

4. BACCM Overview – The Core Concepts that form Business Analysis – Core Concept Model. URL: <https://www.iiba.org/professional-development/knowledge-centre/ba-connection/baccm-overview---the-core-concepts-business-analysis-core-concept-model/> (дата звернення: 25.10.2024).

5. Ibekwe R. Business Analysis during Discovery Phase of Agile Software Delivery. URL: <https://www.linkedin.com/pulse/business-analysis-during-discovery-phase-agile-ralph-ibekwe-phd/> (дата звернення: 25.10.2024).

DOI <https://doi.org/10.30525/978-9934-26-506-8-113>

**USING ORANGE3 TO AUTOMATE DATA MINING PROCESSES  
IN THE MODE OF VISUAL CONSTRUCTION  
OF DATA TRANSFORMATION STAGES**

**ВИКОРИСТАННЯ ORANGE3 ДЛЯ АВТОМАТИЗАЦІЇ ПРОЦЕСІВ  
ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ В РЕЖИМІ  
ВІЗУАЛЬНОЇ ПОБУДОВИ ЕТАПІВ ПЕРЕТВОРЕННЯ ДАНИХ**

**Nefedchenko O.O.,**  
*Student (group 122-22-1p),  
LLC “Technical university  
“Metinvest polytechnic”,  
Zaporizhzhia, Ukraine*

**Нефедченко О.О.,**  
*студент гр. 122-22-1п,  
ТОВ «Технічний університет  
«Метінвест політехніка»,  
м. Запоріжжя, Україна*

**Sahaida P.I.,**  
*DSc (Engineering),  
Associate Professor, LLC “Technical  
university “Metinvest polytechnic”,  
Zaporizhzhia, Ukraine*

**Сагайда П.І.,**  
*д.т.н., доцент,  
ТОВ «Технічний університет  
«Метінвест політехніка»,  
м. Запоріжжя, Україна*

На сучасних підприємствах та в організаціях накопичені корпоративні дані мають великі розміри та стають все більш доступними для фахівців з їх обробки та витягу з них моделей. Відповідно, зростають вимоги до рівня підготовки спеціалістів з Data Science. Якісний розвідувальний аналіз даних (exploratory data analysis) потребує володіння декількома мовами або засобами програмування для розв'язання задач статистичної обробки даних, візуалізації результатів, підготовки звітів. Між тим коло працівників-аналітиків розширюється,

необхідність використання оброблених, з метою підтримки прийняття рішень, даних збільшується. Це потребує підготовки фахівців, які можуть проводити дослідження даних швидко та без необхідності написання і налагодження великих програм, і одночасно забезпечувати якість і надійність результатів проведених досліджень. Це, в свою чергу, вимагає знань і навичок застосування багатьох алгоритмів обробки, особливостей і режимів їх використання. Тому наразі виникає гостра потреба в інструментах, які дозволяють ефективно аналізувати дані та витягувати з них корисну інформацію, з можливістю довідкової та інформаційної підтримки такої діяльності, можливістю автоматизації рутинних активностей.

Одним з таких інструментів є Orange3. Цей візуальний програмний пакет призначений для автоматизації багатьох етапів машинного навчання, крім того, він дає можливість навіть початківцям проводити складні аналізи даних без необхідності написання великої кількості коду. На відміну від аналогічних за функціональністю інструментів від Azure Machine Learning та Google Vertex AI, він не вимагає платного акаунту та суттєвої підготовки в галузі інформаційних технологій.

Orange3 – це вільно розповсюджуваний програмний пакет з відкритим кодом, який використовується для візуального програмування в області машинного навчання та data mining. Він надає інтуїтивно зрозумілий інтерфейс, що дозволяє будувати складні аналітичні pipelines (послідовності операцій обробки даних) шляхом візуального конструювання: перетягування та з'єднання візуальних блоків (віджетів), кожен з яких відповідає за окремий етап перетворення або трансформації даних.

Перевагами використання Orange3 для автоматизації процесу обробки даних та витягу з них моделей є:

- візуальний інтерфейс, простий у використанні, навіть для тих, хто не має глибоких знань програмування;
- підтримка широкого спектру алгоритмів машинного навчання, від класифікації та регресії до кластеризації та асоціативного аналізу.
- можливість інтеграції з іншими пакетами та бібліотеками, за рахунок чого аналітик може працювати з даними з різних джерел, включаючи файли CSV, бази даних, і легко інтегрується з іншими бібліотеками Python (NumPy, SciPy, scikit-learn).
- велика та досить активна спільнота користувачів і розробників, що забезпечує постійний розвиток пакету та підтримку аналітиків-початківців.

– доступність початкового коду програмного комплексу, що дозволяє глибоко зануритися в роботу пакету та адаптувати його під свої потреби.

Приклад використання Orange3. Припустимо, у нас є набір даних про клієнтів підприємства, і ми хочемо передбачити, чи клієнт відмовиться від отримання послуг або товарів. За допомогою Orange3 можна:

- завантажити дані про клієнтів (їх ознаки та результати співробітництва) та в програму;
- виконати попередню обробку даних, видаливши пропущені значення та закодувавши категоріальні змінні;
- обрати алгоритм класифікації, наприклад, логістичну регресію;
- навчити модель на навчальних даних;
- оцінити точність моделі на тестових даних;
- візуалізувати результати класифікації, наприклад, побудувати received operation curve (ROC-криву), яка дозволить оцінити якість навченої моделі та можливість її використання для прогнозів.

Всі ці операції та активності не потребують багато часу, дозволяють швидко освоєння користувальницького інтерфейсу, та дають можливість у короткий термін провести багатоваріантний інтелектуальний аналіз даних.

**Висновки.** Orange3 – це потужний і зручний інструмент для аналізу даних, який дозволяє автоматизувати багато рутинних задач і зосередитися на інтерпретації результатів. Він особливо корисний для тих, хто хоче швидко отримати перші результати аналізу без глибоких знань програмування. Однак, для більш складних задач і пристосування до особливостей та вимог конкретного підприємства або організації може знадобитися додаткове програмування на Python. Для розв'язання складних задач, наприклад, проведення глибокого навчання, потрібно виконувати інтеграцію Orange3 з фреймворками глибокого навчання, такими як TensorFlow або PyTorch, а також комбінування Orange3 з іншими інструментами для аналізу даних, такими як Jupyter Notebook, PyCharm.