THE BASICS OF WORKING WITH KNIME ANALYTICS PLATFORM



Barbora Štětinová, 2024

1st Edition

The basics of working with KNIME Analytics Platform

Written and illustrated by:

Barbora Štětinová

Publisher:

Barbora Štětinová

Date of 1.edition: 2021

Edition: 2.

Date of 2.edition: 2024

Ing. Barbora Štětinová, MBA

The basics of working with KNIME Analytics Platform

© Barbora Štětinová, 2024

978-80-909371-1-6 (online, pdf)
978-80-909371-2-3 (online, ePub)
978-80-909371-3-0 (online, MOBI)
9798345512098 (paperback)

Book Content

Introduction
KNIME ANALYTICS PLATFORM
Introduction to software
Installation
Menu options
KNIME environment
CREATING A WORKFLOW AND WORKING WITH IT
LOADING DATA
INFORMATION ABOUT THE DATA
DATA MANIPULATION
Data joining
Data filtering
Data aggregation
Conditional column
Missing values
DATA VISUALIZATION
Histogram
Table View
Color Manager
Time Series
Data Relation, outliers
METANODES AND COMPONENTS
MACHINE LEARNING
Classification
Regression

The basics of working with KNIME Analytics Platform

Clustering

Association analysis

Machine Learning in KNIME

Conclusion

About the author

Introduction

Today, digital transformation is a heavily discussed topic, driven by the exponential increase in data generated through continuous technological advancements. Data today play a key role in understanding and analyzing situations in various sectors, from commerce and healthcare to banking, telecommunications, manufacturing, and customer behavior. However, data alone are not enough; their true potential lies in how we can turn them into value. To achieve efficiency and ensure competitiveness, it is crucial to set up the right processes and corporate culture that focus on effective data utilization.

This book was created after a long search for a suitable tool for data processing. In our work, like most of you, we processed data and data analyses in MS Excel. While we consider this software to be great, efficient, and very user-friendly, we began to encounter its limitations. Problems arose with processing large amounts of data and working with real-time data. Additionally, Excel does not offer the ability to process machine learning and deep learning techniques, or only to a limited extent.

So we embarked on a journey to find a suitable tool, during which we tried various software and programming languages

(e.g., Python, R) for data analysis. The tool that impressed us the most was the **KNIME Analytics Platform**, which offers a vast array of possibilities for working with data - from basic data analyses to the application of artificial neural networks - all within a user-friendly environment without the need for programming.

Our goal is to provide you with information about:

- The KNIME tool
- Possibilities for working in the KNIME environment
- Ways to work with data
- Tools for data transformation
- Data visualization
- Options for creating and using a basic machine learning model

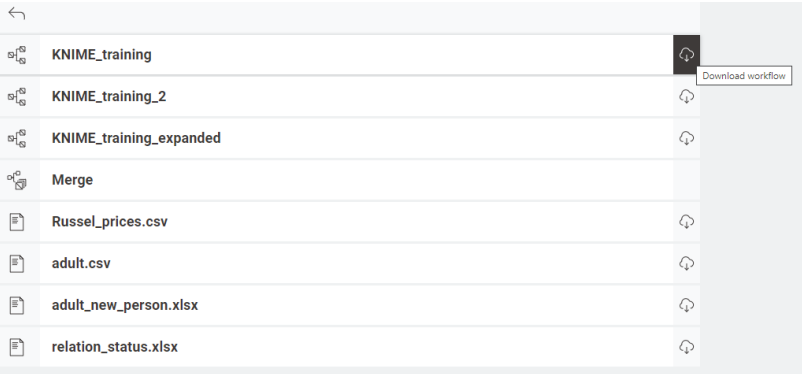
We have prepared a basic guide to introduce you to the KNIME Analytics Platform software!

Open for Innovation
KNIME

You can find all the underlying files and created WorkFlows on KNIME Hub:

https://kni.me/s/9MZ6-WMMTx4I-VBr

Simply download the files using the download icon.



We will come back to this download file later.

KNIME ANALYTICS PLATFORM

Introduction to the software

KNIME software is easily accessible (free in the desktop version) and is one of the frequently used platforms for data analysis, data automation, data visualization, and the creation of Machine Learning and Deep Learning (deep learning using neural networks) in large international organizations. Its advantage lies in its simplicity and the ability to create models without the need to use a programming language.

As already mentioned, the KNIME Analytics Platform software falls into the category of freeware, so you can download and use it completely free of charge. It was developed in an academic environment in Switzerland with the aim of rapid application of Machine Learning and Deep Learning methods and quickly found practical use in the commercial sector as well. Its great advantage is its simplicity, allowing you to create models by inserting and connecting nodes (work packages in the form of icons) into workflows within the working environment.

Therefore, if programming and coding are foreign to you, you don't necessarily have to despair that data analysis and data science aren't for you. With each node representing an activity, you have the option to set specific parameters for that activity. KNIME also offers a considerable number of

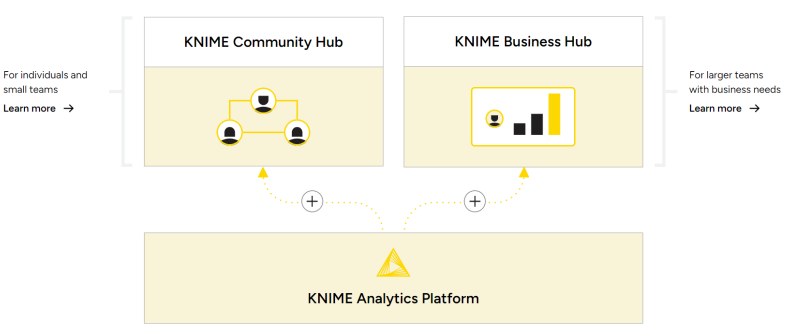
sample workflows that can be used and adapted to your specific task.

KNIME provides two products: KNIME Analytics Platform and KNIME Server, now known as KNIME Business Hub.

The KNIME Analytics Platform includes algorithms for data operations necessary for data analysis projects—that is, from data loading, data transformation, visualization, all the way to building machine learning and deep learning models. As already mentioned, it is open-source software (free to use) for all your data processing needs. Downloading from the KNIME website is free (https://www.knime.com/downloads) and free to use. KNIME is based on visual programming solutions. Visual programming is a key feature of the KNIME Analytics Platform, and thanks to it, using the tool is very easy. The visual programming graphical user interface (GUI) guides you through all the necessary steps to create workflows of individual nodes. KNIME was designed to be open to various data formats, data types, data sources, and even external tools such as Python and R.

The final step in data analysis is deployment into production. That is, easy, convenient, and secure deployment. This process of moving the application into the real world is called "Deploy to production." In this case, it is appropriate to utilize

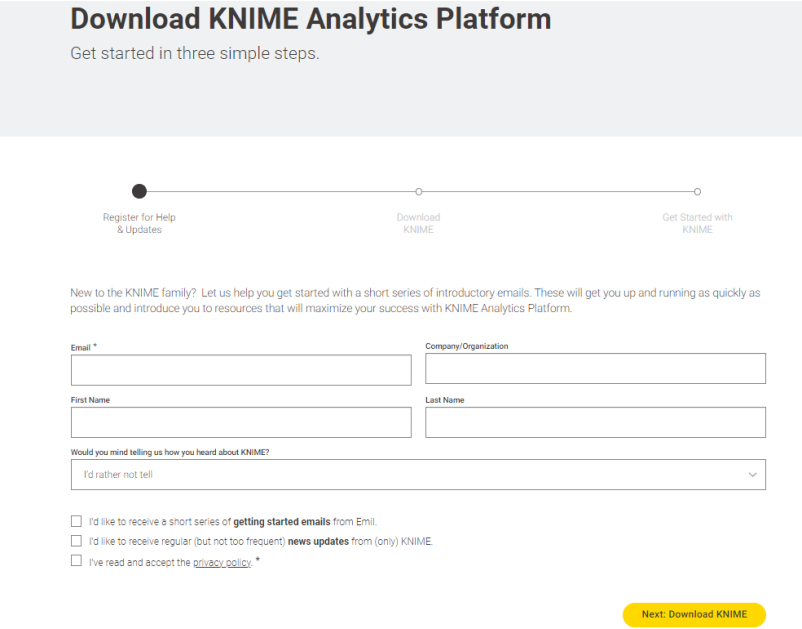
another option from KNIME, namely KNIME Server, now known as **KNIME Business Hub**, which comes with an annual license fee. It offers a protected environment for collaboration within the company, automation, and allows access to workflows and interactive reports through a web application.



Source: https://www.knime.com/software-overview

Installation

To install the KNIME Analytics Platform, visit [http://www.knime.com/downloads], complete the registration, and download the version that matches your operating system.

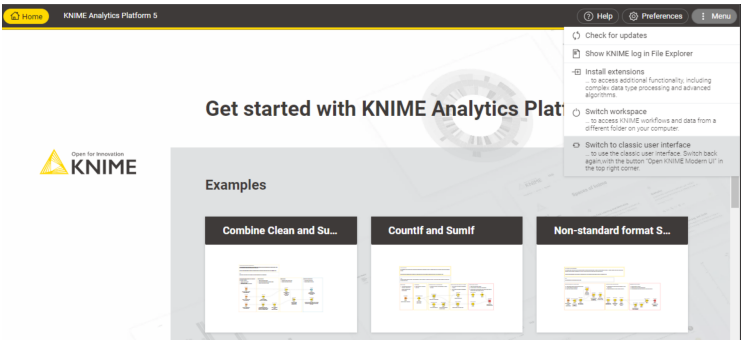


Source: https://www.knime.com/downloads

After downloading the package, find and open it (wherever you chose to place it during installation) and follow the

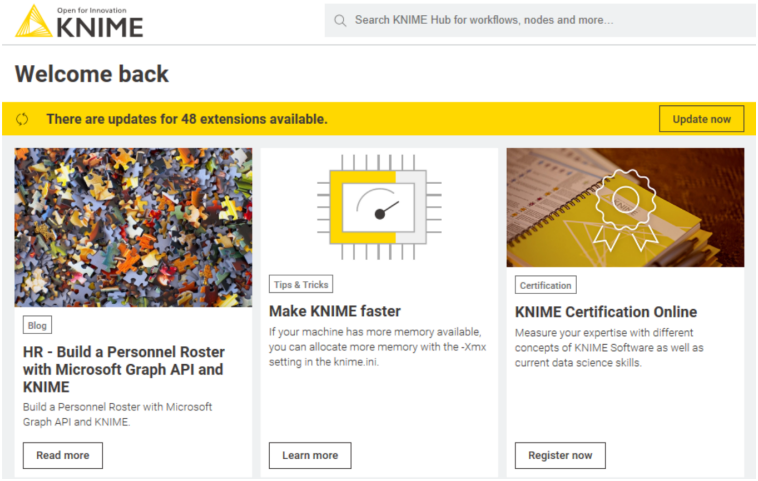
instructions. Install it into any directory for which you have write permissions.

For our purposes, after first opening the KNIME Analytics Platform, I recommend switching the user interface to the classic interface, which allows for more advanced functions. While the new UX design is indeed better in terms of design, we will be working in the standard environment, which you can access via the Menu by selecting **"Switch to classic user interface"**.



IMPORTANT:

IMPORTANT: DUE TO THE FAST-PACED ADVANCEMENTS IN DATA SCIENCE AND NUMEROUS CONTRIBUTORS TO KNIME SOFTWARE, VERSION AND PLUGIN UPDATES ARE FREQUENT.



So, what is KNIME used for?

KNIME helps you with:

- Data loading (e.g., from Excel, CSV, tables, tables created in KNIME, SQL databases, text editors, MySQL, PostgreSQL databases, XML, JSON, etc.)
- Data transformation (using transformations with entire tables, rows, or columns)
- Data visualization (charts, statistics)
- Building Machine Learning and Deep Learning models for data prediction

- Saving results (e.g., into Excel, CSV, databases, Power BI Service, etc.)
- Automation of data science and data analysis processes
- Public sharing via KNIME Hub
- Corporate sharing and model deployment using KNIME Server and KNIME WebPortal, now under KNIME Business Hub

If you're unsure or need assistance, tips, tricks, and advice from KNIME experts or advanced users can be found on the blog: https://www.knime.com/blog.

Your work in KNIME is made easier by existing templates directly within the platform and numerous examples available on KNIME Hub, which is public and free for all KNIME users. KNIME supports full-text search, and the creation of workflows is user-friendly - often referred to as "click and drag" using drag-and-drop nodes.

Let's go through the individual parts of the platform.

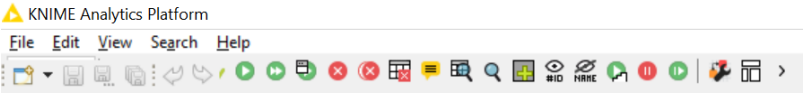
IMPORTANT

Due to frequent updates, we have released this second edition for you, where nodes that have undergone changes will be updated with additional images, including examples from KNIME 5.x versions.

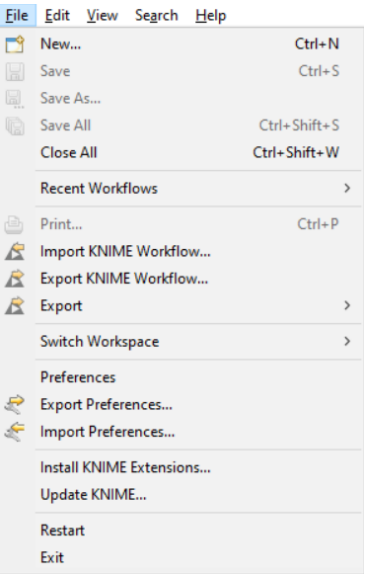
You can recognize the modified nodes by the green border.

Menu options

The menu for working in KNIME is located, just like in most software - for example, from Microsoft - in the top bar of the platform.



The first menu option is the **File** menu, where you can save, open, create a new one, or import and export an existing workflow.



Workflows are saved as `.knwf` files, and an entire group containing multiple workflows - or even data sources,

exported models, and many other items - is then saved under the `.knwr` file type.

Additionally, this menu offers the installation of add-ons via the **Install KNIME Extensions option**, which are not part of the basic version of the platform but can be installed for free if needed. For example, if you need to send processed data in KNIME to a MS Power BI and continue with visualization in the Power BI tool, it is sufficient to install the Power BI extension (and, of course, you need to have the mentioned Power BI account from Microsoft for example, as part of Microsoft Office 365 or a purchased MS Power BI Pro account) and then send the data to this Power BI platform via appropriate node.

A wide range of extensions is available from various areas of data work, and they can also be found through full-text search.

The **Edit** menu allows classic editing actions when working within a file, such as undo, redo, copy, paste and so on.

Vážení čtenáři, právě jste dočetli ukázku z knihy ***The basics of working with KNIME Analytics Platform*.**

Pokud se Vám ukázka líbila, na našem webu si můžete zakoupit celou knihu.