

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ  
СІКОРСЬКОГО»

Факультет інформатики та обчислювальної техніки

Кафедра інформатики та програмної інженерії

Практикум №1

з курсу «Програмування інтелектуальних інформаційних систем»

на тему: «Моніторинг та дослідження платформи інтелектуального  
інформаційного аналізу KNIME»

Викладач:  
Курченко О.А.

Виконав:  
Хільчук А.В.  
студент 3 курсу  
групи ІІ-14 ФІОТ

Київ-2023

## Практична робота №1

**Тема:** Моніторинг та дослідження платформи інтелектуального інформаційного аналізу KNIME.

**Мета:** провести ґрунтовний аналіз платформи ПА KNIME

### **Виконання:**

#### 1. Характеристика платформи:

KNIME (Konstanz Information Miner) – це кроссплатформний open-source інструмент для аналізу даних та їх інтеграції, який надає можливість розробки та виконання комплексних аналітичних процесів. KNIME характеризується високою модулярністю та розширюваністю, що дозволяє користувачу налагоджувати його функціональність для вирішення задач різного характеру.

KNIME є інструментом візуального програмування, що дозволяє створювати робочі потоки (workflow) шляхом організації та з'єднання вузлів обробки та інтелектуального аналізу даних.

Розробник платформи - KNIME AG – Швейцарська компанія, що зосереджена на розробці ПЗ аналізу даних. Головний офіс KNIME AG розташований у Цюриху, однак вони також мають офіси у Берліні й Констанці в Німеччині, та Остіні в Америці.

#### 2. Можливості платформи.

Зокрема за рахунок своєї модулярності, KNIME надає можливість реалізовувати надзвичайно широкий спектр потреб інтелектуального аналізу даних.

Перш за все, KNIME має потужний вбудований інструментарій для реалізації ETL процесів. Він надає значний набір вузлів, що дозволяють зчитувати дані з різноманітних джерел, починаючи простими CSV файлами, та закінчуючи базами даних й REST API. Наявний і набір вузлів для власне очищення та трансформації даних: можна одразу видаляти рядки з відсутніми значеннями, формувати стрічки й інші типи даних, створювати нові стовпці й заповнювати їх даними з інших.

Також існує вбудований масив нодів для візуалізації даних різноманітними способами: стовпчикові діаграми, лінійні графіки,

графіки щільності розподілу даних, тощо. Багато вузлів підтримують різнобічну конфігурацію візуалізації, як-от відображення одразу кількох сутностей для різних значень параметрів певного стовпця набору даних.

Підтримується і, власне, хліб і сіль процесу інтелектуального аналізу даних – тренування та випробування моделей машинного навчання. KNIME містить вузли для всіх завдань інтелектуального аналізу даних: класифікації, кластеризації, асоціативного аналізу, регресії, моделі прогнозування для часових послідовностей, аналіз текстів та зображень.

Ну й урешті-решт, кінцевий етап “end-to-end” процесу аналізу даних – створення продукту для зовнішнього споживача. KNIME надає вузли інтегрованого розгортання, що дозволяють користувачу охоплювати частини workflow, необхідні для роботи у продакшені, саму модель або бібліотеку, а також підготовку даних. Ці захоплені підмножини автоматично зберігаються як робочі процеси з усіма відповідними параметрами та перетвореннями. Результатом розгортання може бути як REST API, так і Data app.

### **3. Переваги платформи**

#### **3.1.Знижений поріг входження**

Не завжди спеціаліст може мати досвід з кодуванням задля створення додатків за допомогою відповідних бібліотек мовою Python чи R. Особливо це стосується професій дослідників в областях медицини, нейрохімії, чи тому подібних, де головний фокус компетенції ставиться на розумінні самої галузі.

Тут такий інструмент, що дозволяє інтуїтивно створити додатки, оперуючи радше високорівневими коцепціями, і стає в нагоді.

Користувачу не потрібно розуміти деталі користування об’єктно-реляційними маперами або SQL для того, аби витягти масив інформації з БД- достатньо всього скористатися потрібним вузлом, налаштувати з’єднання, і все. Не потрібно мати розуміння специфіки конфігурації моделі, щоб вона запрацювала – достатньо просто взяти її вузол.

#### **3.2.Універсальність**

Логічною перевагою платформи, виходячи з її можливостей, є її універсальність. KNIME забезпечує інфраструктуру для:

- Зчитування даних
- Очищення даних
- Дослідження даних
- Створення дашбордів
- Тренування та випробування моделей
- Аналіз часових послідовностей
- Обробка природньої мови (NLP)
- Аналіз зображень
- Управління вхідними/вихідними потоками інформації

### **3.3.Високий ступінь налаштовуваності та розширюваності**

KNIME, попри таку високорівневість, вдається зберігати простір для налаштування багатьох вузлів. Моделям, наприклад, можна підкручувати широкий спектр гіперпараметрів. Не обділені й вузли обробки даних – вузлом обробки стрічок, наприклад, можна задавати окремим синтаксисом правила обробки стрічок за допомогою вкладених функцій.

Немає потрібного вузла? Біди теж немає – можна написати свій власний. Власні вузли можна описувати значною кількістю сучасних мов прогармування. Достатньо лиш установити бібліотеку Apache Maven, логіку вузла – й кінцевий результат готовий. Вузол потім можна буде експортувати на відповідний ресурс спільноти – KNIME hub.

### **3.4.Наявність потужної спільноти довкола платформи**

Звісно, KNIME користуються сотні тисяч експертів, що готові як допомогти тобі з твоїми питанням, так і поділитися своїми рішенням на форумах.

Однак всього лиш наявності таких людей недостатньо, щоб озвучувати ком'юніті як окрему перевагу. KNIME AG докладає особливих зусиль для підтримки даного ком'юніті, забезпечуючи окрему інфраструктуру для користувачів платформи.

Окрім забезпечення обширної документації та бібліотеки розширень, компанія створила окремий форум та платформу для контролю версій окремих workflow користувачів та їх публікації, а

також розповсюдження кастомних компонентів, розширень та вузлів користувачів.

#### **4. Недоліки платформи**

##### **4.1.Ціна візуального програмування**

Усі засоби візуального програмування рано чи пізно зіштовхуються з проблемами у керуванні й KNIME не є винятком.

Для невеликих проектів на рівні індивідуальних дослідників це не становитиме особливої проблеми, однак по мірі зростання проектів проблема управління процесами, а також відслідковування окремих елементів такого пайплайну, ставатиме все більшою, а з нею зростатиме й вартість утримання проекту, що може стати вирішальним фактором для багатьох стартапів малого та середнього масштабу.

##### **4.2.Незручності високорівневості**

Звісно, з грамотним підходом процес користування високорівневими рішеннями можна згладити, однак колізії уникнути фактично ніколи не можна.

За високорівневості завжди доведеться платити, і це аксіома, дійсна всюди.

##### **4.3.Становище платформи на ринку**

На превеликий жаль, платформа надзвичайно недопредставлена на IT-ринку.

З одного боку, її непопулярність частково можна пояснити нішовістю її цільової аудиторії – користувачі, яким потрібно проводити аналіз даних, але недоступне/непотрібне кодування.

З іншого ж боку, її недопредставленість доволі часто пояснюється тим, що її було затьмарено конкурентом – Alteryx. У онлайн дискусіях про KNIME дане ПЗ надзвичайно часто порівнюють, однак останнє має набагато більше популярність.

#### **5. Галузі застосування**

KNIME надає можливість створювати end-to-end продукт, отож, теоретично може застосовуватися в будь-якій індустрії, що покладається на аналіз даних. Однак задля встановлення конкретики можна звернути увагу на найбільших партнерів KNIME. Серед

компаній/кластерів у понад 10000 робітників даною платформою користуються:

- Atos – компанія IT-послуг, що обслуговує широкий спектр сервісів у сферах управління, охорони здоров'я, інформаційних технологій та телекомунікацій
- PwC – компанія, що надає професійні послуги у двох сегментах: довірчі рішення та консалтингові рішення. Індустрії, на яких фокусується компанія – охорона здоров'я, автоматизація, фінанси, енергетика та фармацевтика
- CGI – компанія, що надає IT та бізнес консалтингу в області фінансів та рітейлу
- Capgemini engineering – сукупність компаній, що надає інженерні та R&D послуги в областях банкінгу, фінансів, автоматизації та виробництва.
- LTI- глобальна компанія з надання консалтингу та цифрових рішень в галузі IT-технологій

Таким чином можна виділити наступні галузі застосування даної платформи:

- Фінансові послуги
- Рітейл
- Виробництво
- Енергетика
- Публічний сектор
- Маркетинг

На відносно ж індивідуальному рівні можна відмітити популярність у наступних галузях:

- Медицина
- Нейрохімія
- Когнітивна психологія
- Біоінформатика

## **Висновок**

Отож, у KNIME- це модулярна платформа інтелектуального аналізу даних засобами візуального програмування, що надає можливість створювати end-to-end рішення, покриваючи процес ETL, візуалізацію даних, тренування та застосування моделей, а також розгортання утворених робочих потоків для кінцевого користування. KNIME характеризується низьким порогом входження, що є особливо релевантно для ряду ніш, універсальністю, високою налаштовуваністю та розширюваністю, а також наявністю потужної спільноти довкола даного засобу, що зокрема забезпечується підтримкою відповідної інфраструктури самим розробником. Однак попри це варто пам'ятати, що KNIME, як і будь-який інший засіб візуального програмування, може страждати від проблем організації великих проектів, а високий ступінь високорівневості не завжди може гарантувати гладкий досвід користування ПЗ. Найбільшою ж проблемою KNIME є потужна недопредставленість на ринку, зумовлена зокрема наявністю сильного конкурента, що, безумовно, скасується на релевантності технології. На рівні великих корпорацій KNIME застосовується у галузях фінансів, ритейлу, енергетики, маркетингу публічного сектору. На рівнях індивідуальних дослідників або невеликих дослідницьких груп KNIME найчастіше застосовується у царинах: медицини, нейрохімії, поведінкової психології, біоінформатики.