



Beginner's guide for KNIME: What, Why, How

By : Team Cilans

Introduction

Data Science, AI and Machine Learning are now becoming integral part of many industry verticals including healthcare, Fintech, retail, manufacturing, utilities, media and many more. To develop model in Machine Learning, one needs to understand Linear Algebra, Statistics, and other important concepts in Mathematics. Even though you are already comfortable with the subjects mentioned, you still need to learn not only 'HOW TO CODE' (in Python/R/Go) but also several important concepts in Computer Science like Data structure, Algorithm and Database. Additionally, in order to start a project in Machine Learning, one needs to learn how to install and set up the whole coding environment, as well as how to use the command line. These issues are tedious and become one of the biggest challenges for newcomers and beginners in learning several important concepts simultaneously in order to get into Data Science, Data Analysis, and Machine Learning spaces. Unfortunately, some people give up and never look back to Machine Learning again.



However, good news is that there are some solutions. With great development in GUI-based applications, the introduction of KNIME is a major game changer for non-programmers. We can call Knime as a drag and drop data science tool. The major benefit of KNIME is that no programming knowledge is required. If you know how to use Microsoft Excel, KNIME gives you the same feeling. All you need to do is just get into KNIME website and download the [application](#) and, you are ready to get into world of KNIME without any further setups required and can start exploring following activities.

- To load a dataset (excel, csv, file, database etc), one needs to just drag and drop your file in!
- To clean and pre-process data, just click the functions through various preloaded nodes!
- To select, train, and test Machine Learning models, just drag and drop any model you want!
- To visualize your interesting findings, just drag and drop any kind of graph you want!



Cilans Systems Pvt Ltd

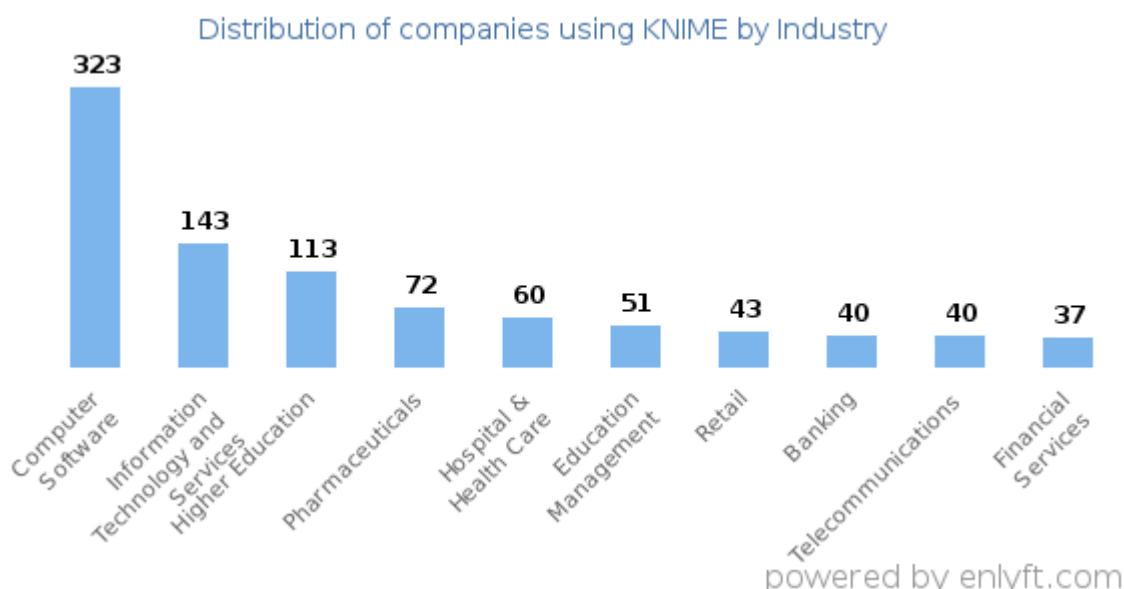
This allows you to focus your efforts on applying Machine Learning algorithms and techniques to your problems and subjects you are interested in the first day of your work!

In summary, to use KNIME, all you need is to just simply define the Workflow between a variety of predefined nodes which are already provided in its repository. This is very convenient since KNIME already provides several predefined components called “Nodes” for numerous different tasks such as Reading data, Cleaning data, Applying ML algorithms, Visualizing data in different formats, and Analyzing results.

What is KNIME?

KNIME stands for “Konstanz Information Miner” which was developed at the University of Konstanz in Germany in early 2004. It is an Open-Source software written in Java on the Eclipse SDK platform. KNIME platform relies on pre-defined components called ‘Nodes’ for building and executing ‘Workflows’. Its core functionality is available for tasks such as machine learning, data mining, analysis, and manipulation. Additionally, the extra features and functionality are available in KNIME through various extensions and supports from numerous community support groups and vendors.

Who uses Knime



As chart (source: enlyft.com), KNIME is gaining popularity in a number of industry verticals. Computer Software (18%), Information Technology and Services (8%) and Higher Education (6%) are the largest segments.



Cilans Systems Pvt Ltd

Alteryx (Based on R, Not an open-source, License Fee and offer enterprise grade support) and RapidMiner (limited functionality in the free version) are alternative software. Also IBM SPSS, SAS or FOUNDRY from Palantir offer some modelling capabilities as KNIME. We expect some active user base of the above software, will shift to KNIME in coming years, as KNIME community is continue to grow.



As per Gartner 2019 Magic Quadrant for Data Science and Machine Learning Platforms, KNIME is rated highly on a leadership scale.

Why use KNIME?

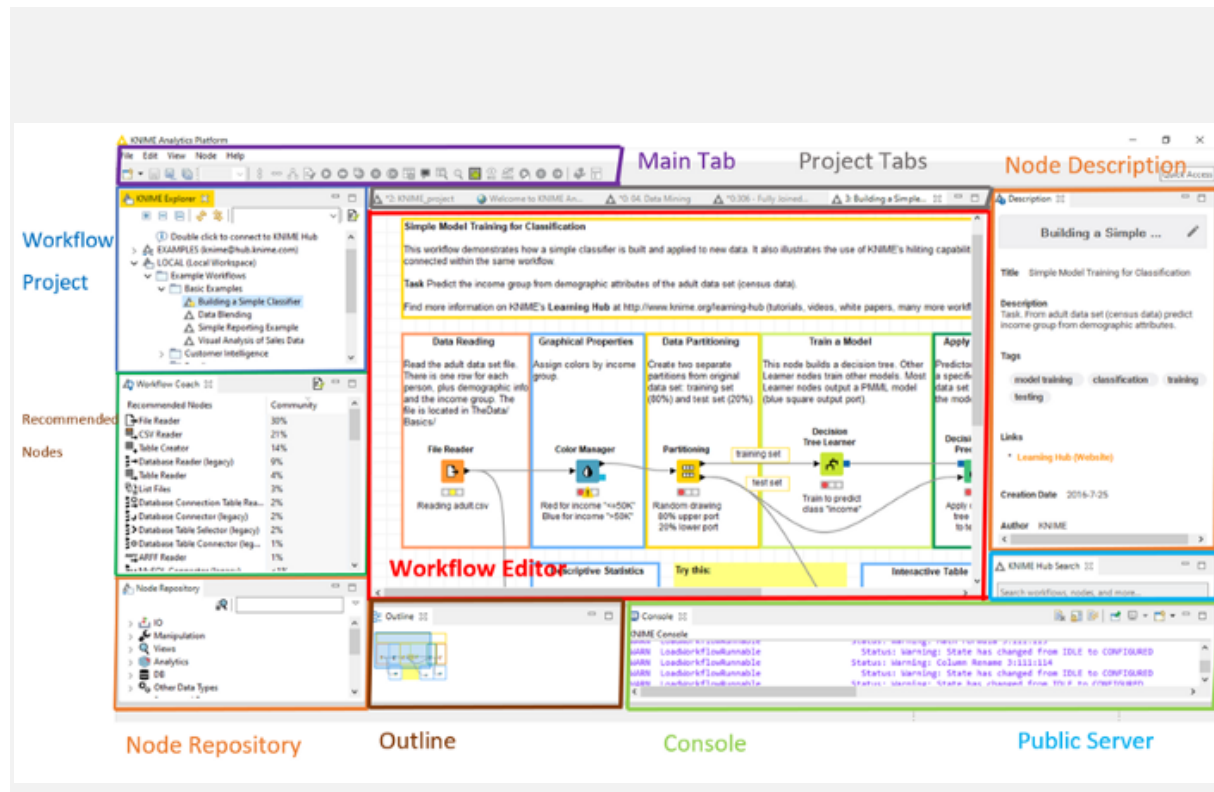
KNIME is a GUI-driven platform for data analytics. This means that knowledge of coding is not a requirement (though sometimes writing code is needed but minimal if you want to add more complexity into your workflow.) KNIME is an open-source application, meaning that it is free to use. Also, it has a growing community which is providing innovative and powerful functions. Through KNIME, it is easier to understand and



Cilans Systems Pvt Ltd

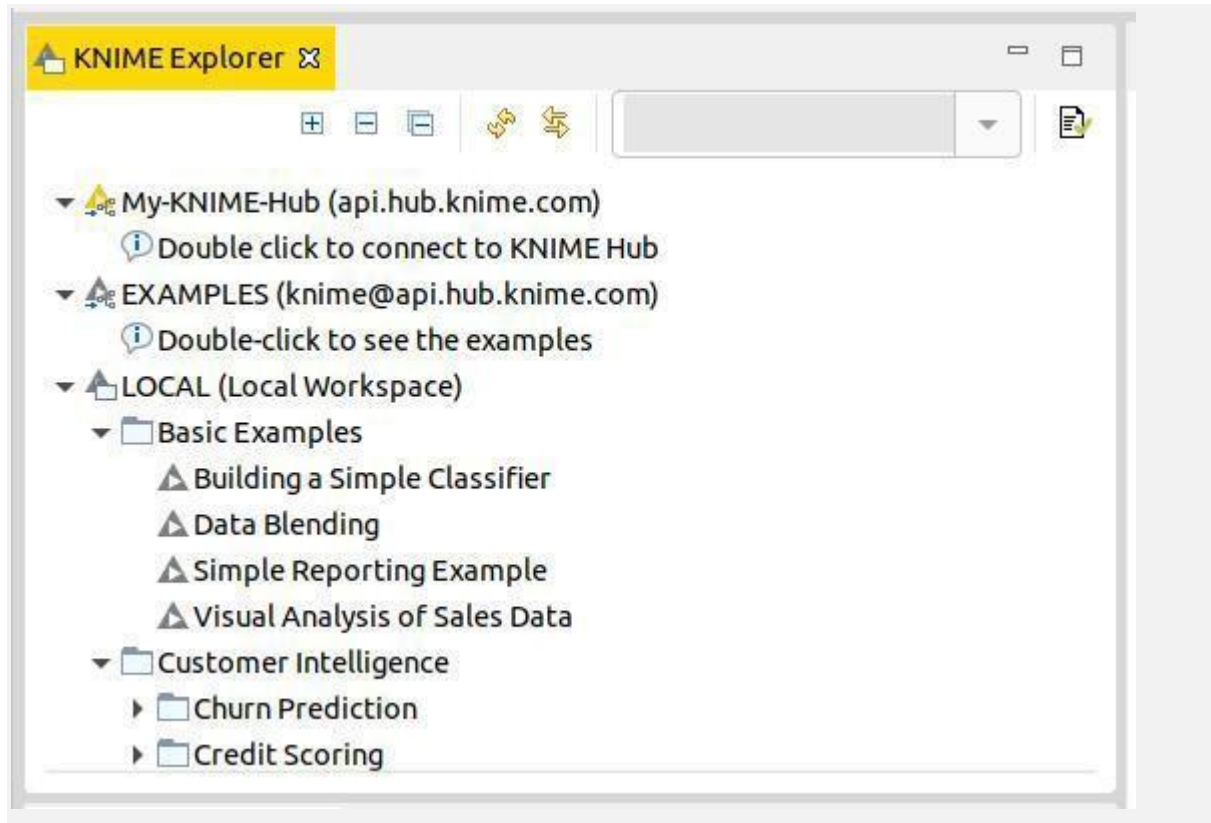
deploy the whole complex processes of Machine Learning from start to finish by means of creating, editing, annotating, visualizing, and sharing workflows. Furthermore, it allows us to integrate data from many potential sources (files, database, web services). It enable to perform several essential ML related algorithms and functions ranging from basic I/O to data manipulations, data transformations, and data mining. In summary, KNIME helps consolidate the combination of various different processes into one single understandable Workflow.

Basic of overview of KNIME Platform and Workflow



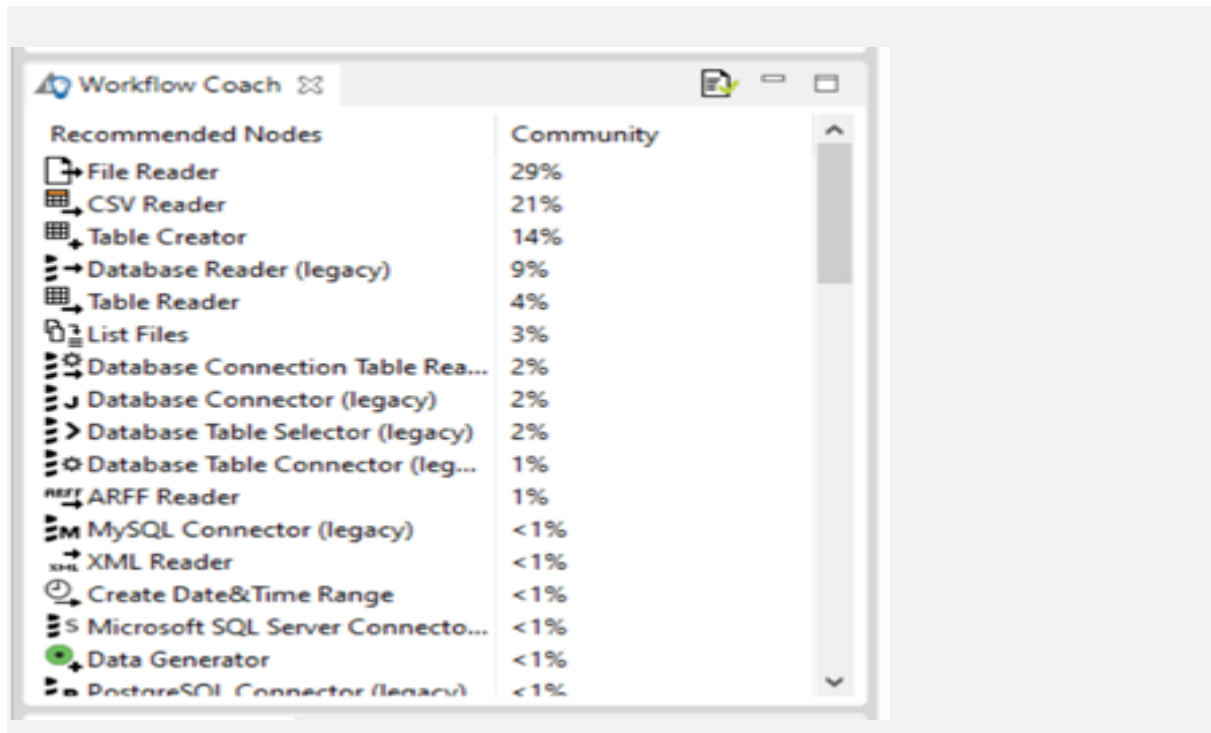
I). Workflow Project:

It consists of LOCAL workspaces which comprise all workspaces you have created from your own local machine, KNIME hub where you can connect with KNIME online server and community, and EXAMPLE workspace where you can get example projects that have already been created by KNIME community and ready to be used.



II). Recommended Nodes or Workflow Coach:

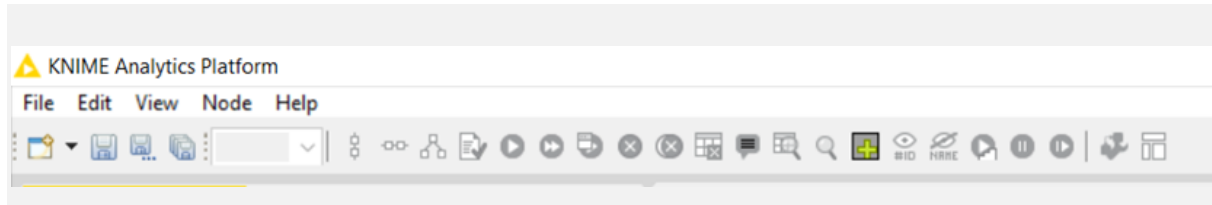
It lists nodes recommended based on the workflows built by the wide community of KNIME users.





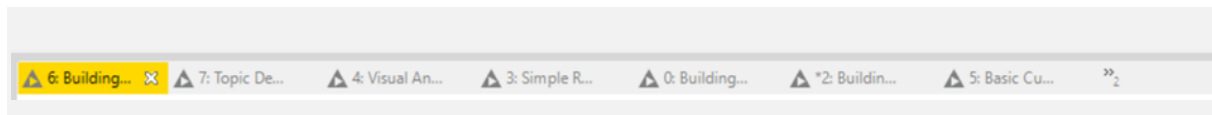
III). Main Tab:

We can call it a 'ToolBar' tab as well. It consists of various basic functions for operating KNIME such as a function to execute and cancel selected nodes.



IV). Project Tabs:

It shows our current projects as you can create and execute several projects at the same time.



V). Node Repository:

It consists of all the available nodes in core KNIME Analytics Platform and in the extensions (Also, the nodes you have installed are listed here). The nodes are nicely organized by categories based on the node functions. Under each main node category, you can expand and select specific nodes with your desired functions.

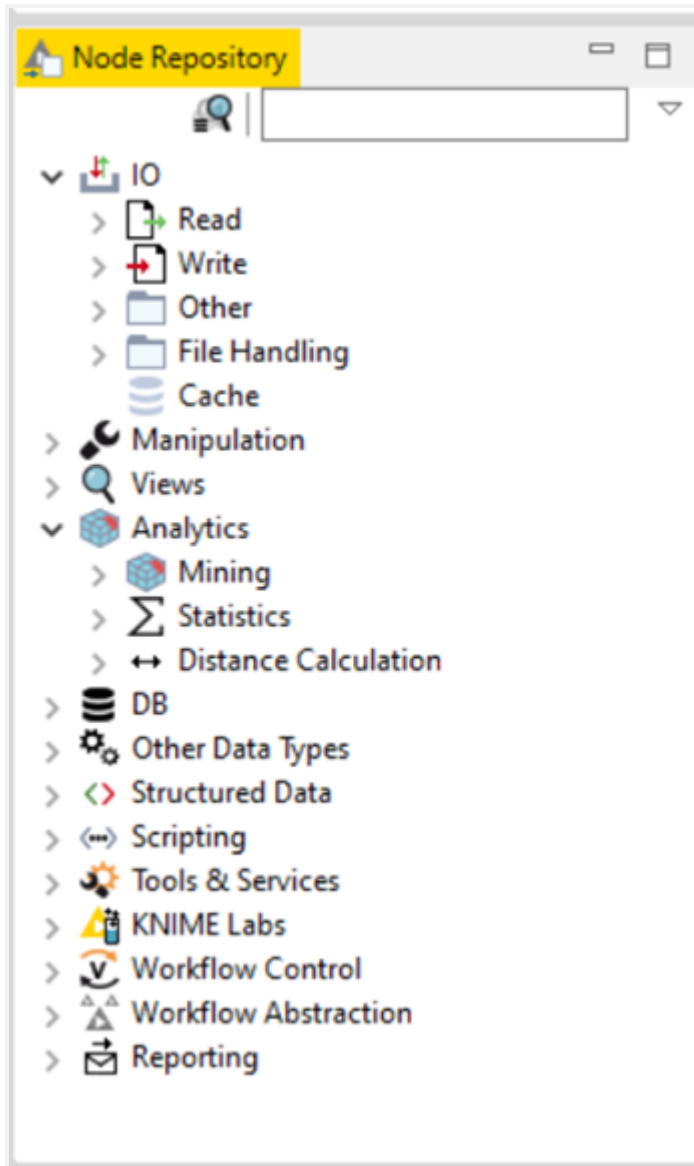
User can also use the Search box on the top of the node repository to find specific nodes.

Each **Nodes** can have 3 states.

Red: "Not Ready/Idle" state which means that the node is not yet configured and can not be executed with its current settings.

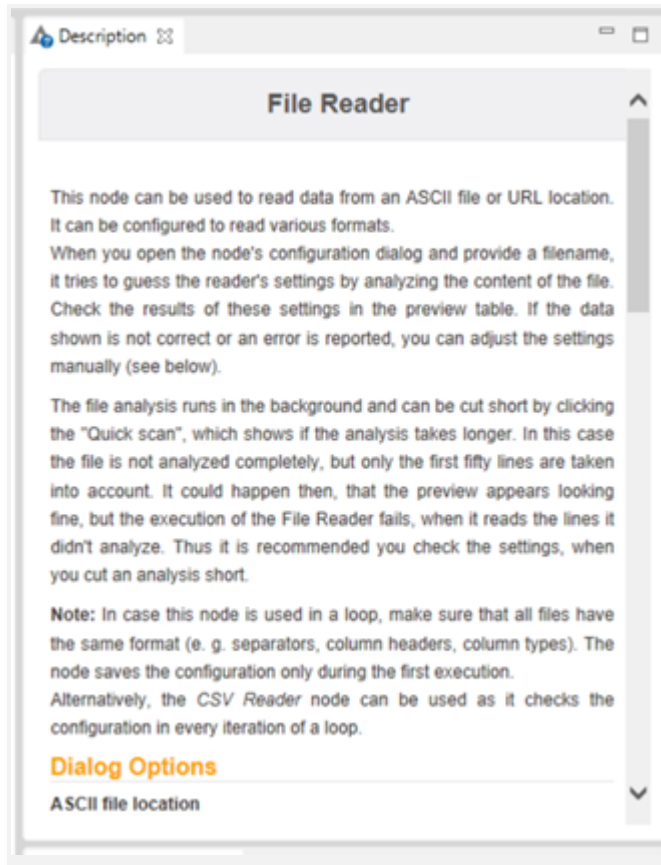
Yellow: "Ready/Configured" state which means that the node has been set up correctly and can be executed at any time

Green: "Executed" state which means that the node has been successfully executed and we can see the results at the final nodes (downstream nodes).



VI). Node Description:

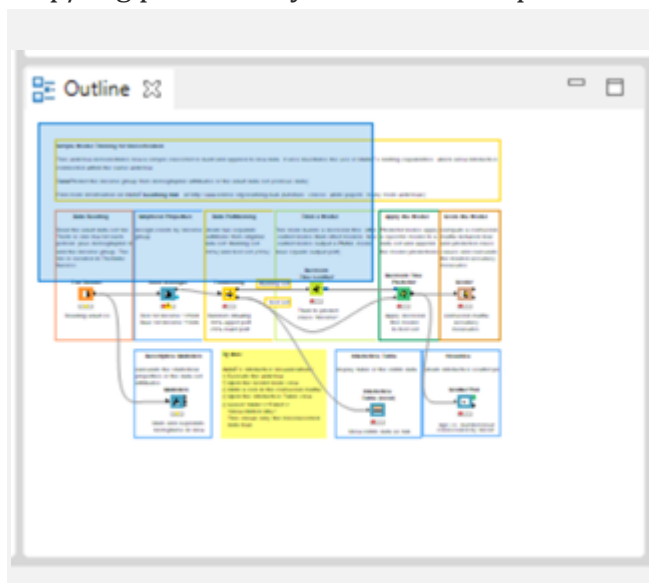
It shows the description of the currently active workflow or a selected node in the Workflow Editor or Node Repository. **For a beginner**, it is very useful in the initial stages of learning, who may not have deeper knowledge of Machine Learning, or forget the purposes of each node in the Workspace or in the Node Repository.



VII). Outline:

It is an overview of the currently active workflow.

it is very useful, particularly when your workflow is very big. The outline will work as a map/ big picture for your Workflow space.

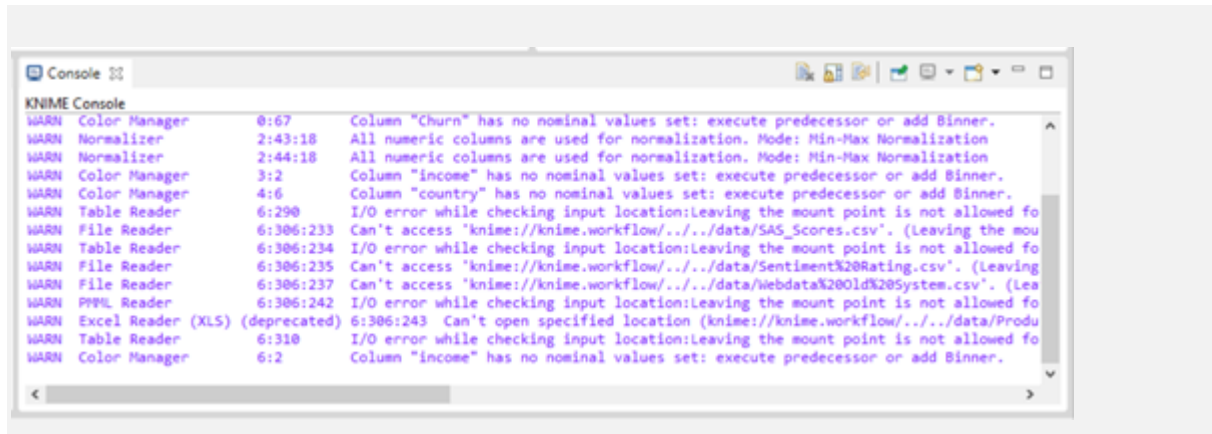




VIII). Console:

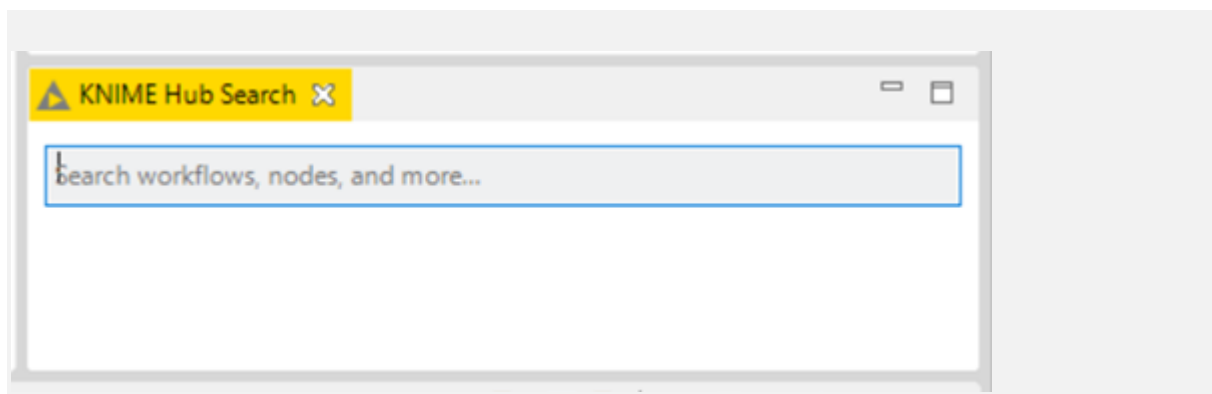
It shows the execution message and status which help indicate what is going on at the current workflow state such as successful operation, error in file, and so on.

It is very useful in helping diagnose the workflow and examine the analytics results.



IX). Public Server:

This tab helps you connect to KNIME server in case you want to search something on KNIME online hub.





Cilans Systems Pvt Ltd

Summary:

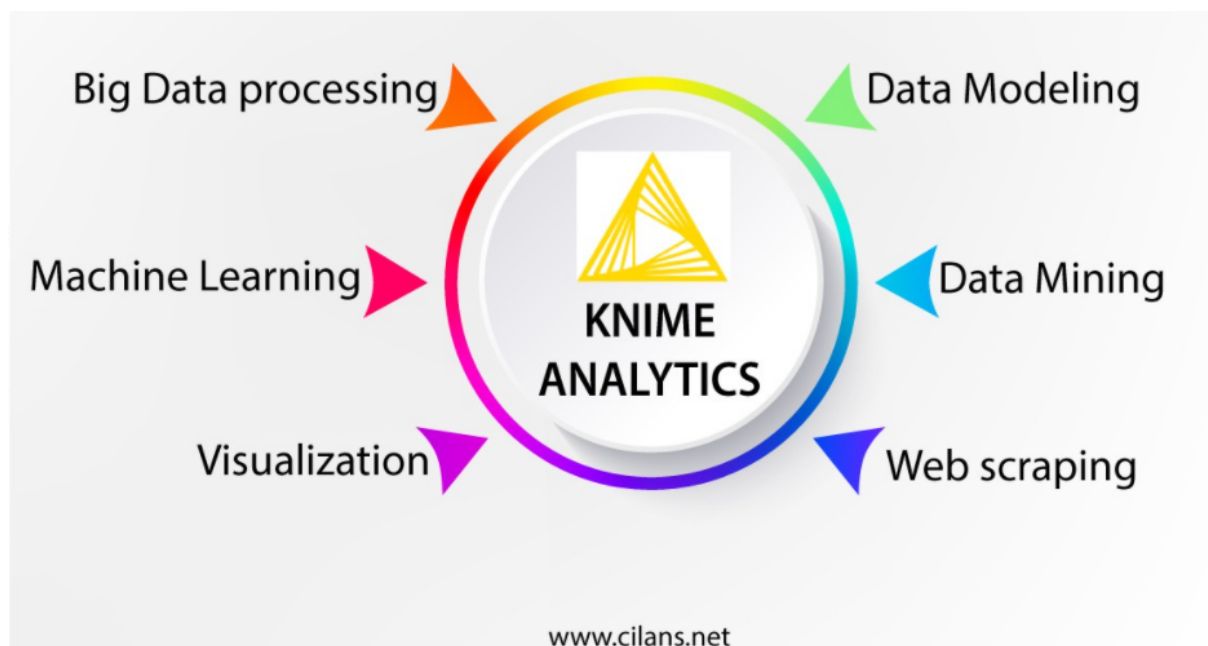
Since its launch in 2004, KNIME has evolved a lot. As highlighted below, (Source: Gartner peer insight), KNIME is very well rated on multiple fronts.

Ratings Overview

1: Evaluation & Contracting	<div><div></div><div></div><div></div><div></div><div></div></div>	4.6
2: Integration & Deployment	<div><div></div><div></div><div></div><div></div><div></div></div>	4.5
3: Service & Support	<div><div></div><div></div><div></div><div></div><div></div></div>	4.5
4: Product Capabilities	<div><div></div><div></div><div></div><div></div><div></div></div>	4.6

Poor 1 2 3 4 5 Excellent

Now, KNIME is one of the most comprehensive, user friendly and free analytics platform for performing machine learning, statistics, and ET tasks, as captured in the image below.



Once can start the journey at [Learning](#) page of KNIME.
Welcome to KNIME Club !

Authored by:

(Team Cilans):
Nikhil, Chintan, Kashyap



Cilans Systems Pvt Ltd

Kindly refer to Cilans Systems (www.cilans.net) for additional articles on Knime Blog series. Please contact us at info@cilans.net or what's app at +91 95588 19097 for any questions.