**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:** *“Розуміння задачі і області застосування”*

**Лектор:**

викладач каф. ПЗ

Матвіїв Б. Д.

**Виконав:**

ст. гр. ПЗ-44

Ройк О. В.

**Прийняв:**

асист. каф. ПЗ

Угриновський Б. В.

« \_\_\_\_ » \_\_\_\_\_\_\_\_ 2020 р.

∑= \_\_\_\_\_ .

Львів – 2020

**Тема роботи:** Розуміння задачі і області застосування.

**Мета роботи:** Обрати та мотивувати обраний набір даних.

**ЗАВДАННЯ**

1. Обрати та обґрунтувати набір даних. Набір повинен відповідати таким умовам:
   * Табличне представлення даних (не зображення або часовий ряд);
   * Мінімум 250 записів;
   * Мінімум 4 колонки з даними за якими можна буде виконувати аналіз, моделювання і візуалізацію (колонки ID, номер або службові не рахуються);
   * Один варіант набору даних на групу.
2. Відповісти на такі питання:
   * Який набір даних ви обрали?
   * Які у вас ідеї, теорії, припущення що у вибраному домені можна покращити з аналізом та видобуванням даних?
   * Які проблеми є у області з якої ви взяли дані? Що є "bottleneck" у даному напрямі, що можна автоматизувати?
   * Що за ризики та складності ви припускаєте можуть очікувати вас в процесі роботи з даними?
   * В рамках добування даних яким методом ваша проблема або ідея може бути вирішена?

**ХІД ВИКОНАННЯ**

**Який набір даних ви обрали*?***

Для виконання лабораторних робіт було обрано набір даних «Steam Store Games», Цей датасет містить інформацію про цифрову платформу для продажу відеоігор Steam. Даними є набір ігор розташованих на цій платформі з інформацією про розробників, видавців, платформами, жанрами, категоріями, тегами, ціною, відгуках та кількістю продаж на платформі.

**Які у вас ідеї, теорії, припущення що у вибраному домені можна покращити з аналізом та видобуванням даних?**

За допомогою аналізу обраного набору даних можна проаналізувати популярність тих чи інших жанрів відеоігор на ринку, отримати список конкурентів їх популярність, оцінку покупців, ціни. Побачити популярні платформи для тих чи інших жанрів, що може допомогти молодим розробника оцінити ризики нового проекту.

**Які проблеми є у області з якої ви взяли дані? Що є "bottleneck" у даному напрямі, що можна автоматизувати?**

Проблемою в даній області є необізнаність молодих розробників в дійсних трендах ігрового ринку, важкість впри виборі вигідної ціни, що здатна вистояти проти конкурентів, неправильний вибір платформ. Дані проблеми можна проаналізувати за допомогою обраного набору даних.

**Що за ризики та складності ви припускаєте можуть очікувати вас в процесі роботи з даними?**

В наборі використані дані отримані за допомогою Steam-Spy API, сервісу не від розробників Steam. Через політику компанії Valve цей сервіс був змушений відмовитись від точних методів отримання даних в 2019 році, і зараз за словами розробників використовує замість цього певні алгоритми, які на жаль не є настільки ж точними, тому інформація є актуальною для ігор, що вийшли до цього моменту, інша ж інформація може бути неактуальною.

**В рамках добування даних яким методом ваша проблема або ідея може бути вирішена?**

Дана задача може бути вирішена методом прогнозування. Можна можемо спрогнозувати, чи буде хорошою ідеєю розробляти гру певного жанру в даний момент, також визначити найвигіднішу цінову політику на основі конкурентів. Також можна використати методи кластеризації, щоб з’ясувати які фактори впливають на успішність гри і які навпаки заважають.

**ВИСНОВКИ**

Отже, виконуючи дану лабораторну роботу, я проаналізував набір даних «Steam Store Games», визначив основні ідеї, теорії, припущення, проблеми, ризики та складності щодо аналізу даного набору даних, обрав методи, яким дана задача може бути вирішена.