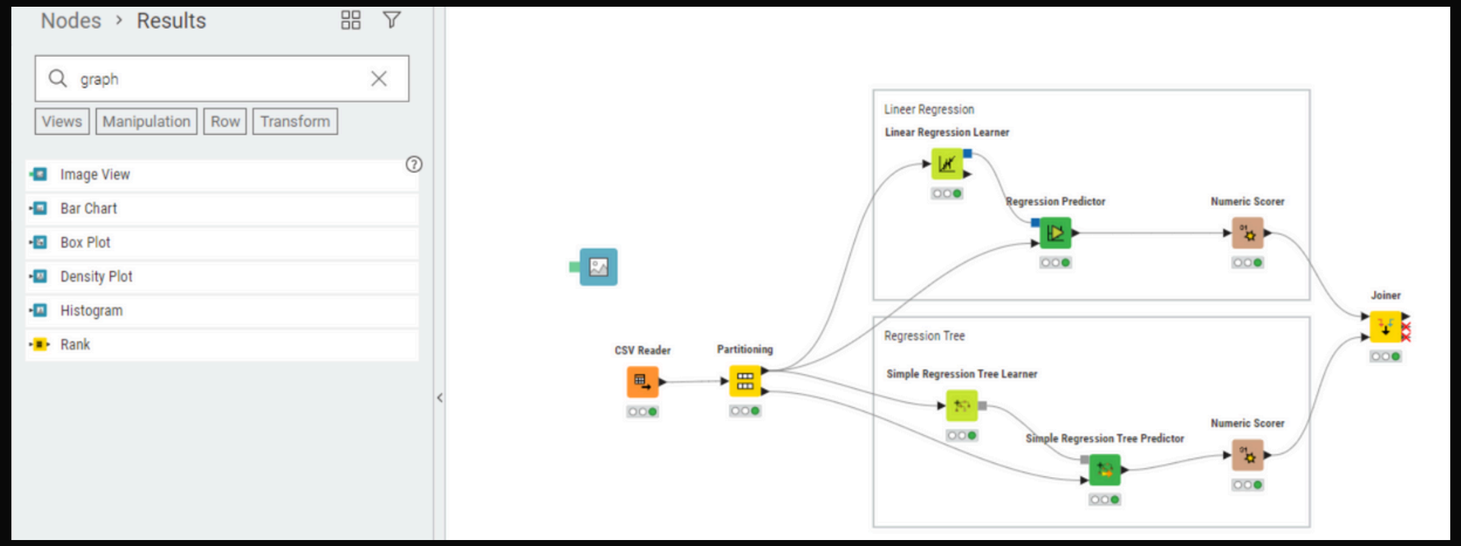


Sample KNIME workflow

✳ Creating a Workflow

1

- Start with a new KNIME workflow project.
- Drag nodes from the node repository to the workflow editor.
- Connect nodes to define the data flow.



Data Preprocessing in KNIME

Preparing data is crucial for accurate analysis and modeling.

Data Cleaning

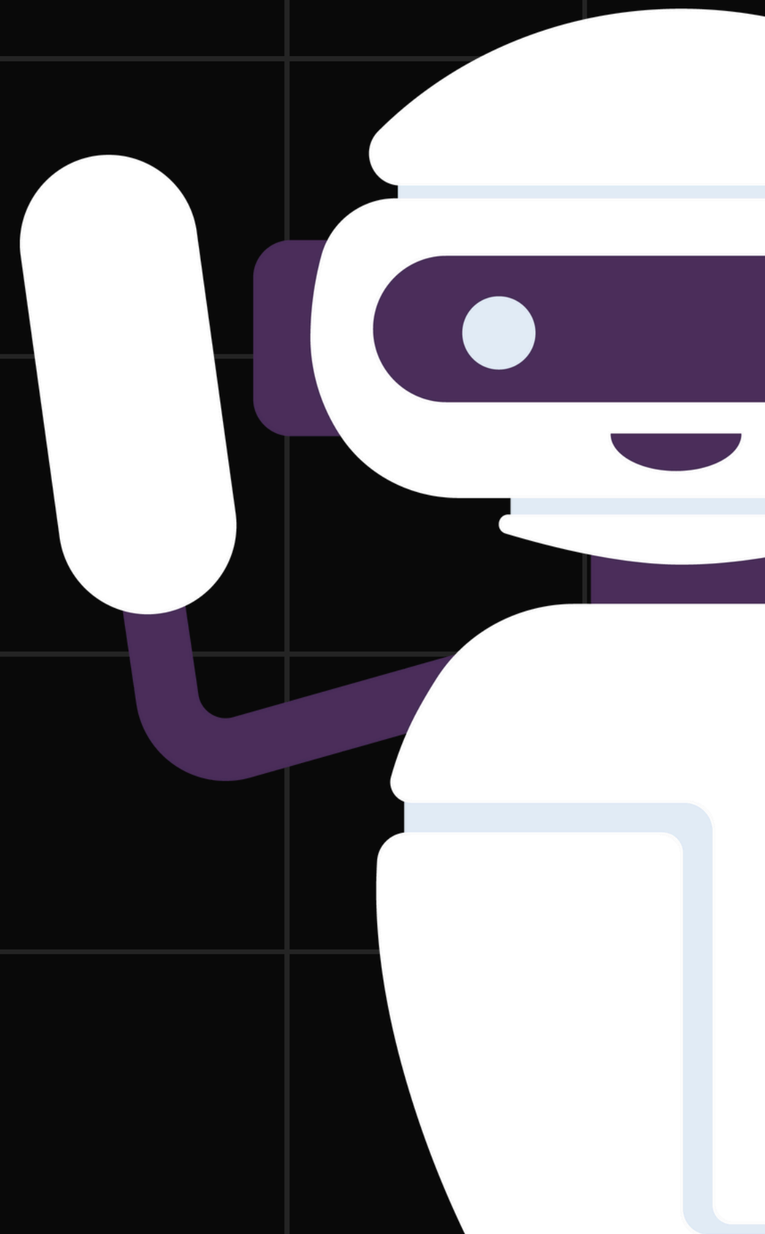
Handling missing values, removing duplicates.

Data Transformation

Normalization, scaling, encoding categorical variables.

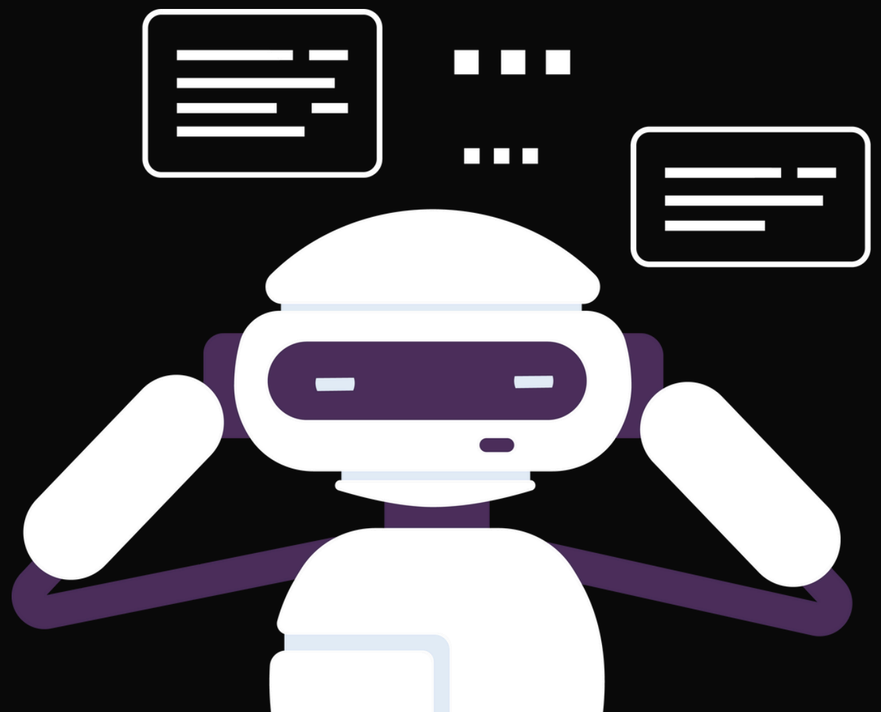
Feature Selection

electing important features for modeling.



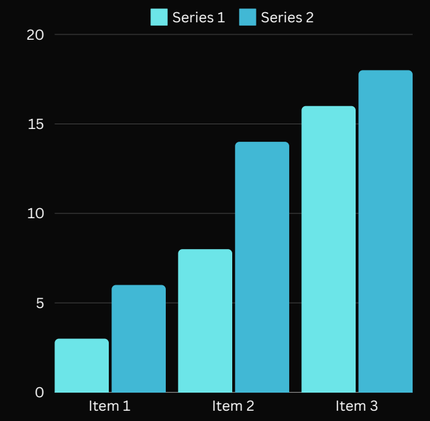
Visualizations in KNIME

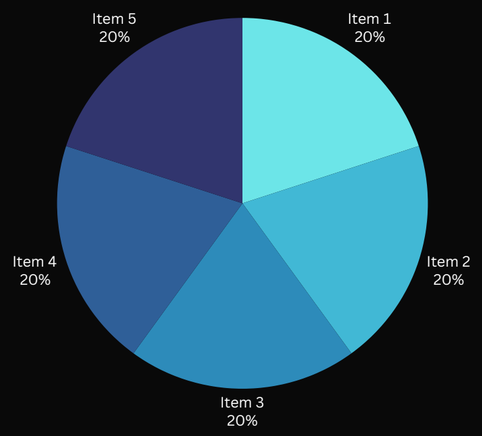
KNIME provides various nodes for data visualization.

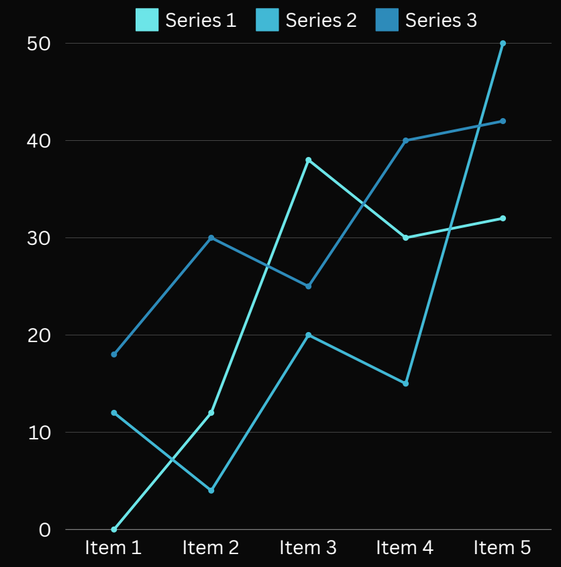


Types of Visualizations:

Bar Charts, Line Plots, Scatter Plots, Histograms, Heatmaps.







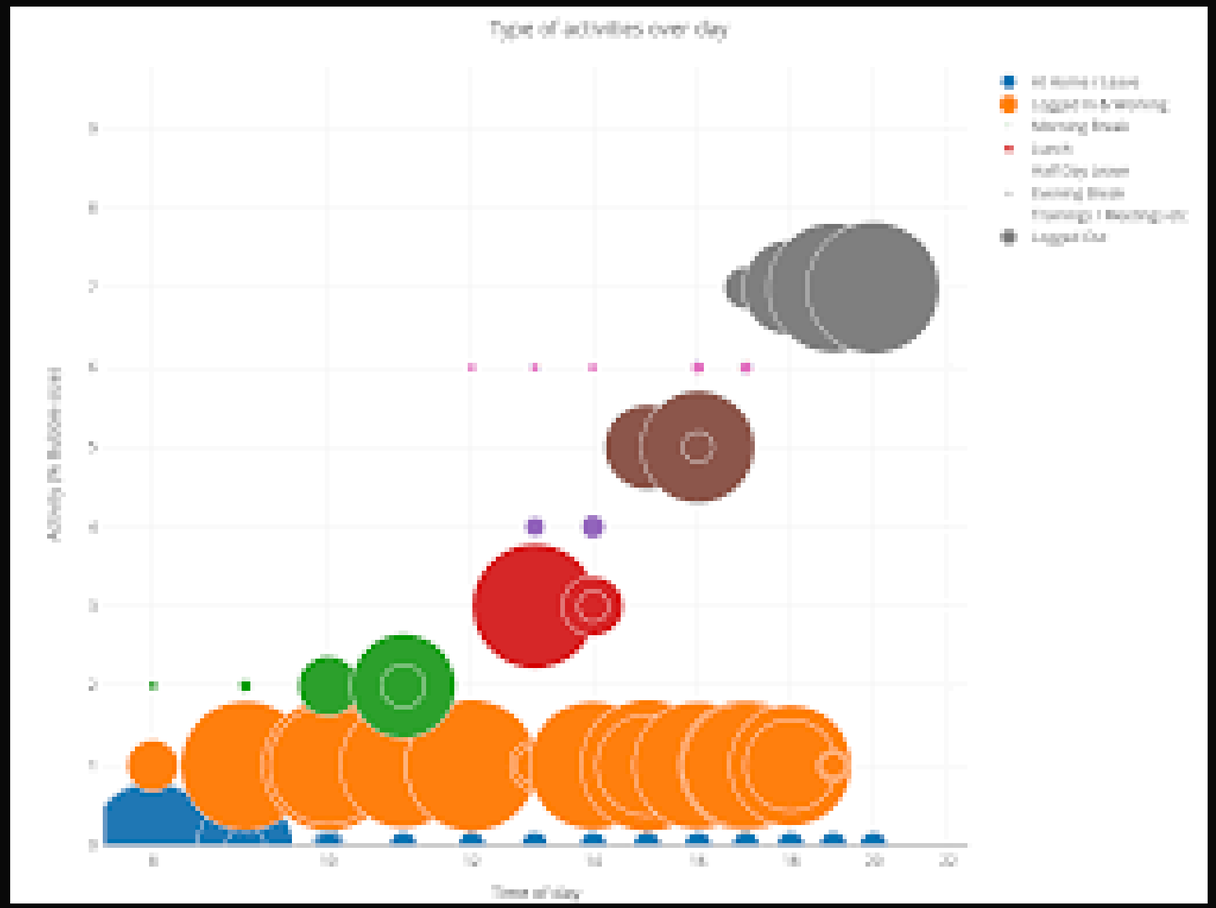
Visualizations in KNIME

Interactive Dashboards:

Using the KNIME WebPortal for interactive visualization.

Steps:

- **Select appropriate visualization nodes from the repository.**
- **Configure node settings to customize the visual output.**
- **Connect to data nodes to visualize the data.**



Modelling in KNIME

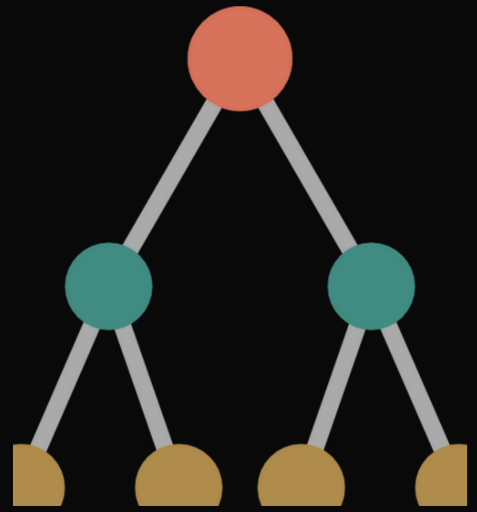
Types of models

Supervised Learning

- Regression
- Classification

Unsupervised Learning

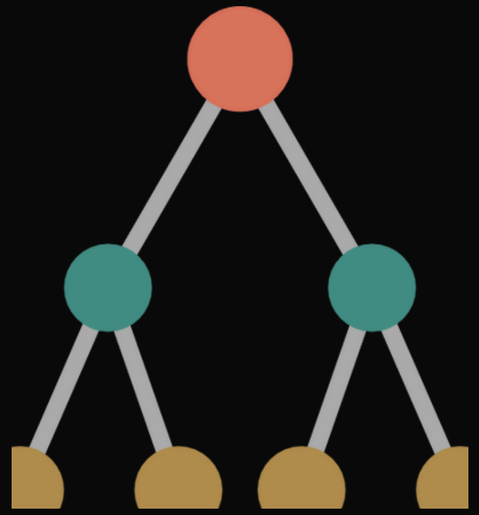
- Clustering
- Association Rules



Modelling in KNIME

Workflow

1. Splitting data into training and test sets
2. Training the model using various algorithms (e.g., Decision Trees, SVM, k-Means).
3. Evaluating model performance.



Modelling in KNIME

Learner Nodes

- Decision Tree Learner
- Logistic Regression Learner

Predictor Nodes

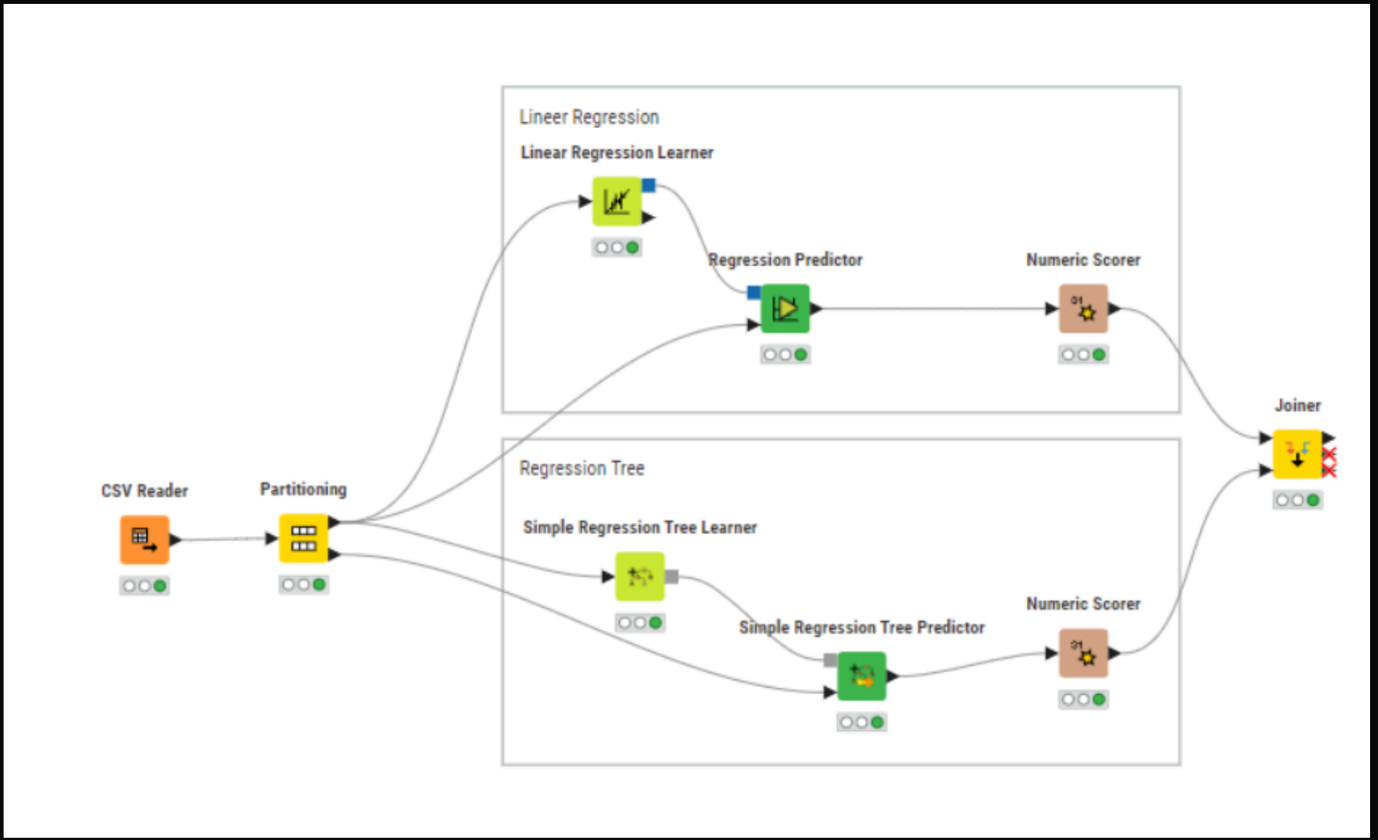
- Decision Tree Predictor
- Logistic Regression Predictor

Predictor Nodes

- Scorer
- ROC Curve

Modelling in KNIME

Example



Deployment

Benefits

- enabling automated
- regular execution
- integration
- scheduling
- control access with user roles and permissions

Thankyou

@reallygreatsite