

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ "ЛЬВІВСЬКА ПОЛІТЕХНІКА"**

**Інститут КНІТ
Кафедра ПЗ**

ЗВІТ

До лабораторної роботи № 1

З дисципліни: *“Видобування та опрацювання даних”*

На тему: *“Етап розуміння бізнесу за CRISP-DM”*

Лектор:

асист. каф. ПЗ
Угриновський Б.В.

Виконав:

ст. гр. ПЗ-45
Хруставчук М.Л.

Прийняв:

асист. каф. ПЗ
Симець І.І.

« ____ » _____ 2026 р.

Σ= _____

Львів – 2026

Тема роботи: Етап розуміння бізнесу за CRISP-DM.

Мета роботи: Вибрати набір даних та обґрунтувати актуальність його дослідження з точки зору бізнесу та науки про дані.

ТЕОРЕТИЧНІ ВІДОМОСТІ

Добування даних (Data Mining) – це процес виявлення прихованих закономірностей, залежностей та структур у великих масивах інформації з використанням методів статистики, машинного навчання та аналізу баз даних. Основною метою добування даних є отримання корисних знань, які можуть бути використані для прийняття обґрунтованих рішень, прогнозування майбутніх подій або оптимізації процесів. У сучасних умовах обробка великих обсягів інформації є невід’ємною частиною бізнесу, науки, медицини, фінансів та спортивної аналітики.

Однією з найпоширеніших методологій реалізації проєктів з аналізу даних є CRISP-DM (Cross-Industry Standard Process for Data Mining). Дана методологія визначає послідовність етапів роботи з даними та забезпечує структурований підхід до розв’язання задач. Вона складається з шести основних фаз: розуміння бізнесу, розуміння даних, підготовка даних, моделювання, оцінювання та розгортання. Кожен етап має чітко визначену мету та забезпечує логічний перехід до наступного кроку.

Перший етап – розуміння бізнесу – є критично важливим, оскільки саме на цьому етапі формулюється проблема, визначаються цілі дослідження та критерії успіху. Неправильне трактування бізнес-потреб може призвести до побудови технічно коректної, але практично непотрібної моделі. У межах даного етапу необхідно проаналізувати предметну область, сформулювати гіпотези, оцінити доступність даних та визначити потенційну бізнес-цінність майбутнього аналізу.

Таким чином, етап розуміння бізнесу закладає основу для подальших аналітичних дій та визначає напрям дослідження.

ЗАВДАННЯ

1. Який набір даних обрано? Коротко опишіть ПО.
2. Які ідеї чи гіпотези можна перевірити або покращити за допомогою аналізу даних?
3. Які проблеми та «вузькі місця» існують у цій області? Що можна автоматизувати?
4. Які можливі ризики чи труднощі під час роботи з даними?
5. Який метод (кластеризація, прогнозування, класифікація) підходить для вирішення задачі?

ХІД ВИКОНАННЯ

1. Який набір даних обрано? Коротко опишіть ПО.

Для виконання лабораторної роботи було обрано набір даних [*FIFA 23 Complete Player Dataset*](#). Даний датасет містить детальну інформацію про професійних футболістів, включаючи їхній загальний рейтинг (overall), потенціал розвитку (potential), вік, фізичні параметри, позицію на полі, технічні характеристики, ринкову вартість та заробітну плату.

Предметна область – спортивна аналітика та футбольний менеджмент. Дані відображають реальні характеристики гравців і можуть використовуватись для оцінки їх ефективності, планування трансферів та формування складу команди. Таким чином, набір даних має практичну бізнес-цінність у сфері управління футбольним клубом.

2. Які ідеї чи гіпотези можна перевірити або покращити за допомогою аналізу даних?

Основна гіпотеза полягає в тому, що трансферна вартість гравця залежить від його рейтингу, віку та технічних характеристик. Можна перевірити, які саме

показники мають найбільший вплив на `value_eur` та чи існує переоцінка або недооцінка певних гравців.

Також можна дослідити відповідність гравця його позиції. На основі технічних і фізичних характеристик можливо визначити, чи оптимально обрана позиція, або яка позиція є найбільш підходящою для конкретного профілю.

Таким чином, датасет дозволяє реалізувати як прогностичні, так і аналітичні задачі.

3. Які проблеми та «вузькі місця» існують у цій області? Що можна автоматизувати?

У футбольному менеджменті складно об'єктивно оцінити вартість і перспективність гравця. Трансферні рішення часто базуються на експертній думці, що містить суб'єктивність і ризик переплати.

Основне вузьке місце полягає в необхідності аналізу великої кількості характеристик одночасно. Людина не може ефективно врахувати всі параметри та їх взаємозв'язки без аналітичних інструментів.

Автоматизувати можна:

- Прогнозування трансферної вартості гравців.
- Оцінювання відповідності гравця певній позиції.
- Порівняння кандидатів для підсилення складу.

Застосування методів машинного навчання дозволяє зменшити фінансові ризики та підвищити ефективність трансферної політики.

4. Які можливі ризики чи труднощі під час роботи з даними?

Датасет містить велику кількість ознак, тому необхідний відбір релевантних характеристик. Можливі пропущені значення або некоректні типи даних, що потребують попередньої обробки.

Існує ймовірність високої кореляції між характеристиками гравців, що може впливати на стабільність моделей. Трансферна вартість може мати нерівномірний

розподіл, що ускладнює застосування деяких алгоритмів без додаткового перетворення даних.

Також рейтинги FIFA формуються за власною методологією, тому вони можуть не повністю відображати реальну спортивну ефективність.

5. Який метод (кластеризація, прогнозування, класифікація) підходить для

У межах даного набору даних доцільно застосувати два підходи.

Перший – регресія для прогнозування трансферної вартості (`value_eur`) на основі характеристик гравця. Оскільки `value_eur` є числовою неперервною змінною, регресійні моделі дозволяють оцінити очікувану ринкову ціну та визначити можливу переоціненість або недооціненість футболіста.

Другий – класифікація для визначення найбільш підходящої позиції гравця на основі його технічних і фізичних характеристик. У цьому випадку цільовою змінною є категорія (позиція), а модель дозволяє оцінити відповідність гравця його ролі на полі або запропонувати альтернативну позицію.

Таким чином, регресія використовується для фінансового прогнозування, а класифікація – для аналізу спортивної ефективності та оптимального розподілу ролей.

ВИСНОВКИ

У ході виконання лабораторної роботи було обрано та обґрунтовано набір даних у сфері спортивної аналітики, що має практичну бізнес-цінність. Було проаналізовано предметну область, сформульовано гіпотези та визначено можливі напрями застосування методів машинного навчання. Етап розуміння бізнесу дозволив чітко визначити цілі подальшого аналізу та закласти основу для наступних етапів CRISP-DM.