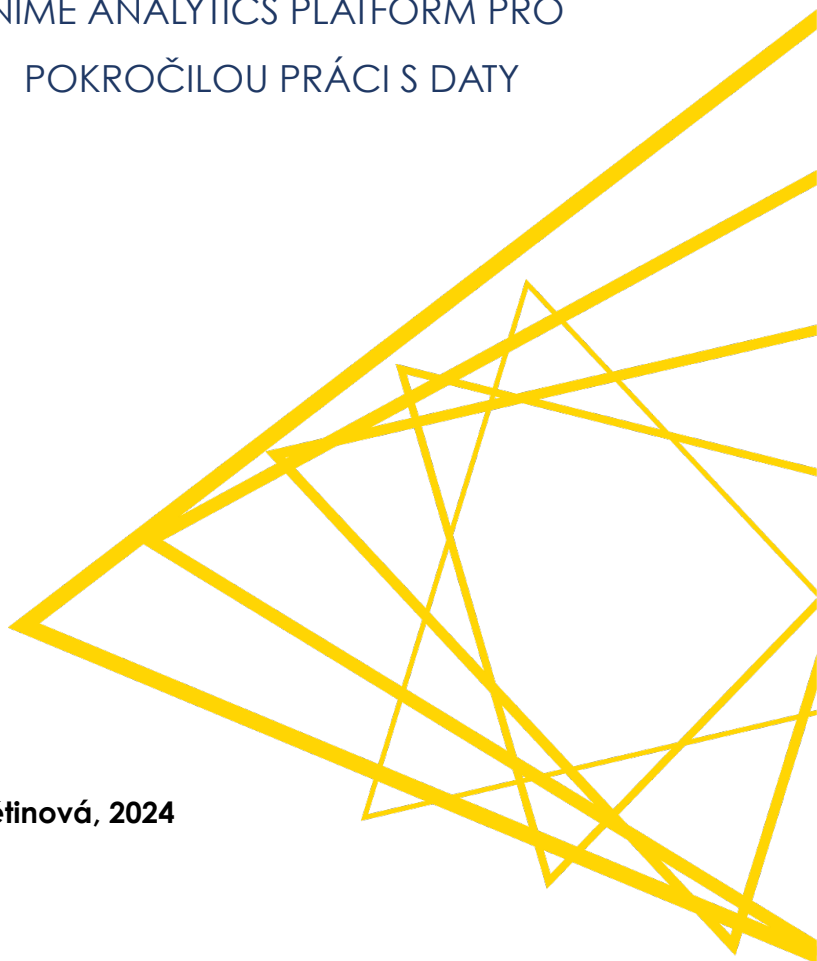


ZÁKLADY PRÁCE S KNIME ANALYTICS PLATFORM

KNIME ANALYTICS PLATFORM PRO
POKROČILOU PRÁCI S DATY

Barbora Štětinová, 2024

2.vydání



Základy práce s KNIME Analytics Platform

Název: Základy práce s KNIME Analytics Platform : KNIME Analytics Platform pro pokročilou práci s daty

Napsal a ilustroval:

Barbora Štětínová

Vydal:

Bečov nad Teplou : Barbora Štětínová

Datum 1.vydání: 2021

Vydání: 2.

Datum 2.vydání: 2024

Ing. Barbora Štětínová, MBA

© Barbora Štětínová, 2024

978-80-909348-8-7 (online, pdf)

978-80-909348-9-4 (online, ePub)

978-80-909371-0-9 (online, MOBI)

Základy práce s KNIME Analytics Platform

OBSAH KNIHY

ÚVOD	5
O KNIME ANALYTICS PLATFORM	8
Úvod do software	8
Instalace	11
Možnosti z Menu	15
Popis prostředí	18
TVORBA WORKFLOW A PRÁCE S NÍM	23
NAČTENÍ DAT	30
INFORMACE O DATECH	41
OPERACE S DATY	46
Spojení dat	46
Filtrování dat	58
Agregace dat	67
Podmíněný sloupec	75
Chybějící hodnoty	78
VIZUALIZACE DAT	85
Histogram – Histogram uzel	87
Tabulka – Table View uzel	96
Přiřazení barev – Color Manager uzel	99
Časové řady – Line Plot uzel	102
Relace mezi daty – Scatter Plot uzel, Outliers uzel	112
KONTEJNEROVÉ UZLY	119
STROJOVÉ UČENÍ (MACHINE LEARNING) V KNIME	136
Úvod do strojového učení	136
Klasifikace	139

Základy práce s KNIME Analytics Platform

Regrese	140
Clustering	141
Asociační analýza	141
UKÁZKA TVORBY MACHINE LEARNING MODELU	143
ZÁVĚREM	166
O AUTOROVI	170

Základy práce s KNIME Analytics Platform

ÚVOD

Digitální transformace se dnes stala jedním z nejčastěji diskutovaných témat díky rychlému technologickému pokroku, který vedl k exponenciálnímu nárůstu objemu generovaných dat. Data dnes hrají klíčovou roli v porozumění a analýze situací v různých odvětvích – od obchodu a zdravotnictví až po bankovníctví, telekomunikace, výrobu nebo zákaznické chování. Avšak data sama o sobě nestačí; jejich skutečný potenciál spočívá v tom, jak je umíme proměnit v **hodnotu**. Abychom dosáhli efektivity a zajistili si konkurenceschopnost, je zásadní nastavit správné procesy a firemní kulturu, která se zaměří na efektivní využití dat.

Tato kniha vznikla po dlouhodobém pátrání a hledání vhodného nástroje pro zpracování dat. Při své práci jsme, tak jako určitě většina z vás, zpracovávali data a datové analýzy v MS Excel. Ač tento software považujeme za skvělý, efektivní a uživatelsky velice příjemný, začali jsme narážet na jeho limity. Problém vznikl se zpracováním velkých množství dat a práce s daty v reálném čase. Zároveň Excel nenabízí možnosti zpracovávat techniky strojového a hlubokého učení, nebo jen omezeně.

Vydali jsme se tak na cestu hledání vhodného nástroje, kdy jsme vyzkoušeli různé software a programovací jazyky (např.

Základy práce s KNIME Analytics Platform

Python, R) pro datové analýzy. Nejvíce nás oslovil nástroj KNIME Analytics Platform, jež nabízí obrovskou paletu možností práce s daty od základních datových analýz až po aplikace umělých neuronových sítí, a to vše v uživatelsky příjemném prostředí bez nutnosti programování.

Naším cílem je předat vám informace o:

- ☀️ Nástroji KNIME
- ☀️ Možnostech práce v KNIME prostředí
- ☀️ Způsobech, jak pracovat s daty
- ☀️ Nástrojích pro transformaci dat
- ☀️ Vizualizaci dat
- ☀️ Možnostech tvorby a použití základního modelu strojového učení

**Připravili jsme základního průvodce, kde
vám představujeme software**

KNIME Analytics Platform!



Základy práce s KNIME Analytics Platform

Veškeré podkladové soubory a vytvořená WorkFlow naleznete



na KnimeHub:

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

Postačí stáhnout soubory pomocí ikony na stahování.

K tomuto souboru se vrátíme později.

Icon	File/Folder Name	Download Icon
Folder icon	KNIME_training	Download icon (with 'Download workflow' tooltip)
Folder icon	KNIME_training_2	Download icon
Folder icon	KNIME_training_expanded	Download icon
Folder icon	Merge	Download icon
File icon	Russel_prices.csv	Download icon
File icon	adult.csv	Download icon
File icon	adult_new_person.xlsx	Download icon
File icon	relation_status.xlsx	Download icon

Základy práce s KNIME Analytics Platform

O KNIME ANALYTICS PLATFORM

Úvod do software

KNIME software je snadno dostupný (zdarma v desktopové verzi) a je jednou z často používaných platform pro datové analýzy, automatizaci dat, vizualizaci dat a tvorbu *Machine Learning* (strojové učení) a *Deep Learning* (hluboké učení pomocí neuronových sítí) ve velkých mezinárodních organizacích. Jeho výhoda tkví v jednoduchosti a možnosti tvořit modely bez nutnosti použití programovacího jazyka.

Software KNIME Analytics Platform je, jak již bylo řečeno, z kategorie freeware, tudíž jej můžete stáhnout a používat zcela zdarma. Byl vyvinut v akademickém prostředí ve Švýcarsku za účelem rychlé aplikace *Machine Learning* a *Deep Learning* metod a rychle našel aplikační využití i v komerční sféře. Jeho velkou výhodou je jeho jednoduchost, s níž můžete vytvořit model vkládáním a propojováním uzlů (pracovní balíčky ve formě ikon) do workflow v pracovním prostředí.

Proto, pokud je vám programování a kódování cizí, nemusíte si nutně zoufat, že analýzy dat a datová věda není pro vás. U každého z uzlů představujícího aktivitu máte možnost nastavení specifických parametrů dané aktivitě. KNIME zároveň nabízí značné množství ukázkových workflow, jež je možno použít a adaptovat na váš konkrétní úkol.

Základy práce s KNIME Analytics Platform

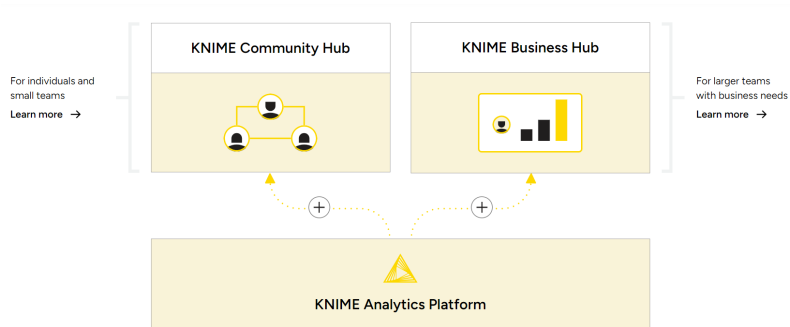
KNIME poskytuje dva produkty: KNIME Analytics Platform a KNIME Server, nově znám jako KNIME Business Hub.

Platforma KNIME Analytics Platform zahrnuje algoritmy pro datové operace potřebné pro projekty datových analýz, tedy od načtení dat, transformace dat, vizualizace až po tvorbu modelů strojového a hlubokého učení. Jak již bylo řečeno, jedná se o open-source software (k použití zdarma) pro všechny vaše potřeby práce s daty. Stažení z webu KNIME je zdarma (<https://www.knime.com/downloads>) a zdarma k použití. KNIME je založen na vizuálním řešení programování. Vizuální programování je klíčovou vlastností KNIME Analytics Platform a díky tomu je použití nástroje velmi snadné. Vizuální programování, grafické uživatelské rozhraní (GUI) vás provede všemi nezbytnými kroky k vytvoření workflow jednotlivých uzlů. KNIME byl navržen tak, aby byl otevřený různým datovým formátům, datovým typům, zdrojům dat, a dokonce i externím nástrojům jako je například Python a R.

Posledním krokem v datových analýzách je nasazení do produkce, tedy snadné, pohodlné a bezpečné nasazení. Tento proces přesunu aplikace do reálného světa se nazývá Deploy to production. V tomto případě je vhodné využít další možnosti KNIME, a to **KNIME Server**, nově známý jako **KNIME Business Hub**, který je dodáván s ročním licenčním poplatkem. KNIME Server nabízí prostředí chráněné pro

Základy práce s KNIME Analytics Platform

spolupráci v rámci firmy, pro automatizaci a umožňují přístup k workflow a interaktivním reportům přes webovou aplikaci.



Zdroj: <https://www.knime.com/software-overview>

Základy práce s KNIME Analytics Platform

Instalace

Pro instalaci KNIME Analytics Platform přejděte na <http://www.knime.com/downloads>, uveďte podrobnosti o sobě a stáhněte verzi vhodnou pro váš operační systém.

Download KNIME Analytics Platform

Get started in three simple steps.



New to the KNIME family? Let us help you get started with a short series of introductory emails. These will get you up and running as quickly as possible and introduce you to resources that will maximize your success with KNIME Analytics Platform.

Email *	Company/Organization
<input type="text"/>	<input type="text"/>
First Name	Last Name
<input type="text"/>	<input type="text"/>
Would you mind telling us how you heard about KNIME?	
<input type="text" value="I'd rather not tell"/>	
<input type="checkbox"/> I'd like to receive a short series of getting started emails from Emil.	
<input type="checkbox"/> I'd like to receive regular (but not too frequent) news updates from (only) KNIME.	
<input type="checkbox"/> I've read and accept the privacy policy . *	

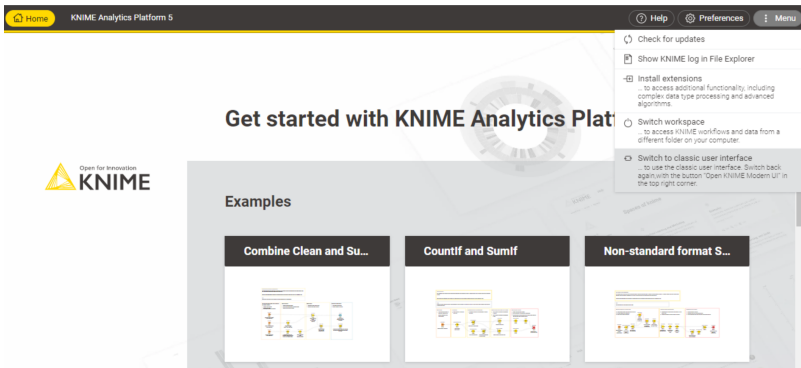
Next: Download KNIME

Zdroj: <https://www.knime.com/downloads>

Po stažení balíčku jej vyhledejte a postupujte podle pokynů. Nainstalujte jej do libovolného adresáře, pro který máte oprávnění k zápisu.

Základy práce s KNIME Analytics Platform

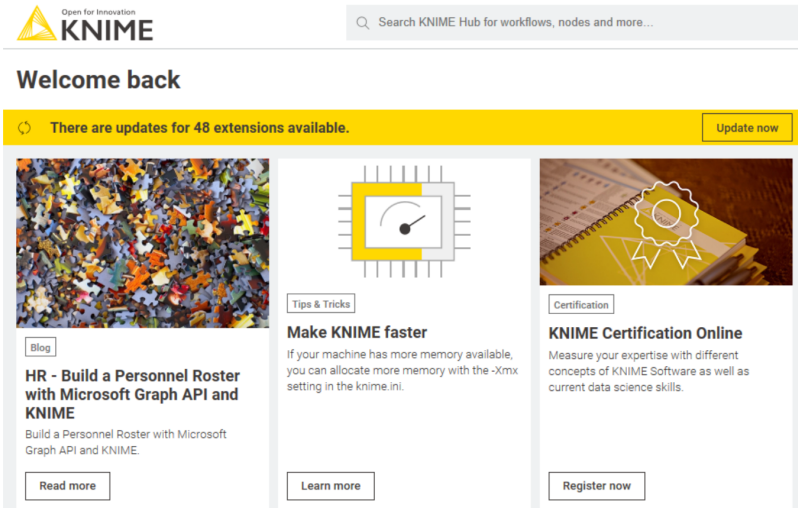
Pro naše účely, po prvním otevření KNIME Analytics Platform doporučuji přepnout si uživatelské prostředí na klasický interface, který umožňuje pokročilejší funkce. Nový UX design je sice designově lepší, ale my pracovat budeme ve standardním prostředí, do nějž se dostaneme pomocí Menu a zvolením „**Switch to classic user interface**“.



Důležité:

Z důvodu rychlého a dynamického vývoje datové vědy obecně a velkému počtu přispěvatelů do KNIME software, dochází často k aktualizaci verzí nebo doplňků. Aktualizace softwaru je možná přímo v platformě při jejím otevření.

Základy práce s KNIME Analytics Platform



Open for Innovation
KNIME

Search KNIME Hub for workflows, nodes and more...

Welcome back

There are updates for 48 extensions available. [Update now](#)

Blog

HR - Build a Personnel Roster with Microsoft Graph API and KNIME

Build a Personnel Roster with Microsoft Graph API and KNIME.

[Read more](#)

Tips & Tricks

Make KNIME faster

If your machine has more memory available, you can allocate more memory with the -Xmx setting in the knime.ini.

[Learn more](#)

Certification

KNIME Certification Online

Measure your expertise with different concepts of KNIME Software as well as current data science skills.

[Register now](#)

K čemu tedy KNIME slouží?

KNIME vám pomůže při:

- ☀ Načtení dat (např. z Excel, csv, tabulky, vytvořené tabulky v KNIME, z SQL databáze, textového editoru, MySQL, PostgreSQL databází, XML, JSON apod.)
- ☀ Transformaci dat (pomocí transformace s celými tabulkami, řádky, sloupci)
- ☀ Vizualizaci dat (grafy, statistiky)
- ☀ Tvorbě modelů Machine learning a Deep learning pro predikci dat

Základy práce s KNIME Analytics Platform

- ☀ Uložení výsledků (např. do Excel, csv, databází, Power BI Service apod.)
- ☀ Automatizaci procesů datové vědy a datových analýz
- ☀ Pro veřejné sdílení přes KNIME Hub
- ☀ Pro firemní sdílení a produkci modelů pomocí KNIME Server a KNIME WebPortal nově pod KNIME Business Hub

Nevíte-li si rady, tipy, triky a rady od KNIME expertů či KNIME pokročilých uživatelů naleznete na blogu <https://www.knime.com/blog>.

Práci v KNIME ulehčí existující šablony přímo v platformě, mnoho příkladů dostupných na KNIME Hub, jež je veřejný a zdarma pro všechny uživatele KNIME. KNIME pracuje s full-textovým vyhledáváním a tvorba workflow je takzvaně „na klikáčku“ a pomocí drag a drop uzlů.

A jak tedy KNIME vypadá? Pojdme si projít jednotlivé části platformy.

Důležité!

Z důvodů časté aktualizace jsme pro vás vydali toto 2.vydání, kde budou aktualizovány dodatečnými obrázky uzly, které prošly změnou, ukázkami z verzí KNIME 5x.

Tyto uzly poznáte dle zeleného ohraničení.

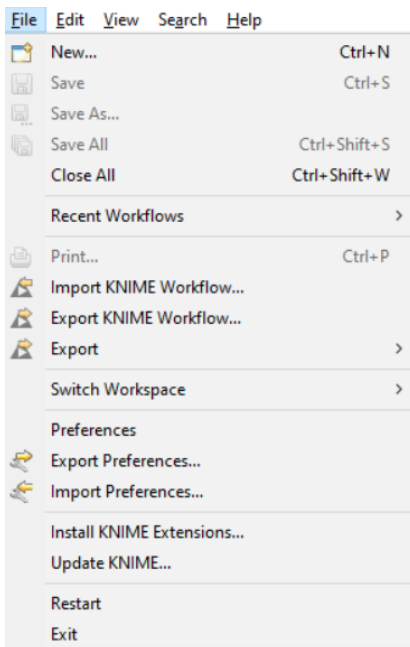
Základy práce s KNIME Analytics Platform

Možnosti z Menu

Menu pro práci v KNIME se nachází, stejně jako ve většině softwarů například od firmy Microsoft, v horní liště platformy.



První možností menu je nabídka práce se souborem v nabídce **File**, kde jej můžete uložit, otevřít, otevřít nový nebo importovat či exportovat již existující workflow.

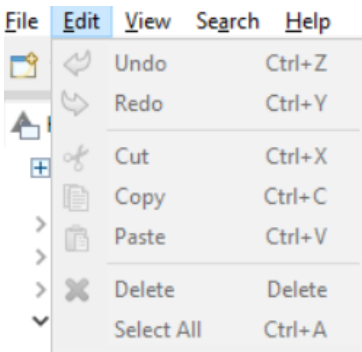


Základy práce s KNIME Analytics Platform

Workflow jsou ukládána jako soubor .knwf a celá skupina obsahující více workflow, nebo i datové zdroje, exportované modely a mnohé další, je pak ukládána pod typem .knwr.

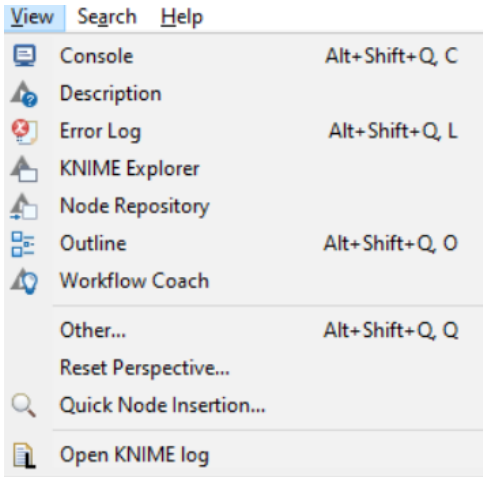
Zároveň toto menu nabízí instalaci doplňků v možnosti Install KNIME Extensions..., jež nejsou součástí základní verze platformy, ale je možno je zdarma doinstalovat, je-li potřeba. Například, potřebujete-li poslat zpracovaná data v KNIME na Power BI účet a navázat tak vizualizací v Power BI nástroji, postačí si nainstalovat Power BI extension (a samozřejmě je nutné mít k dispozici zmíněný Power BI účet od Microsoft – např. v rámci Microsoft Office 365 nebo zakoupený MS Power BI Pro účet) a data tak zaslat na tuto Power BI platformu. Doplňků je k dispozici celá řada z různých oblastí práce s daty a lze je vyhledat i přes full-textové vyhledávání.

Menu **Edit** umožňuje klasické editovací akce při práci v souboru, jako akce zpět, vpřed apod.



Základy práce s KNIME Analytics Platform

Záložka **View** obsahuje možnosti nastavení pohledu neboli zobrazení vaší platformy.



Je zde možné nastavit, co bude a co nebude obsahovat pohled vaší platformy. A k tomu se právě dostáváme.

Co tedy KNIME platforma nabízí?