

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук  
(повна назва)

Кафедра Системотехніки  
(повна назва)

**КВАЛІФІКАЦІЙНА РОБОТА**  
**Пояснювальна записка**

Рівень вищої освіти другий (магістерський)

Дослідження застосування методів інтелектуального аналізу даних  
в системах фріланс-бірж для 2D/3D Generalist  
(тема)

Виконав: студент 2 курсу, групи ІТІМ-22-1  
Різник А.О.  
(прізвище, ініціали)

Спеціальність 122 Комп'ютерні науки  
(код і повна назва спеціальності)

Тип програми освітньо-професійна  
(освітньо-професійна або освітньо-наукова)

Освітня програма ОПП Інформаційні  
технології проектування  
(повна назва освітньої програми)

Керівник доцент Ребезюк Л.М.  
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри

\_\_\_\_\_  
(підпис)

Гребеннік І.В.  
(прізвище, ініціали)

2024 р.



Харківський національний університет радіоелектроніки

(назва закладу вищої освіти)

Факультет Комп'ютерних наук

Кафедра Системотехніки

Рівень вищої освіти другий (магістерський)

Спеціальність 122 Комп'ютерні науки  
(код і повна назва)

Тип програми освітньо-професійна

Освітня програма ОПП Інформаційні технології проектування  
(повна назва освітньої програми)

**ЗАТВЕРДЖУЮ:**

Зав. кафедри СТ

проф. Гребеннік І.В.

" \_\_\_\_\_ " \_\_\_\_\_ 2023 р.

## **ЗАВДАННЯ**

### **НА КВАЛІФІКАЦІЙНУ РОБОТУ**

студентові Різнику Артему Олександровичу  
(прізвище, ім'я, по батькові)

Тема роботи Дослідження застосування методів інтелектуального аналізу даних в системах фриланс-бірж для 2D/3D Generalist

затверджена наказом по університету від " 20 " листопада **2023р.** № 1373Ст

**2. Термін подання студентом роботи** 17 січня 2024 р.

**3. Вихідні дані до роботи.** дослідження застосування методів інтелектуального аналізу даних у системах фриланс-бірж для 2D/3D Generalist з метою об'єктивного встановлення вартості поточних робіт цифрових митців. Перелік використовуваних програмних засобів: ОС Microsoft Windows v.7 або вище. Технічнезабезпечення: IBM-сумісний ПК з МП Pentium II або вище.

**4. Перелік питань, що потрібно опрацювати в роботі** \_\_\_\_\_

4.1 Вступ. 4.2 Аналіз предметної області. 4.2.1 Загальна характеристика діяльності сервісів, що визначається предметною областю 4.2.2 Аналіз існуючих ІС фриланс-бірж для митців. 4.2.3 Постановка задач дослідження 4.3 Дослідження методів інтелектуального аналізу даних та застосування їх в системах фриланс-бірж 4.3.1 KDD та модель інтелектуального аналізу даних. 4.3.2 Інтелектуальні методи отримання інформації для аналізу 4.3.3 Методи інтелектуального аналізу даних 4.4 Структурне проектування програмного фільтру для визначення вартості виконання проекту фрилансером 4.4.1 Кластерний аналіз в контексті задачі прогнозування вартості виконання проекту фрилансером 4.4.2 Метод асоціативних правил для розрахунку виконання проекту фрилансером 4.5 Програмна реалізація 4.5.1 Вибір мови програмування 4.5.2 Програмна розробка фільтру для визначення рекомендацій вартості роботи фрилансера 4.6 Висновки

**5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (слайдів) 5.1 Об'єкт, предмет дослідження, мета роботи (слайд 1). 5.2 Постановка задач дослідження та розробки (слайд 2,3). 5.3 Аналіз предметної області. Аналіз існуючих систем-аналогів (слайд 4). 5.4 Діаграма прецедентів (слайд 5) 5.5 Опис архітектури ІС фріланс-біржі (слайд 6). 5.6 Популярні методи інтелектуального аналізу даних (слайд 7). 5.7 Приклад застосування методу кластеризації (слайд 8). 5.8 Приклад застосування методу асоціативних правил (слайд 9). 5.9 Приклади типів робіт для замовлень із 2D графікою (слайд 10). 5.10 Приклади типів робіт для замовлень із 3D графікою (слайд 11). 5.11 Приклади типів робіт для замовлень із Відео (слайд 12) 5.12 Асоціативні правила діапазону вартості проекту для роботи із 3D графікою (слайд 13) 5.13 Принцип роботи фільтра (слайд 14,15) 5.14 Програмний фільтр для визначення рекомендаційої вартості роботи фрілансера (слайд 16,17) 5.15 Висновки (слайд 18)**

**6. Дата видачі завдання** 20 листопада 2023 р.

### КАЛЕНДАРНИЙ ПЛАН

Пор. №	Назва етапів роботи	Термін виконання етапів Роботи	Примітка
1.	Отримання завдання на виконання кваліфікаційної роботи.	20.11.23	
2.	Аналіз завдання та предметної області щодо діяльністю фріланс-біржі	21.11 – 23.11.23	
3.	Аналіз існуючих систем-аналогів та постановка задач задачі дослідження	24.11 – 30.11.23	
4.	Дослідження методів інтелектуального аналізу даних та застосування їх в системах фриланс-бірж	01.12 – 08.12.23	
5.	Кластерний аналіз в контексті задачі прогнозування вартості виконання проекту фрілансером	09.12 – 15.12.23	
6.	Метод асоціативних правил для розрахунку вартості проекту	16.12 – 24.12.23	
7.	Програмна розробка фільтру для визначення рекомендацій вартості роботи фрілансера	25.12 – 08.01.24	
8.	Оформлення пояснювальної записки та документації та презентаційних матеріалів комп'ютерного захисту	09.01 – 16.01.24	
9.	Представлення роботи на рецензування	17.01.24	

Студент



(підпис)

Різник А.О.

Керівник роботи



(підпис)

доц. Ребезюк Л.М.

## ABSTRACT

Master's thesis: 89 p., 8 tabl., 20 fig., 1 app., 13 sources.

FREELANCE EXCHANGE, 2D/3D GENERALIST, METHODS OF INTELLIGENT DATA ANALYSIS, COST OF CURRENT WORK, CLUSTERIZATION

The object of research – data analysis in freelance exchange systems for 2D/3D Generalist.

The subject of the study – methods of intellectual data analysis in freelance exchange systems for 2D/3D Generalist with the aim of objectively establishing the value of current works of digital artists.

The purpose of the work – to study the application of intelligent data analysis methods in freelance exchange systems for 2D/3D Generalist in order to objectively establish the value of current works of digital artists.

Research methods – systematic approach, methods of functional analysis, methods of intellectual data analysis, method of associative rules.

To solve the task, an analysis of the subject area was carried out, on the basis of which the requirements for a modern freelance exchange were determined, and an analysis of existing freelance exchanges was carried out and their shortcomings were identified. As part of the study of methods of intelligent data analysis and their application in freelance exchange systems, KDD and the model of intelligent data analysis, intelligent methods of obtaining information for analysis, methods of intelligent data analysis were analyzed. To determine the cost of executing the current project by a freelancer, a cluster analysis was performed in the context of the task of forecasting the cost of executing a current project by a freelancer, and the method of associative rules for calculating the cost of executing a project by a freelancer was considered. A software filter was developed to determine the recommended value of a freelancer's current work.

## РЕФЕРАТ

Кваліфікаційна робота: 89 стор., 8 табл., 20 рис., 1 дод., 13 джерел.

ФРИЛАНС-БІРЖА, 2D/3D GENERALIST, МЕТОДИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ, ВАРТІСТЬ ПОТОЧНОЇ РОБОТИ, КЛАСТЕРІЗАЦІЯ

Об'єкт дослідження – аналіз даних в системах фриланс-бірж для 2D/3D Generalist.

Предмет дослідження – методи інтелектуального аналізу даних в системах фриланс-бірж для 2D/3D Generalist з метою об'єктивного встановлення вартості поточних робіт цифрових митців.

Мета роботи – дослідження застосування методів інтелектуального аналізу даних у системах фриланс-бірж для 2D/3D Generalist з метою об'єктивного встановлення вартості поточних робіт цифрових митців.

Методи дослідження – системний підхід, методи функціонального аналізу, методи інтелектуального аналізу даних, метод асоціативних правил.

Для вирішення поставленої задачі проведено аналіз предметної області, на основі якого були визначені вимоги до сучасної біржі фрилансу та проведено аналіз існуючих фриланс-бірж та виявлені їх недоліки. В рамках дослідження методів інтелектуального аналізу даних та застосування їх в системах фриланс-бірж проаналізовані KDD та модель інтелектуального аналізу даних, інтелектуальні методи отримання інформації для аналізу, методи інтелектуального аналізу даних. Для визначення вартості виконання поточного проекту фрилансером проведено кластерний аналіз в контексті задачі прогнозування вартості виконання поточного проекту фрилансером та розглянуто метод асоціативних правил для розрахунку вартості виконання проекту фрилансером. Розроблено програмний фільтру для визначення рекомендаційної вартості поточної роботи фрилансера.

## ЗМІСТ

Скорочення та умовні позначки .....	6
Вступ .....	7
1 Аналіз предметної області .....	9
1.1 Загальна характеристика діяльності сервісів, що визначається предметною областю.....	10
1.2 Аналіз існуючих ІС фриланс-бірж для митців.....	12
1.3 Постановка задачі дослідження.....	26
2 Дослідження методів інтелектуального аналізу даних та застосування їх в системах фриланс-бірж.....	29
2.1 KDD та модель інтелектуального аналізу даних .....	32
2.2 Інтелектуальні методи отримання інформації для аналізу .....	36
2.3 Методи інтелектуального аналізу даних.....	33
3 Структурне проектування програмного фільтру для визначення вартості виконання проекту фрилансером.....	48
3.1 Кластерний аналіз в контексті задачі прогнозування вартості виконання проекту фрилансером.....	48
3.2 Метод асоціативних правил для розрахунку вартості виконання проекту фрилансером .....	60
4 Програмна реалізація.....	69
4.1 Вибір мови програмування.....	69
4.2 Програмна розробка фільтру для визначення рекомендацій вартості роботи фрилансера.....	71
Висновки .....	78
Перелік джерел посилання .....	79
Додаток А Графічні матеріали кваліфікаційної роботи .....	80

## СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ

- ICN – International Council of Nurses;  
KDD – Knowledge Discovery in Databases;  
NLP – Natural Language Processing,
- БД – база даних;  
ІС – інформаційна система;  
СУБД – система управління базами даних;

## ВСТУП

Фриланс-біржа – це сервіс для взаємодії між замовниками (власниками сайтів, юридичними особами, вебстудіями) та фрилансерами (фахівцями у галузях IT, SEO, SMM, маркетингу, дизайну тощо), які надають послуги віддалено.

Спочатку через обставини, такі як карантин, а зараз і через воєнні події, збільшилася кількість українців, які виявляють інтерес до роботи на фрилансі. Відповідно до аналізу сервісу Freelancehunt, порівняно з літом 2021 року, попит на послуги фрилансерів в Україні літом 2022 року зрос на 30%.

Фрилансер – це самостійний фахівець, який самостійно знаходить клієнтів та може працювати над кількома проєктами одночасно. Часто терміни "фрилансер" та "віддалений співробітник" використовуються як взаємозамінні, але це неправильне уявлення. У відмінну від фрилансера, віддалений співробітник є найманим працівником однієї компанії та працює за фіксованим графіком. Фрилансер співпрацює з різними клієнтами та встановлює графік роботи самостійно.

В сучасному світі багато людей обирають гнучкий графік роботи, який відкриває безліч можливостей. Фриланс, у свою чергу, став повноцінним джерелом заробітку. Використання фриланс-бірж як інструменту для пошуку та просування своїх послуг є дуже зручним рішенням для вирішення проблеми гарантій виконання умов угоди як для виконавців, так і для роботодавців. Біржі вирішують завдання пошуку роботодавця або реклами послуг, пропонуючи простий процес реєстрації та публікації портфоліо робіт (для виконавців) або оголошення про роботу (для фрилансерів). Це можна зробити з будь-якого пристрою, такого як ноутбук, смартфон чи електронний планшет, що дозволяє вам бути активними де завгодно. У світі мережі Інтернет існує безліч фриланс-бірж, кожна з яких має свої особливості, наприклад, можливість створення різних типів акаунтів, оцінки виконаних контрактів, сортування за рейтингом і так далі.

В процесі виконання кваліфікаційної роботи здобувача ступеня «магістр» було виконано дослідження застосування методів інтелектуального аналізу даних для вирішення проблеми демпінгу ціноутворення замовлень для цифрових митців, за допомогою якого розроблені компоненти інформаційної системи вузкоспеціалізованої фриланс-біржі, які дозволяють підвищити ефективність роботи і чітко незалежне формування цін на виконання замовлення.

Для реалізації системи були використані наступні технології: MySQL – як сервер для баз даних це вільна система керування реляційними базами даних. HTML – розмітка окремих частин веб додатку. CSS – оформлення розмітки (анімації, тіні, переходи і т.д.). JavaScript – мова програмування, що дозволяє реалізувати ряд складних рішень в web-документах.

Під час аналізу більшості сучасних онлайн фриланс-бірж виявлено, що більшість сервісів не надають можливості для конструктивної та об'єктивної оцінки вартості роботи. Зазвичай ця вартість визначається відповідно до бюджету роботодавця або складності та важливості завдання для нього. Таким чином, основною метою цього дослідження є вивчення застосування методів інтелектуального аналізу даних у системах фриланс-бірж для 2D/3D Generalist з метою об'єктивного встановлення вартості робіт цифрових митців.

## 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

Фриланс-біржа представляє собою інтернет-платформу, яка надає зручний механізм взаємодії між роботодавцями та фахівцями, уникнувши особистих зустрічей. Цей формат співпраці дозволяє обговорювати всі деталі проекту та оплачувати виконану роботу навіть в разі географічної віддаленості сторін. Термін "фрилансер" виник у зв'язку із вільними найманцями військового періоду. У сучасному розумінні це вільний фахівець, який обрав роботу віддалено замість офісного приміщення, і не обов'язково має офіційну освіту, головне - має відповідні навички для виконання завдань.

Weblancer, як фриланс-біржа, відіграє ключову роль в забезпеченні можливості здобувачам пропонувати свої послуги, а роботодавцям - знаходити відповідних виконавців. На платформі зібрано великий каталог вакансій та портфоліо фахівців з України та інших країн. Роботодавці можуть розміщувати свої вакансії або конкретні проекти, вибираючи виконавців для вирішення завдань. Робота на фриланс-біржах для фахівців відкриває можливість не лише для одноразових проектів, а й для довгострокових перспектив. Weblancer є ідеальним майданчиком для знайомства із потенційними роботодавцями, особливо для новачків, які тільки починають будувати свою клієнтську базу. Для фахівців, які тільки розпочинають свій шлях віддалено, ця платформа є надзвичайно цінним ресурсом для професійного росту без значних витрат, володіючи рядом переваг.

Однією з ключових переваг участі на фриланс-біржі є забезпечення фінансової безпеки. Заключаючи угоду в рамках Безпечного правочину, обидві сторони мають гарантії щодо виконання завдання та своєчасної оплати. Виведення коштів з платформи здійснюється швидко та без зайвих труднощів, існують підтримка найпопулярніших платіжних систем.

Додатково, завдяки рейтингам, легко оцінити професіоналізм виконавців, оглядаючи їхні портфоліо, робочий досвід та наявність рекомендацій. Аспіранти, в свою чергу, можуть долучити до репутації роботодавця та

визначити, чи варто надійти на його замовлення.

Технічна підтримка, яка відіграє важливу роль для будь-якого веб-ресурсу, на Weblancer швидко реагує на запити користувачів, ефективно вирішуючи виниклі труднощі або консультуючи з питань функціонування порталу.

Ще однією приємною можливістю є форум, де завжди можна звернутися за підтримкою від більш досвідчених колег. Крім того, цікаву та корисну інформацію про віддалену роботу можна знайти в блозі, який регулярно оновлюється новими статтями.

Таблиця 1.1 – Порівняльна таблиця вигод та переваг

№ н\п	Дія	На фриланс-бірже	Поза біржею
1.	Пошук замовлень	Замовлення та вакансії на одному сайті згруповані за темами	За оголошеннями в інтернеті, за допомогою холодних дзвінків
2.	Оплата праці	Гарантована оплата за використання безпечної угоди. Можливість переконатися в надійності замовника завдяки відгукам	Працюючи без передоплати, є ризик зіткнутися з несумлінними наймачами
3.	Підвищення репутації	З кількістю успішно виконаних завдань та кількістю хороших відгуків зростає рейтинг фрилансеру	Щоразу доводиться доводити свою профпридатність новим клієнтам, виконувати тестові завдання
4.	Аналітика	У профілі виконавця зберігається історія замовлень та оплат, а також ділове листування	Потрібно самостійно вести облік замовлень, на що не завжди вистачає часу

1.1 Загальна характеристика діяльності сервісів, що визначається предметною областю

На початку цього століття з'явилися платформи віддаленої праці, які виступають посередниками між замовниками та виконавцями. Ці платформи швидко набули популярності та довіри, оскільки призначені для забезпечення

безпеки обох сторін та спрощення співпраці. Вони створюють зручні умови для пошуку працівника, детального ознайомлення з його професійними якостями та досвідом.

Робота через такі платформи дозволяє укласти безпечні угоди, гарантуючи обом сторонам отримання бажаного результату. Крім того, вони надають можливість фрилансерам створити своє портфоліо, що допомагає роботодавцям краще оцінити стиль та професійний рівень працівника під час вибору.

Платформа, виступаючи посередником, приймає участь у вирішенні суперечок і розбіжностей як незалежна сторона. У випадку неналежного виконання завдання замовник може повернути свої кошти, що практично неможливо при прямій співпраці.

Оскільки більшість платформ функціонують на основі рейтингів, роботодавець може ознайомитися з відгуками, які інші роботодавці залишили про потенційного співробітника, що служить додатковим підтвердженням якості роботи фрилансера.

Раніше фрилансерам доводилося самотійно шукати замовників на основі довіри, але зараз завдяки появі платформ обидві сторони знаходять одна одну набагато швидше, а їхня співпраця стала безпечнішою та ефективнішою.

Сучасні біржі фрилансу повинні відповідати таким вимогам:

- зручність використання (юзабіліті): основні розділи повинні бути легкими для сприйняття та зрозумілими;
- портфоліо: у зареєстрованих фрилансерів повинно бути доступне портфоліо, де представлені приклади їхніх виконаних робіт;
- спеціалізація: фрилансери повинні бути відсортовані за чітко вказаною спеціалізацією;
- відгуки: формування рейтингу як для фрилансерів, так і для роботодавців, з веденням реєстру позитивних та негативних відгуків.

На сьогоднішній день біржі фрилансу стали зручним способом знаходження віддаленої роботи в Інтернеті, існує багато платформ, серед яких

деякі особливо популярні серед українських фрилансерів.

## 1.2 Аналіз існуючих ІС фриланс-бірж для митців

### 1.2.1 Фриланс-біржа Upwork

Біржа була створена у 2015 році після об'єднання двох бірж Elance та oDesk. Офіційна сторінка сайту за посиланням [upwork.com](http://upwork.com) [1]. Для того, щоб почати працювати, потрібно зареєструватися як клієнт (роботодавець) або фрилансер.

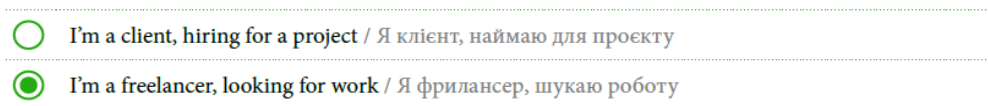


Рисунок 1.1 – Скриншот сторінки Реєстрація на біржі

Далі для фрилансера потрібно вказати наступні дані:

- First Name (Ім'я);
- Last Name (Прізвище);
- Email ;
- Country (країна);
- Адреса;
- Порогодинний рейт.

Також можна включити в профіль фрилансера короткий опис, а також інформацію про досвід роботи, якщо такий є, освіту та мовні вміння. Під час реєстрації слід вибрати 10 категорій, які на кращий спосіб визначають види робіт, які ви можете виконувати. Це дозволить Upwork належним чином відображати профіль фрилансера для відповідного типу клієнтів у результатах пошуку.

Upwork – це конкурентний ринок із великою кількістю учасників.

Додавання сертифікатів до свого профілю та створення портфоліо – це ефективний спосіб продемонструвати свої навички та кваліфікацію. Портфоліо може допомогти залишити враження на потенційних клієнтів, представляючи реальні приклади навичок та досвіду. Біржа встановлює наступні технічні вимоги для медіафайлів:

- документи розміром до 25 МВ;
- зображення (.jpg, .gif, .png, до 10 МВ);
- відеофайли (.mp4, .mov, .webm, .ogm, ogv, до 100 МВ, максимум 2 файлів, до 60 секунд);
- аудіофайли (.mp3, .wav, до 10 МВ, максимум 20 файлів).

З точки зору зручності пошуку необхідного контракту Upwork має такі налаштування пошуку:

- рівень виконавця – Experience Level (Entry, Intermediate, Expert);
- кількість пропозицій – Number of Proposals (Less than 5, 5 to 10, 10 to 15, 15 to 20, 20 to 50);
- інформація по клієнту – Client Info (My Previous Clients клієнт, з яким вже була співпраця, Payment Verified клієнт із підтвердженням методом оплати);
- категорія роботи – Category;
- тип роботи – Job Type (почасова оплата чи фіксована вартість роботи);
- довготривалість проекту – Project Length (Less than 1 month, 1 to 3 months, 3 to 6 months, More than 6 months);
- кількість годин на тиждень – Hours per Week (Less than 30 hrs/week, More than 30 hrs/week);
- історія контрактів роботодавця – Client History (No hires, 1 to 9 hires, 10+ hires);
- геолокація роботодавця – Client Location;
- часовий пояс роботодавця – Client time zones.

Категорії опису робіт фрилансера наведено на рисунку 1.2, а вікно налаштування пошуку – на рисунку 1.3.

<b>Accounting &amp; Consulting</b>		<b>Design &amp; Creative</b>	
<input type="checkbox"/> Accounting & Bookkeeping	<input type="checkbox"/> Financial Planning	<input type="checkbox"/> Art & Illustration	<input type="checkbox"/> Audio & Music Production
<input type="checkbox"/> Human Resources	<input type="checkbox"/> Management Consulting & Analysis	<input type="checkbox"/> Branding & Logo Design	<input type="checkbox"/> Game Art & AR/VR Design
<input type="checkbox"/> Other - Accounting & Consulting		<input type="checkbox"/> Graphic, Editorial & Presentation Design	<input type="checkbox"/> Performing Art
<b>Admin Support</b>		<b>Engineering &amp; Architecture</b>	
<input type="checkbox"/> Data Entry & Transcription Services	<input type="checkbox"/> Market Research & Product Reviews	<input type="checkbox"/> 3D Modeling & CAD	<input type="checkbox"/> Buildings & Landscape Architecture
<input type="checkbox"/> Project Manager	<input type="checkbox"/> Virtual Assistance	<input type="checkbox"/> Chemical Engineering	<input type="checkbox"/> Civil & Structural Engineering
<b>Customer Service</b>		<input type="checkbox"/> Contract Manufacturing	<input type="checkbox"/> Electrical & Electronic Engineering
<input type="checkbox"/> Community Management & Tagging	<input type="checkbox"/> Customer Experience & Tech Support	<input type="checkbox"/> Energy & Mechanical Engineering	<input type="checkbox"/> Interior & Trade Show Design
<b>Data Science &amp; Analytics</b>		<input type="checkbox"/> Physical Sciences	
<input type="checkbox"/> AI & Machine Learning	<input type="checkbox"/> Data Analysis & Testing		
<input type="checkbox"/> Data Design & Visualization	<input type="checkbox"/> Data Extraction/ETL		
<input type="checkbox"/> Data Mining & Management			
<b>IT &amp; Networking</b>		<b>Web, Mobile &amp; Software Dev</b>	
<input type="checkbox"/> Database Management & Administration	<input type="checkbox"/> DevOps & Solution Architecture	<input type="checkbox"/> Desktop Application Development	<input type="checkbox"/> Ecommerce Development
<input type="checkbox"/> ERP/CRM Software	<input type="checkbox"/> Information Security & Compliance	<input type="checkbox"/> Game Design & Development	<input type="checkbox"/> Mobile Development
<input type="checkbox"/> Network & System Administration		<input type="checkbox"/> Other - Software Development	<input type="checkbox"/> Product Management
<b>Legal</b>		<input type="checkbox"/> QA & Testing	<input type="checkbox"/> Scripts & Utilities
<input type="checkbox"/> Corporate & Contract Law	<input type="checkbox"/> Finance & Tax Law	<input type="checkbox"/> Web & Mobile Design	<input type="checkbox"/> Web Development
<input type="checkbox"/> International & Immigration Law	<input type="checkbox"/> Public Law		
<b>Sales &amp; Marketing</b>		<b>Writing</b>	
<input type="checkbox"/> Display Advertising	<input type="checkbox"/> Email & Marketing Automation	<input type="checkbox"/> Content & Copyrighting	<input type="checkbox"/> Creative Writing Services
<input type="checkbox"/> Lead Generation & Telemarketing	<input type="checkbox"/> Marketing & Brand Strategy	<input type="checkbox"/> Grant & Proposal Writing	<input type="checkbox"/> Other - Writing
<input type="checkbox"/> SEO & SEM Services	<input type="checkbox"/> Social Media & PR Services	<input type="checkbox"/> Resumes & Cover Letters	<input type="checkbox"/> Technical Writing
<b>Translation</b>			
<input type="checkbox"/> Legal, Medical & Technical Translation	<input type="checkbox"/> Translation & Localization		

Рисунок 1.2 – Категорії опису робіт фрилансера

На жаль, разом із такими перевагами, як широкі можливості пошуку контрактів та зручний автоматизований сервіс для виконання та оплати

контрактів, Upwork має суттєві обмеження, особливо для митців та дизайнерів. Основними недоліками є відсутність оцінки рівня майстерності митців та відсутність чіткої визначеності щодо вартості їхньої роботи. Зауважимо, що дизайнери та митці становлять менше половини ринку фриланс-працівників, тому ці аспекти є критичними для успішної роботи фриланс-біржі.

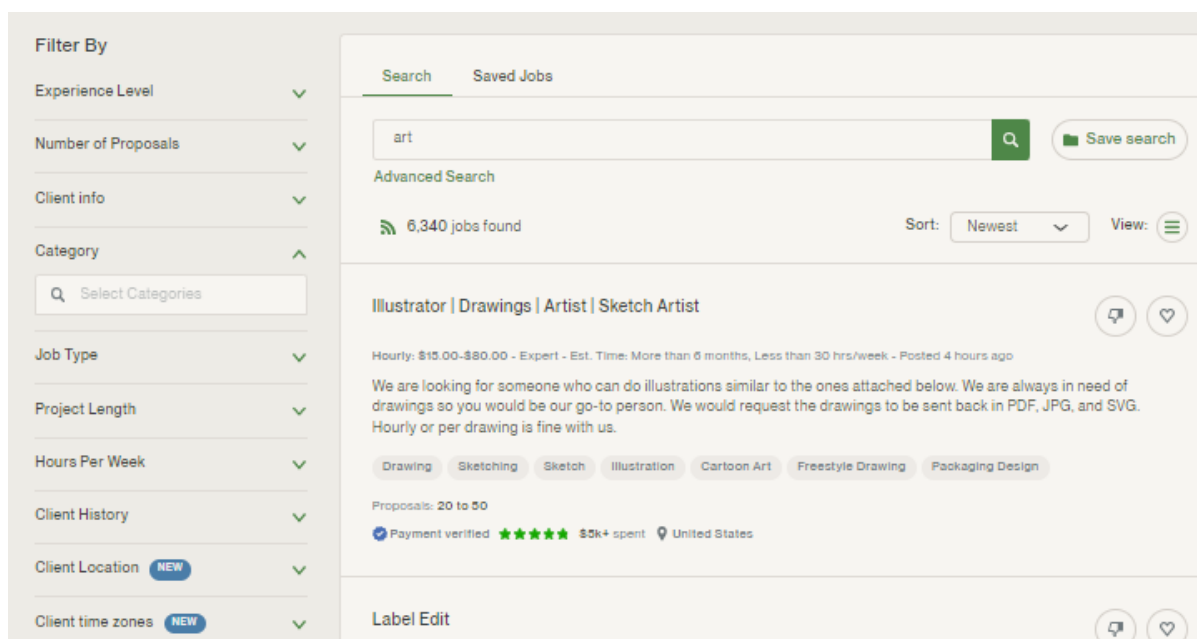


Рисунок 1.3 – Вікно налаштування пошуку

Для прикладу розглянемо замовлення для створення різної анімації 2D. Вартість виконання замовлення роботодавець визначає сам. Хтось спирається на свій особистий досвід, хтось проводить аналіз вже існуючих подібних замовлень, хтось взагалі ставить вартість інтуїтивно. У будь-якому випадку, платформа не надає змоги в корегуванні хоча б інтервалу вартості завдань. Через це відбувається демпінг вартості та знецінення рівня обізнаності та досвіду митців.

На рисунку 1.4 можна побачити замовлення, яке не має вартості взагалі, полягає у створенні 3D мініатюр, де замовник зазначив Середній рівень володіння навичками, а серед знань необхідного ПЗ зазначені Illustrator, Photoshop, YouTube, які не є ПЗ для створення трьохвимірних моделей: Illustrator – це застосунок компанії Adobe для роботи із векторною графікою (2D),

Photoshop – це застосунок компанії Adobe для роботи із растровою графікою (2D), YouTube застосунок для програвання відеофайлів. Тобто, ми бачимо відсутність технічно правильного опису завдання. Додатково, тривалість проекту зазначена в інтервалі, тобто, є вірогідність того, що замовник не може оцінити кількість часу, необхідного для виконання завдання.

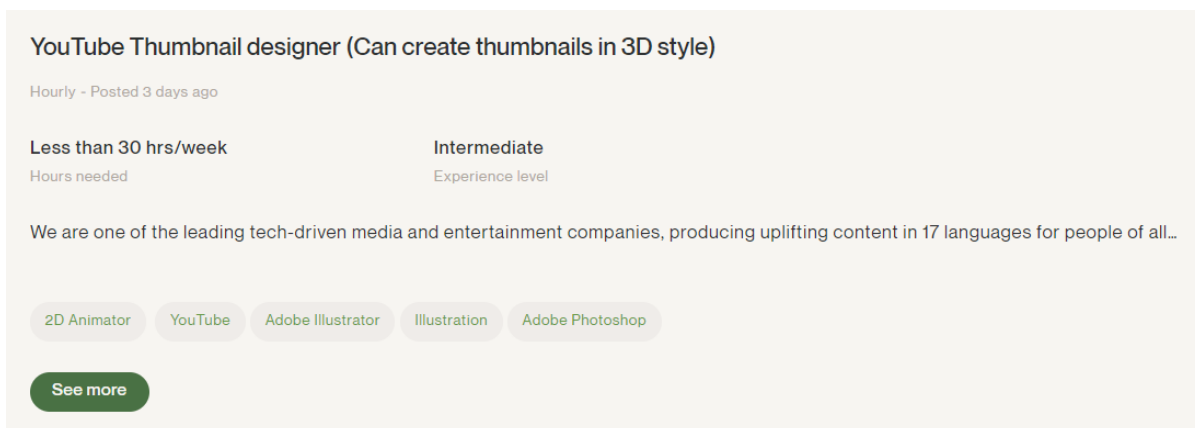


Рисунок 1.4 – Вікно замовлення за пошуком «2D Animator»

На рисунку 1.5 можна побачити замовлення, яке оцінено у 300\$, полягає у створенні 2D анімації, а саме короткого відео у стилі вестерн або аніме, де замовник зазначив рівень володіння навичками Експерт, а серед знань необхідного ПЗ зазначені Clip Studio Paint, Krita, Toon Moom Harmony та Adobe Animate.

Clip Studio Paint (раніше відомий як Manga Studio) - це програмне забезпечення для малювання та створення коміксів та ілюстрацій. Воно широко використовується художниками, мангаками (творцями манга) та іншими професіоналами в галузі мистецтва. Clip Studio Paint також включає функції, спеціально розроблені для створення манги, такі як можливість розміщення кадрів, обробка тексту та інші.

Krita – це вільний та відкритий графічний редактор, призначений для роботи з растровими зображеннями. Він розробляється спеціально для художників, ілюстраторів та художників-коміксів, а також включає різноманітні інструменти для малювання та редагування.

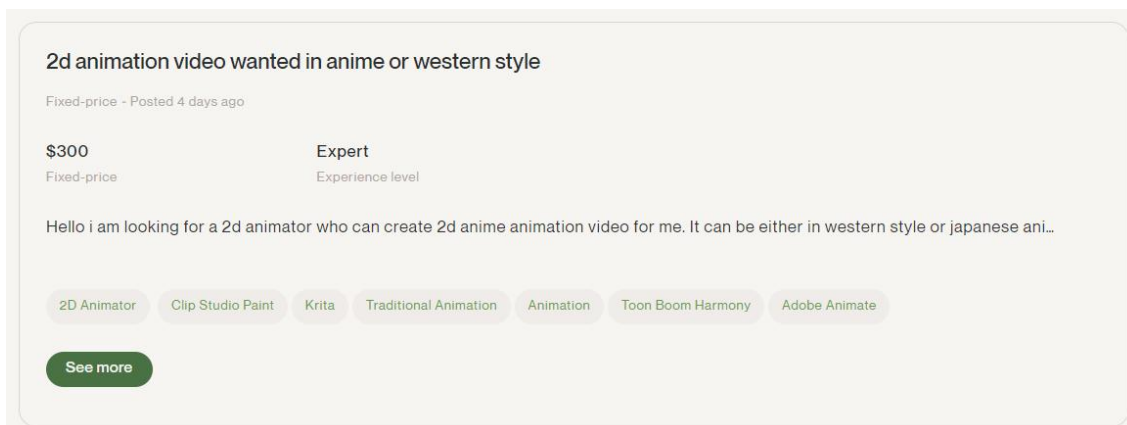


Рисунок 1.5 – Вікно замовлення за пошуком «2D Animator»

Toon Boom Harmony – це професійний програмний продукт для створення анімації, особливо призначений для анімаційного виробництва у галузі мультфільмів, телебачення та веб-контенту. Деякі ключові можливості Toon Boom Harmony включають створення складних персонажів, розгортання та редагування рухів, роботу з 2D та 3D анімацією, використання системи скелетної анімації для зручного управління персонажами, а також інтеграцію із зовнішніми програмами для обробки звуку та інших завдань.

Adobe Animate – це програмний продукт від компанії Adobe, спеціально розроблений для створення анімацій та інтерактивних мультимедійних контентів. Цей інструмент використовується для створення анімацій, інтерактивних веб-елементів, ігор, анімованих рекламних банерів та інших мультимедійних проектів.

В цьому випадку стек застосунків більш вдало підібраний, ніж в попередньому. Тривалість проекту не зазначена, як і технічні вимоги до готового відео (роздільна здатність, кількість кадрів, формат, тощо).

Розглянемо третій приклад завдання за запитом «2D Animator», який представлено на рисунку 1.6 можна побачити замовлення, яке оцінено у 30\$, полягає у створенні 2D анімації, а саме короткого відео, яке буде використано для постингу в соціальній мережі Instagram (це дає нам інформацію про розмір

відео, можлиий формат та роздільну здатність), тут також присутній референс, що допомагає оцінити митцям об'єм роботи. Замовник зазначив Середній рівень володіння навичками, а серед знань необхідного ПЗ зазначений Adobe After Effect. Adobe After Effects – це програмний продукт від компанії Adobe, який призначений для створення та редагування відео та анімаційних ефектів. Цей інструмент широко використовується в індустрії відеопродакшну для створення професійних візуальних ефектів, анімації, титрів, та інших видів графічно-візуального контенту. Тобто, в цьому завданні лпис має більше інформації про необхідне готове рішення, але стек застосунків підібраний некоректно, бо анімація, яка створюється в Adobe After Effects занадто складна для анімації графічного дизайну в Instagram.

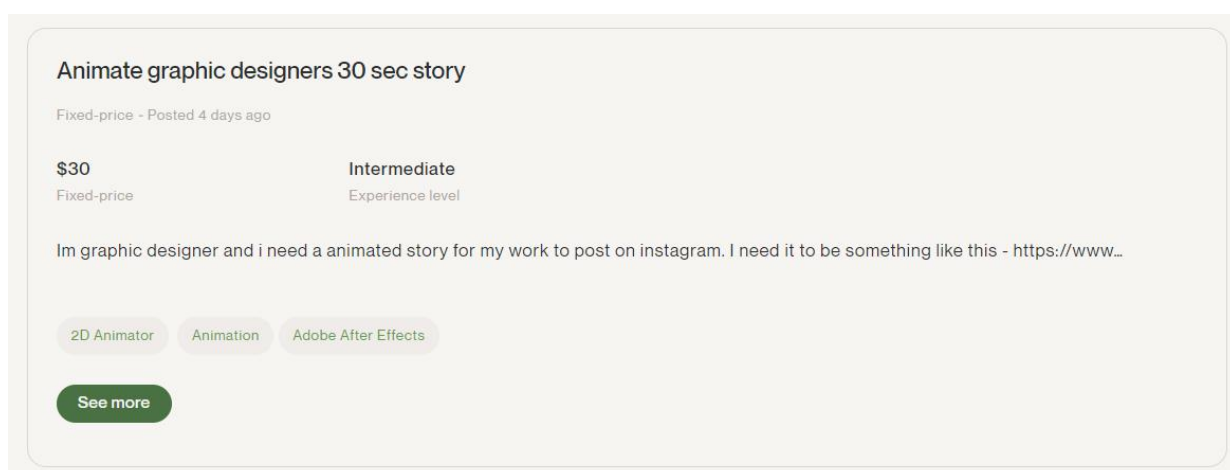


Рисунок 1.6 –Вікно замовлення за пошуком «2D Animator»

Якщо зробити запит пошуку замовлень «3D Modeling» (див.рисунок 1.7), то тут таке ж різноманіття варіантів замовлень та опис цих замовлень. У першому випадку хоч і зазначено серед необхідних скілів «3D Modeling», але задача полягає у експорті трьохвимірної моделі із програми Blender, де попередньо необхідно виконати процедуру ригінгу (нанесення основних вузлів скелету на модель), а потім виконати цієї підготовленої моделі у ігровий рушій Unreal Engine. Серед необхідно ПЗ також зазначено Unity, це аналог ігрового рушія Unreal Engine та не є застосунком для моделювання трьохвимірної моделі.

Вартість виконання замовлення 50\$, тим файлу FBX, рівень експертності виконавця – Середній.

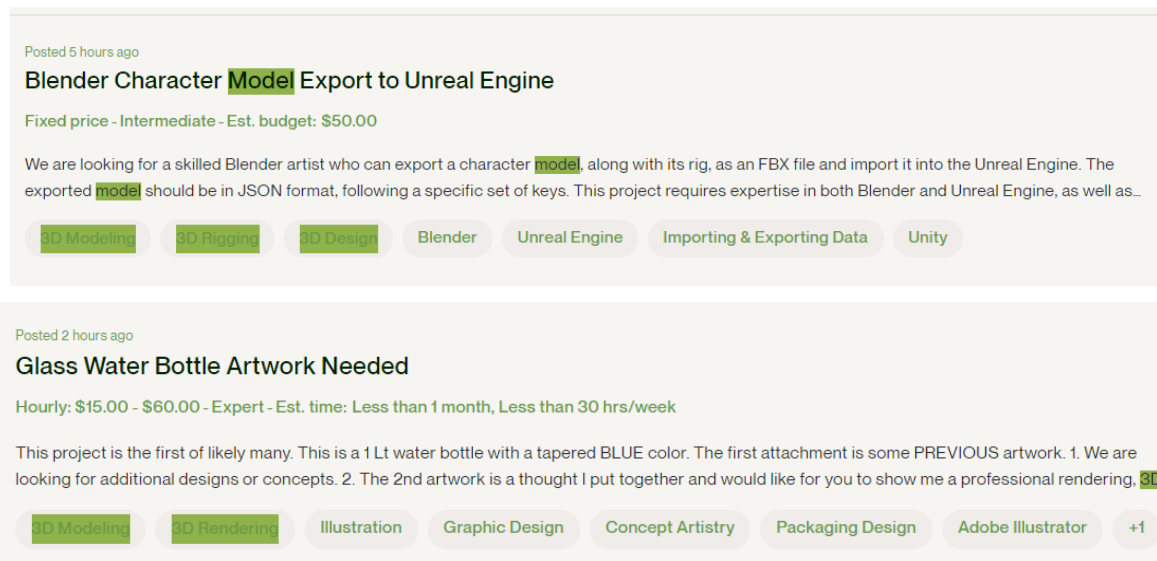


Рисунок 1.7 – Вікно замовлення за пошуком «3D Modeling»

У другому випадку необхідно створити модель пляшки із водою, є референс та опис загального вигляду моделі. Це завдання, по факту, є оплачуваним тестовим, за результатом якого можлива подальша співпраця із митцем. Хоч і є приблизній вигляд готового результату роботи, але вартість зазначена у інтервалі 15-60\$. Серед ПЗ зазначені програми для растрової графіки, векторної графіки та роботи з графічним дизайном, що не є релевантними для роботи із трьохвимірним моделюванням.

Пошук замовлення (див.рисунок 1.8) можна здійснити в тому числі за назвою та необхідними навичками, проте якщо замовник не знає, як коректно називається вид митця, який може виконати замовлення, можлива помилка у його оформленні й загальному розумінні сутності роботи, не кажучи про оцінку її собівартості.

Urwork, хоча є популярною біржею для фрілансу, має деякі недоліки, особливо для цифрових митців:

- відсутність Оцінки Рівня Митця: У розробників, дизайнерів та інших цифрових митців часто немає оцінки їхнього рівня кваліфікації або досвіду на

платформі. Це робить важчим для клієнтів оцінити навички та обрати найбільш відповідального виконавця;

– відсутність Фіксованих Вартостей Робіт: Upwork часто сприяє роботам на годину, що може бути неідеальним для цифрових митців, які звикли до фіксованих тарифів за конкретні завдання;

– конкурентна Спрямованість: За рахунок великої кількості виконавців на платформі конкуренція може бути дуже великою, що робить складнішим отримання замовлень для нових виконавців;

– обмежений Контроль Над Платформою: Виконавці мають обмежений контроль над функціоналом та умовами платформи, оскільки всі правила встановлюються самою Upwork.

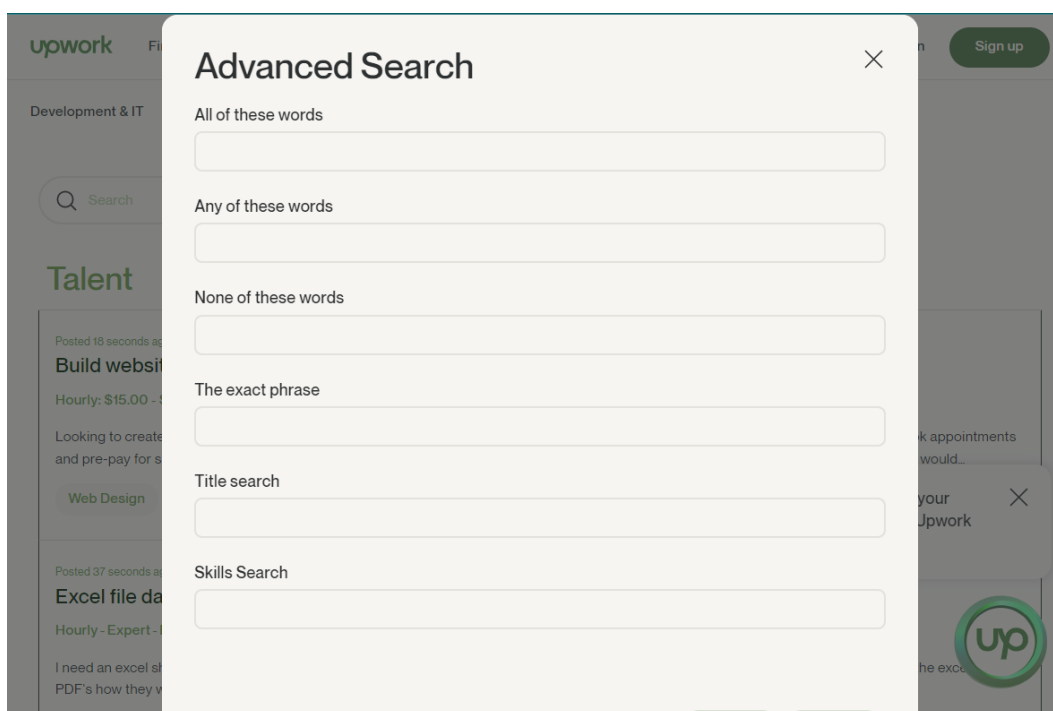


Рисунок 1.8 – Пошук замовлення на Upwork

Хоча Upwork може бути важливою платформою для знаходження роботи, ці недоліки варто враховувати, особливо якщо ви цифровий митець і шукаєте специфічні умови роботи.

## 1.2.2 Сервіс Kabanchik.ua

Сервіс "Кабанчик" (див.рисунок 1.9) не є типовою фриланс-біржею в традиційному розумінні [2]. Це повноцінний маркетплейс, де можна замовити різноманітні послуги, більшість з яких пов'язані з виконанням робіт у сфері онлайн.

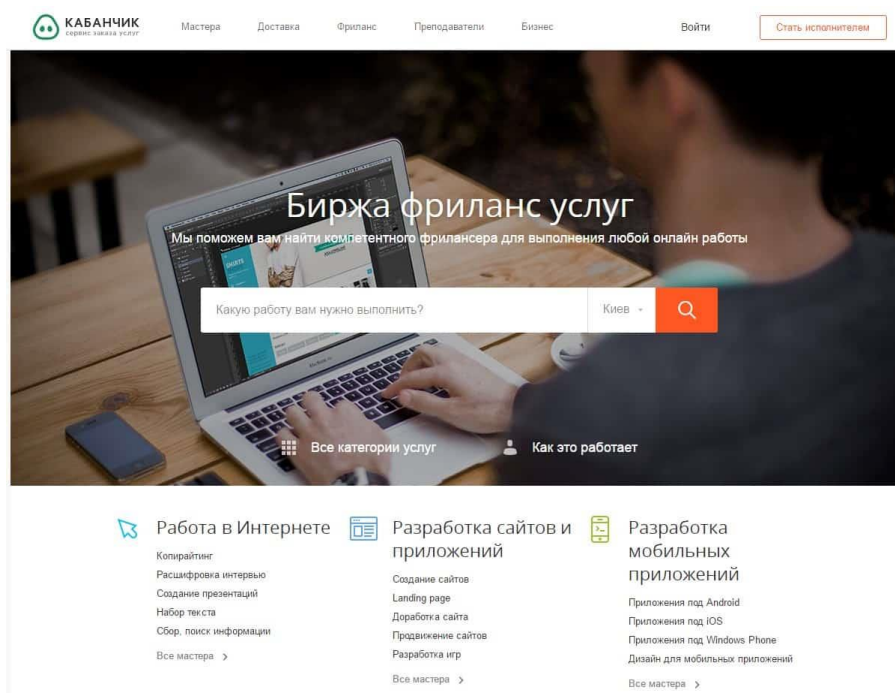


Рисунок 1.9 – Вікно пошуку сервісу "Кабанчик"

Процес взаємодії між учасниками має простий алгоритм:

- а) замовник розміщує завдання з докладним описом необхідних деталей;
- б) виконавці подають свої заявки, і замовник обирає одного з них;
- в) обраний виконавець виконує завдання;
- г) замовник оцінює результати, виставляє свою оцінку виконавцеві.

Профіль виконавця має такі критерії (див.рисунок 1.10):

- ім'я та прізвище;
- геолокація;
- спеціалізація;
- короткий опис робітника;
- дані з верифікації працівника.

Також на сторінці фрилансера є можливість перегляду вже виконаних контрактів та відгуків.

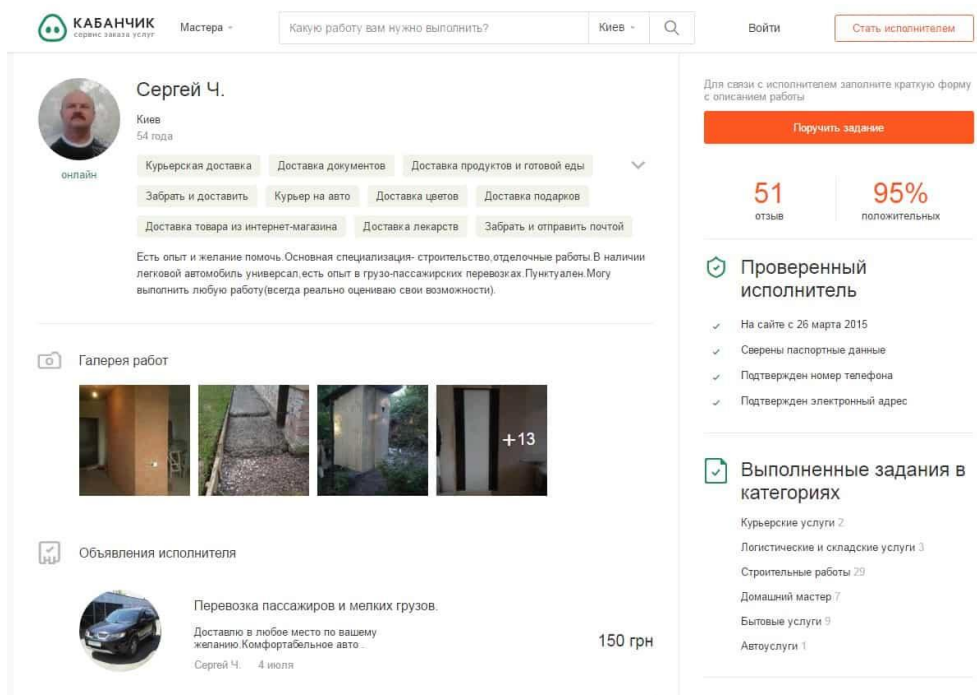


Рисунок 1.10 – Профіль виконавця сервісу "Кабанчик"

Для дизайнерів та митців Кабанчик надає можливість створення окремого розділу цифрового профілю з портфоліо.

Формування рейтингу виконавця на сервісі KabanChik.ua є важливим етапом, який визначає його можливості отримання якісних замовлень. Замовники, після завершення проекту, залишають відгук та оцінюють виконавця за якістю роботи, ввічливістю та пунктуальністю за 5-бальною шкалою. Тільки позитивні відгуки враховуються у рейтингу, і топові виконавці мають не менше 90% позитивних відгуків.

Однією з вигід сервісу KabanChik.ua є програма захисту Замовників. Ця програма надає можливість Замовникам, які постраждали від неналежних Фахівців, вимагати компенсацію до 1 000 грн за одне замовлення, створене на платформі. Після розгляду звернення Адміністрація сервісу приймає рішення щодо можливості надання компенсації.

Мобільний додаток є додатковою перевагою, дозволяючи швидко створювати, приймати та виконувати контракти. Також існує можливість придбати топ-статус фрилансера, що підвищує ймовірність успішного укладення контракту з фрилансером.

Для досягнення топ-статусу на платформі, необхідно відповідати певним критеріям:

- прийняти не менше 10 замовлень у категоріях першого рівня, таких як "Кур'єрські послуги", "Клінінгові послуги", "Домашній майстер", "Ремонт техніки";

- заробити рейтинг не нижче 95% за останніми 10-20 замовленнями в даній категорії, при отриманні лише позитивних відгуків з максимальною кількістю балів;

- мати щонайменше 50% виконаних замовлень від середнього показника по категорії (якщо середній показник, наприклад, 54%, то виконавець повинен мати не менше 27%);

- не мати заборгованості на балансі спеціаліста, який вводився на початку 2020 року як один із варіантів оплати роботи;

- звертати увагу на бюджет виконаних замовлень, де в топ потрапляють ті, у кого бюджети є найвищими. Тобто важливо не просто виконувати багато завдань, а виконувати їх за більш високими сумами.

При створенні нового замовлення немає змоги додати ПЗ або навичкі, які необхідні для виконання, проте можна вибрати категорію, під яку підпадає замовлення. Можна вважати плюсом те, що категорії розділені достатньо детально (див.рисунок 1.11). Також можна вказати термін виконання, додати опис, що треба зробити, та файли референси. Також сайт дозволяє визначити рекомендовану вартість роботи, яка є виключно рекомендацією, роботодавець має змогу зазначити будь-яку суму (див.рисунок 1.12).

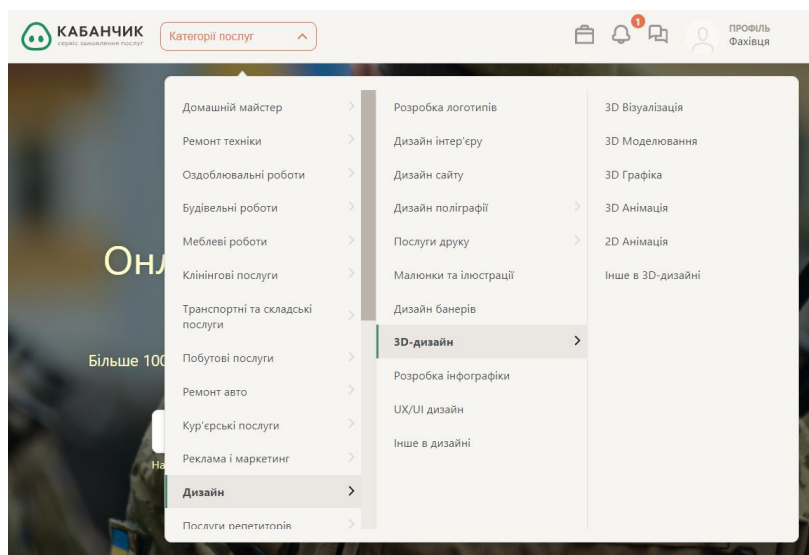


Рисунок 1.11 – Сторінка вибору категорії замовлення у сервісі "Кабанчик"

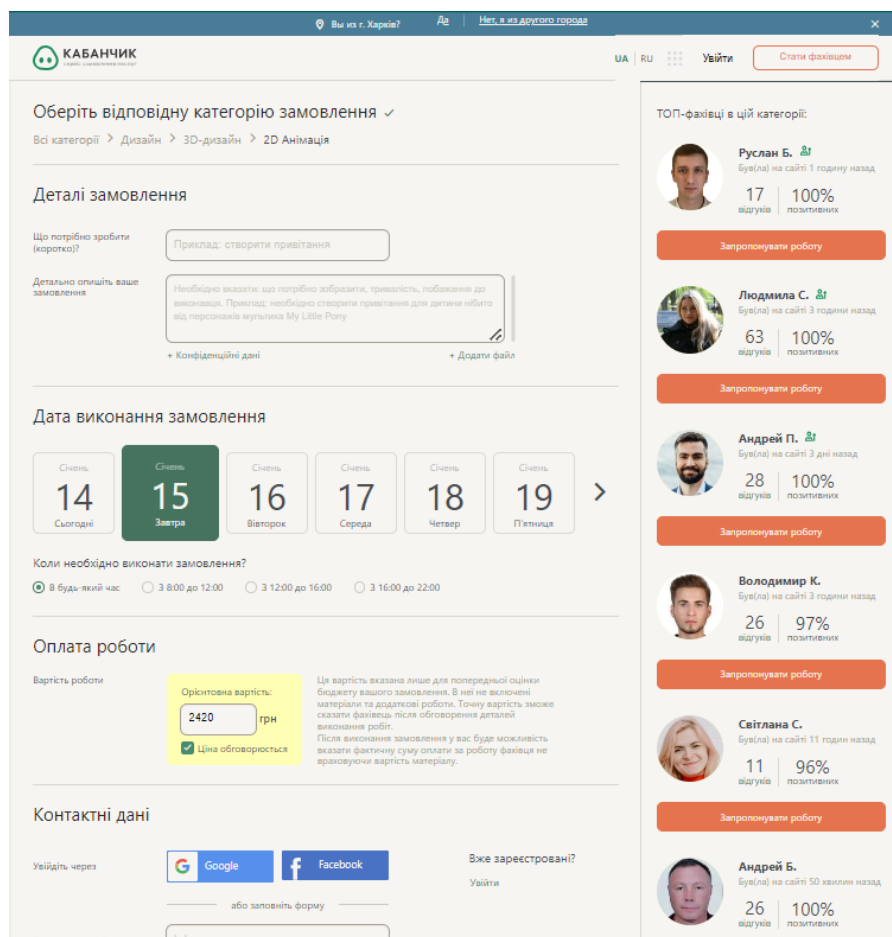


Рисунок 1.12 – Сторінка створення замовлення у сервісі "Кабанчик"

Розглянемо приклади завдання за запитом «2D Анімація» [3]. В першому прикладі (див.рисунок 1.13) необхідно виконати анімування для короткого

ролика соціальної мережі Instagram, зазначена кінцева дата виконання та вартість у 500 грн. Сервіс "Кабанчик" є українським, тому розрахунок виконується в національній валюті. В описі немає ані прикладів, ані референсів, ні опису, що саме має бути зроблено, тому зробити висновок, наскільки відповідає фінансова винагорода обсягу роботи, дуже складно.

У другому прикладі (див.рисунок 1.14) необхідно створити анімацію вже існуючої моделі, вказан тип анімації, тривалість, є приближний опис, що саме необхідно зробити. Проте оцінити, наскільки відповідає фінансова винагорода обсягу роботи, дуже складно.

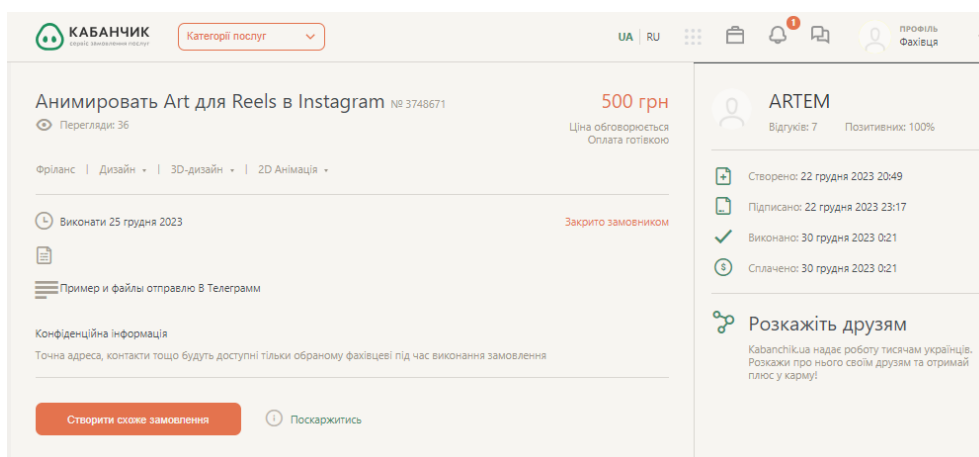


Рисунок 1.13 – Вікно замовлення за пошуком «2D Animator» у сервісі "Кабанчик"

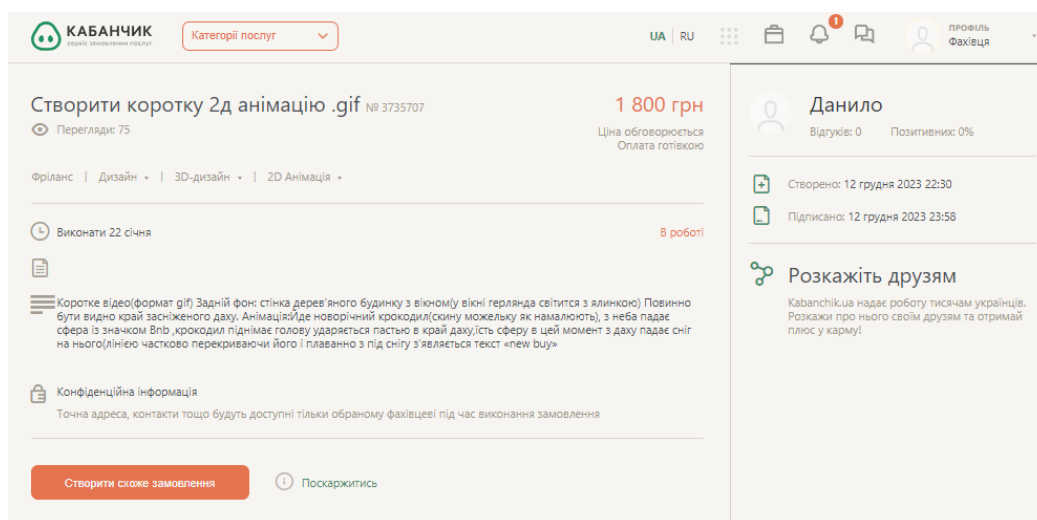


Рисунок 1.14 – Вікно замовлення за пошуком «2D Animator» у сервісі "Кабанчик"

До недоліків сервісу можна віднести наступні моменти:

- сервіс виставляє середні ціни на той чи інший вид послуг або ціни від замовника, який не завжди може оцінити обсяг роботи та її складність. Це не дозволяє оцінити справжній рівень фрилансера у грошовому еквіваленті;
- сервіс дуже широкоспеціалізований, тому оцінка контрактів для дизайнерів та цифрових митців не реалізована, не можна зазначити, наприклад, необхідне ПЗ для виконання замовлення та рівень його володіння. Це робить сервіс не дуже привабливим для постійної роботи митців та дизайнерів;
- сервіс не дозволяє завантажувати файли у великій роздільній здатності, що є важливим критерієм для демонстрації рівня роботи митця чи дизайнера.

### 1.3 Постановка задачі дослідження

Основними напрямками діяльності будь-якої фриланс-біржі є організація бази контрактів від роботодавців та реєстрація фрилансерів і роботодавців. Цей процес включає в себе завдання з обліку, класифікації, зберігання контрактів і профілів користувачів біржі, а також їхнє обслуговування. Рівень обслуговування біржі в значній мірі залежить від своєчасності та повноти формування бази контрактів, роботодавців та виконавців.

Обслуговування користувачів фриланс-біржі здійснюється різним шляхом укладання контрактів поміж роботодавцем та виконавцем і допомоги окремим користувачам в підборі необхідного їм контракту та виконавців. Термін виконання контракту для різних категорій митців і кількість видаваних контрактів для кожного митця визначається окремо, виходячи з виду контракту і категорії митця. Число контрактів, що виконуються, не обмежується.

Обслуговування користувачів фриланс біржі пов'язане з урахуванням великої кількості операцій, безліч контрактів, виконавців і роботодавців. Складність пошуку потрібного контракту в каталозі займає тривалий час і у більшості виконується методом перебору усіх контрактів обраної галузі.

У біржу приходять багато контрактів від різних роботодавців. Кожен контракт має наступні параметри:

- назва контракту (короткий опис типу завдання);
- унікальний шифр (ICN);
- галузь;
- роботодавець;
- технологічний стек;
- рівень виконавця;
- вартість роботи;
- строк виконання.

Кожен контракт може бути в кількох екземплярах. Всі контракти розрізняються по своєму унікальному шифру – ICN. Класифікація контрактів розподіляє контракти за галузями знання відповідно до їх змісту та за характером завдання. У ній використовується цифро буквенні індекси ступінчастою структури. У системі є база даних (БД). Вона являє собою сукупність даних про всі види робіт, що можуть бути виконані митцями (розробка трьохмірної моделі, ілюстрація, дизайн контекстної реклами і ін.), побудовану у вигляді списків даних. Основне її призначення – зберігання шаблонів, для їх подальшого доцільного використання при створенні контракту. БД зберігається на електронному вигляді

Слід виділити такі етапи взаємодії клієнта та виконавця на біржі фрилансу:

- обговорення деталей проекту;
- визначення термінів виконання;
- формування бюджету;
- укладання "Безпечної угоди";
- резервування оплати замовником;
- виконання проекту фрилансером;
- надання підсумкового результату;
- внесення правок за потреби;

- затвердження клієнтом;
- надходження грошей на рахунок виконавця;
- обмін відгуками.

В даній роботі розглядаються замовлення саме для представників артових професій, це Ui/Ux Designer, Illustrator, Designer, 2D Artist, Concept Artist, 2D Enviroment Artist, Retoucher, 3D Character Artist, 3D Generalist, VFX Artist, 2D Animator, 3D Animator, 3D Modeller, 3D Enviroment Artist, 3D. Більшість бірж не дає змоги оцінити рівень митця та вартість його роботи. Загалом це залишається на розсуд роботодавця, через що виникає дуже великий дисбаланс ринку праці для митців.

Таким чином, необхідно за наявності існуючої інформації у системі фриланс-біржі провести дослідження застосування методів інтелектуального аналізу даних в системі фриланс-біржі для 2D/3D Generalist для об'єктивного встановлення вартості робіт цифрових митців.

Виконаємо математичну постановку задачі дослідження.

Вхідна інформація (дано):

– експертна оцінка вартості аналогічних робіт із портфоліо інших фрилансерів на фриланс-біржі.

Вихідна інформація (необхідно): оціночна вартість поточної роботи фрилансера.

Таким чином, необхідно провести дослідження застосування методів інтелектуального аналізу даних в системі фриланс-біржі для 2D/3D Generalist для об'єктивного встановлення вартості поточної роботи фрилансера.

## 2 ДОСЛІДЖЕННЯ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ТА ЗАСТОСУВАННЯ ЇХ В СИСТЕМАХ ФРИЛАНС-БІРЖ

Інформаційні потоки наскрізь пронизують сучасний світ. Особливостями сучасних вимог до оброблення інформації є величезний обсяг даних та їх різноманіття, що робить неспроможними як статистичні, так й експертні підходи в багатьох практичних аспектах, зокрема економічних, тому для аналізу інформації, накопиченої у сучасних базах даних, методи повинні бути ефективними, тобто простими у використанні, мати значний рівень масштабованості й автоматизму [4].

Люди шукають закономірності в даних з тих пір, як почалось їх осмислене життя. Мисливці шукають закономірності в поведінці тварин, фермери шукають закономірності у рості сільськогосподарських культур, політики шукають закономірності в думці виборців. Робота дослідника полягає в тому, щоб осмислити інформативність та повноту даних, які є в його розпорядженні, виявити зв'язки між ними і побудувати моделі з метою прийняття правильних рішень.

Існує багато наукових напрямків застосування методів інтелектуального аналізу даних. У біології інтелектуальний аналіз застосовують, щоб допомогти ідентифікувати гени в геномах. У біомедицині інтелектуальний аналіз застосовується для прогнозування активності препарату та дозволяє враховувати не тільки хімічні властивості ліків, але і їх тривимірну структуру, що прискорює розробку ліків і знижує їх вартість. У всіх перерахованих додатках методи інтелектуального аналізу даних досягли рівня практичного застосування. Методи інтелектуального аналізу даних можуть врятувати життя, наприклад, при виявленні змін характеристик фізіологічних процесів, які не можна пояснити добовим ритмом пацієнта, дією лікарських препаратів тощо. Також, оскільки сучасний світ спирається на вразливі мережеві комп'ютерні системи і все більше турбується про кібер-безпеку, інтелектуальний аналіз все

частіше застосовується для виявлення атак та розпізнавання незвичайних моделей поведінки інформаційних систем.

Управління сучасним бізнесом немислимо без прогнозування та аналізу даних, який в залежності від цілей дослідження можна розділити на наступні види:

1) інформаційно-пошуковий і візуальний аналіз – у ході такого аналізу, що не набуваючи ніяких нових знань про предмет, маємо можливість розглянути його по частинах і з різних точок зору. Здійснюється це, як правило, шляхом чітко сформульованого запиту до реляційної бази даних. Цей вид аналізу лежить в області деталізованих даних, ніяк їх не узагальнюючи;

2) оперативно-аналітичний аналіз, або OLAP – дані агрегуються, надаючи аналітику можливість отримати будь-яку ступінь узагальнення в будь-якому розрізі. На відміну від інформаційно-пошукового аналізу, тут можемо виявити різного роду закономірності в даних, які інакше були б не видно. OLAP вводить нас у сферу узагальнених даних;

3) інтелектуальний аналіз, або Data Mining – спрямований на виявлення прихованих закономірностей у даних, наприклад повторюваних шаблонів або кластерів. Інакше кажучи, на його основі можна отримати моделі, що дозволяють краще розуміти дані і передбачати їхню поведінку. Data Mining в дійсності припускає безпосереднє виявлення знань [5].

Технологія Data Mining (інтелектуальний аналіз даних) – це видобуток даних з великих масивів. Цей метод виявляє раніше невідомі дані, нетривіальні, практично корисні й доступні для інтерпретації знань, що необхідні для прийняття рішень у різних сферах людської діяльності, тобто інформація, знайдена в процесі застосування методів Data Mining, повинна бути нетривіальною та раніше невідомою. Знання повинні описувати нові зв'язки між властивостями, прогнозувати значення одних ознак на основі інших. Наголосимо на тому, що сфера застосування інтелектуального аналізу даних є необмеженою.

Найбільше цей метод підходить комерційним підприємствам, що розпочинають проекти на основі інформаційних сховищ даних (Data Warehousing). Досвід таких підприємств показує, що віддача від використання методу може досягати десятків разів. Підприємства за допомогою подібних інформаційних методів аналізу можуть одержати відчутні переваги перед конкурентами та у стратегічному плануванні тощо.

Найбільшою перевагою методів інтелектуального аналізу даних є здатність виявлення нових знань, які неможливо одержати звичайними методами економічного аналізу, завдяки розширеним алгоритмам завдань. На теперішній час саме інтелектуальний аналіз є найбільш затребуваним на підприємствах найближчим часом.

Конкретні переваги аналізу даних включають наступне:

- більш ефективний маркетинг і продажі. Інтелектуальний аналіз даних допомагає маркетологам краще зрозуміти поведінку та вподобання клієнтів, що дозволяє їм створювати цільові маркетингові та рекламні кампанії. Подібним чином відділи продажів можуть використовувати результати інтелектуального аналізу даних, щоб підвищити коефіцієнт конверсії потенційних замовників та виконавців і продавати додаткові продукти та послуги існуючим клієнтам. Інтелектуальний аналіз даних і прогнозне моделювання також забезпечують механізми рекомендацій, які пропонують фрілансерам можливі замовлення, а замовникам допомагають зорієнтуватися із вартістю та стровами виконання замовлення;

- краще обслуговування клієнтів. Завдяки аналізу даних є можливість швидше виявляти потенційні проблеми з обслуговуванням клієнтів і надавати актуальну інформацію для використання під час дзвінків і онлайн-чатів із клієнтами;

- посилене управління ризиками. Є можливість краще оцінювати фінансові, правові, кібербезпекові та інші ризики та розробляти плани управління ними;

– менші витрати. Інтелектуальний аналіз даних допомагає заощаджувати кошти за рахунок підвищення операційної ефективності бізнес-процесів і зменшення надмірності та марнотратства корпоративних витрат.

Зрештою, ініціативи з інтелектуального аналізу даних можуть призвести до підвищення доходів і прибутків, а також до конкурентних переваг, які відрізняють компанії від їхніх бізнес-конкурентів.

## 2.1 KDD та модель інтелектуального аналізу даних

Процес пошуку корисних даних у сховищі має назву Knowledge Discovery in Databases (KDD) – відкриття знань у базі даних. KDD включає питання підготовки даних, вибору інформативних ознак, очищення даних, застосування методів «здобичі даних» (Data Mining) [6,7], а також обробки і інтерпретації отриманих результатів. Добуток знань у базах даних – це послідовність дій, яку необхідно виконати для побудови моделі пошуку знання. Ця послідовність не описує конкретний алгоритм або математичний апарат, не залежить від предметної галузі. Це набір атомарних операцій, комбінування яких дає змогу одержати потрібний розв’язок. KDD складається з таких етапів, як підготовка вхідних даних, вибір інформативних ознак, очищення даних, застосування методів інтелектуального аналізу даних, підсумкове оброблення та інтерпретація отриманих результатів.

Процес створення моделі інтелектуального аналізу включає такі етапи:

- засвоєння даних (обстеження первинних даних);
- перевірка даних на повноту;
- аналіз ключових факторів впливу;
- виділення винятків;
- виявлення категорій;
- аналіз сценаріїв;
- прогнозування.

У центрі цього процесу перебувають методи інтелектуального аналізу даних, які дають змогу виявляти закономірності та знання [7,8]. Інтелектуальний аналіз даних використовує різні методи та техніки для отримання корисної інформації та знань з великих обсягів даних. Ось кілька основних методів інтелектуального аналізу даних:

1) методи машинного навчання:

- класифікація: розділення об'єктів на категорії або класи на основі їх характеристик;
- регресія: прогнозування числових значень на основі попередніх даних;
- кластерний аналіз: групування схожих об'єктів у кластери;
- асоціативний аналіз: виявлення взаємозв'язків між елементами великих наборів даних;

2) обробка природної мови (Natural Language Processing, NLP):

- розпізнавання інтенту: визначення намірів користувачів на основі текстових або вимовлених виразів;
- синтаксичний аналіз: розпізнавання структури речень та визначення взаємозв'язків між словами;

3) аналіз великих даних (Big Data Analytics):

- методи обробки великих обсягів даних: використання розподілених систем для аналізу великих обсягів даних;
- методи обробки стрімінгових даних: робота з потоковими даними в реальному часі.

4) інтелектуальні системи підтримки прийняття рішень:

- системи експертів: Використання правил і експертних знань для прийняття рішень;
- методи прийняття рішень на основі знань: Врахування експертних знань та даних для прийняття рішень;

5) виявлення аномалій:

– методи виявлення викидів: аналіз даних для виявлення аномальних або незвичайних відхилень;

б) оптимізація:

– методи математичної оптимізації: знаходження оптимальних значень параметрів або рішень задачі;

7) системи рекомендацій:

– методи фільтрації інформації: надання персоналізованих рекомендацій на основі історії та вподобань користувача.

Ці методи часто використовуються в поєднанні для розв'язання складних завдань аналізу даних та витягування знань. Вибір конкретного методу залежить від типу даних, завдань аналізу та бізнес-задач, які потрібно вирішити.

Інтелектуальним аналізом даних (data mining) зазвичай займаються спеціалісти з обробки даних та інші кваліфіковані спеціалісти з ВІ та аналітики. Але це також можуть виконувати бізнес-аналітики, керівники та працівники, які функціонують як спеціалісти з даних громадян в організації.

Його основні елементи включають машинне навчання та статистичний аналіз, а також завдання керування даними, які виконуються для підготовки даних для аналізу. Використання алгоритмів машинного навчання та інструментів штучного інтелекту автоматизувало більшу частину процесу та полегшило видобуток масивних наборів даних, таких як бази даних клієнтів, записи транзакцій і файли журналів із веб-серверів, мобільних додатків і датчиків.

Процес аналізу даних (data mining) можна розбити на чотири основні етапи:

1) Збір даних. Релевантні дані для аналітичної програми визначаються та збираються. Дані можуть розташовуватися в різних вихідних системах, сховищі даних або озері даних, що стає все більш поширеним у середовищах великих даних, які містять суміш структурованих і неструктурованих даних. Також можна використовувати зовнішні джерела даних. Звідки б не надходили дані,

фахівець із обробки даних часто переміщує їх до озера даних для решти етапів процесу.

2) Підготовка даних. Цей етап включає набір кроків для підготовки даних до видобутку. Він починається з дослідження даних, профілювання та попередньої обробки, після чого йде робота з очищення даних для виправлення помилок та інших проблем із якістю даних.

3) Видобуток даних. Після того, як дані підготовлені, фахівець з даних вибирає відповідну техніку інтелектуального аналізу даних (data mining), а потім реалізує один або кілька алгоритмів для інтелектуального аналізу. У програмах машинного навчання алгоритми, як правило, потрібно випробувати на вибіркових наборах даних, щоб шукати необхідну інформацію, перш ніж їх перевірити на повному наборі даних.

4) Аналіз та інтерпретація даних. Результати аналізу даних використовуються для створення аналітичних моделей, які можуть допомогти в ухваленні рішень та інших бізнес-діях. Спеціаліст із обробки даних або інший член групи з вивчення даних також має донести результати до керівників компаній і користувачів, часто за допомогою візуалізації даних і використання методів оприлюднення даних [6].

Етапи побудови моделі методами Data Mining:

- постановка завдання, ідентифікація бізнес-проблеми, типові «вузькі місця» бізнесу;
- підготовка даних, визначення джерел даних для аналізу, вибір, очищення та попередня обробка даних;
- побудова моделей із зазначенням алгоритмів інтелектуального аналізу даних і їх параметри;
- перевірка моделі: оцінка якості роботи створеної моделі перед початком її використання в "виробничому середовищі" [4].

## 2.2 Інтелектуальні методи отримання інформації для аналізу

До методів отримання і аналізу даних відносять:

- описові - дисперсійний та регресійний аналіз;
- аналіз часових рядів; кластерний аналіз;
- еволюційні - дерева рішень;
- генетичні алгоритми;
- штучні нейронні мережі;
- метод «найближчих сусідів».

В процесі накопичення даних збільшується необхідність застосування комп'ютерних методів інтелектуального аналізу даних як ефективного допоміжного засобу, який дозволяє досліднику отримати додаткові знання з предметної галузі, в якій він працює і має приймати зважені і обґрунтовані рішення. Класичним прикладом, який вимагає застосування методів інтелектуального аналізу даних, є проблема зміни вимог клієнтів в залежності від актуальності типу замовлення на конкурентному ринку [7]. Наприклад, найпопулярнішим типом контенту для просування в соціальних мережах зараз є короткі відео із застосуванням анімації, які потребують навичок роботи із ПЗ Spine, After Effect, а ще два роки тому таким контентом був повністю 3D-анімований ролик за допомогою ПЗ Cinema 4D, 3DMAX. Ключем до розв'язання цієї проблеми є база даних профілів замовлень та виконавців. Моделі замовлень можна проаналізувати, щоб визначити їх відмінні характеристики. Якщо такі характеристики вдалося знайти, їх можна застосувати для ідентифікації нетипових замовлень або класифікувати їх за вартістю роботи.

Нижче наведені типові приклади задач для пояснення головних цілей інтелектуального аналізу даних.

2.2.1. Ухвалення рішення про розміщення та рекомендовану вартість замовлення (задача класифікації)

У системі фріланс-біржф для рекомендації вартості замовлення замовнику необхідно заповнити анкету, яка містить дані про замовлення. Ці дані приймаються до уваги як обґрунтування для прийняття рішення про розміщення замовленн та надання рекомендації щодо його вартості виконання. Такі рішення зазвичай приймаються в два етапи. Спершу застосовують статистичні методи для визначення однозначних рішень про «схвалення» або «відхилення» розміщення замовленн. Якщо замовлення не відхилено на першому етапі, далі включається більш детальний аналіз, який вимагає більш складного аналізу роботи і найчастіше вимагає залучення до процесу прийняття рішень експертів. В очевидних ситуаціях можна застосувати процедуру статистичного прийняття рішень і розрахунку певного числового параметра за інформацією, наданою в анкеті. Заявка вважається «схваленою», якщо статистичний параметр потрапляє у зазначений відрізок вартості і «відхиленою» в іншому випадку. У неочевидних випадках заявка направляється до співробітників фріланс біржі для досягнення компромісного рішення.

### 2.2.2. Прогнозування навантаження замовлень (задача прогнозування)

Стандартна модель навантаження включає три компоненти: базове навантаження за рік, періодичність замовлень протягом року і вплив свят. Кількість замовлень на сайті характеризується періодичністю: добовою – попит має ранній ранковий мінімум, екстремум ввечері і денний максимум; щотижневою – попит нижче по вихідних днях; сезонну – підвищений попит за 2 місяці до свят. Частотні ефекти враховуються шляхом реконструкції річного навантаження у вигляді послідовності днів з урахуванням свят, денормалізації навантаження для обліку загальної тенденції. Модель прогнозу розраховують за допомогою лінійного регресійного аналізу значень типу замовлення, рівня володіння ПЗ, об'єму замовлення, тощо, і знаходження оптимальних параметрів регресійної моделі.

### 2.2.3. Технічна діагностика (задача пошуку асоціативних правил)

Діагностика – одна з основних областей застосування експертних систем. Хоча складені вручну правила, що застосовуються в експертних системах, часто є цілком інформативними, але інтелектуальний аналіз даних може стати в нагоді в ситуаціях, коли об'єднання (додавання) таких правил може стати занадто трудомісткою працею або коли важко передбачити заздалегідь всі можливі комбінації вхідних даних та можливі запити. Правила діагностики можуть генеруватися за допомогою методу пошуку асоціативних правил.

### 2.2.4. Маркетинг та продажі (задача пошуку асоціативних правил)

Деякі з найбільш активних застосувань інтелектуального аналізу даних відносяться до сфери маркетингу і продажів, в якій компанії володіють величезними обсягами точно записаних даних. Інтелектуальний аналіз таких даних може визначати групи клієнтів, для яких будуть цікаві нові послуги. Аналіз передбачає застосування методів пошуку асоціативних правил для визначення груп замовлень, які додаються найчастіше.

## 2.3 Методи інтелектуального аналізу даних

Для інтелектуального аналізу даних для різних програм обробки даних можна використовувати різні методи (див.рисунок 2.1). Розпізнавання шаблонів — це звичайний випадок використання інтелектуального аналізу даних (data mining), який складається з декількох методів, та включає в себе в тому числі й виявлення аномалій, метою чого є визначення викидних значень у наборах даних.



Рисунок 2.1 – Популярні методи інтелектуального аналізу даних

Популярні методи аналізу даних включають наступні:

- визначення асоціацій. У інтелектуальному аналізі даних правила асоціації – це оператори «якщо-тоді», які визначають зв'язки між елементами даних. Критерії підтримки та достовірності використовуються для оцінки зв'язків – підтримка вимірює, як часто пов'язані елементи з'являються в наборі даних, тоді як достовірність відображає кількість разів, коли твердження «якщо-тоді» є точним.
- класифікація. Цей підхід призначає елементи в наборах даних до різних категорій. Деревя рішень, класифікатори Байеса, k-найближчий сусід і логістична регресія є прикладами методів класифікації.
- кластеризація. У цьому випадку елементи даних, які мають спільні характеристики, об'єднуються в кластери. Приклади такого методу включають кластеризацію k-середніх, ієрархічну кластеризацію та моделі Гауса.
- регресія. Це ще один спосіб знайти зв'язки в наборах даних шляхом обчислення прогнозованих значень даних на основі набору змінних. Прикладами є лінійна регресія та багатовимірна регресія. Деревя рішень і деякі інші методи класифікації також можна використовувати для виконання регресії.
- аналіз послідовності та шляху. Дані також можуть бути отримані для пошуку шаблонів, у яких певний набір подій або значень призводить до пізніших.
- нейронні мережі. Нейронна мережа – це набір алгоритмів, що моделює діяльність людського мозку. Нейронні мережі особливо корисні в складних

програмах розпізнавання образів, що включають глибоке навчання, більш просунуте відгалуження машинного навчання [6].

### 2.3.1 Визначення асоціацій

На відміну від методів побудови моделей класифікації, методи пошуку асоціативних правил не потребують вибору атрибуту класу, усі атрибути вважаються атрибутами ознак, класом є комбінація значень окремих атрибутів ознак. Методи побудови асоціативних правил призначені для пошуку комбінацій значень атрибутів, на базі яких за комбінацією значень атрибутів першої множини («умова») можна спрогнозувати значення атрибутів другої множини («наслідок»). Пошук асоціативних правил часто виконують супермаркети під час аналізу споживчого кошику для визначення продуктів, які покупці часто купують разом, після чого знайдені комбінації товарів розміщують поруч, щоб збільшити ймовірність їх купівлі.

Асоціативне правило складається з двох множин  $X$  («умова») і  $Y$  («наслідок») у вигляді конструкції IF-THEN:  $X \rightarrow Y$ , тобто, якщо знайдені значення атрибутів множини  $X$ , тоді з ним ймовірно будуть знайдені значення атрибутів множини  $Y$ .

У таблиці 2.1 представлений спрощений приклад даних про рекомендації вартості роботи та строку виконання. В рядках таблиці наведені дані про замовлення, тобто параметри, на підставі яких робляться висновки (в термінології інтелектуального аналізу даних, атрибути ознак). На підставі цих даних замовник може зробити висновки щодо призначення грошової винагороди та термін виконання замовлення. Рекомендації представлені в останньому стовпчику кожного рядка. Тобто кожен рядок таблиці є прикладом певного набору атрибутів ознак замовлення та рекомендації (атрибуту класу).

Таблиця 2.1 містить 8 рядків, що включають два можливих значення для атрибуту тип замовлення, два значення для атрибуту наявності анімації чи її відсутності, два значення для атрибуту типу ПЗ, яке використовується для

виконання замовлення, і два значення для атрибуту рівня володіння необхідним ПЗ. Кількість можливих комбінацій значень для чотирьох атрибутів дорівнює  $2 \times 2 \times 2 \times 2 = 16$ . Вивчивши дані можна сформулювати асоціативні правила для рекомендованої вартості виконання замовлення (див. Лістинг 2.1).

Таблиця 2.1 – Приклад застосування методу асоціативних правил

Тип замовлення	Анімація	ПЗ	Рівень володіння	Рекомендована вартість/h роботи
Відео	ні	Sony Vegas	basic	10-20
Відео	ні	Sony Vegas	pro	21-40
Відео	ні	After Effect	basic	15-30
Відео	так	After Effect	pro	31-70
Графіка	ні	Adobe Illustrator	basic	5-14
Графіка	ні	Adobe Illustrator	pro	15-30
Графіка	так	Spine	basic	10-20
Графіка	так	Spine	pro	21-40

Лістинг 2.1. Приклад асоціативного правила

*ЯКЩО Тип замовлення = = Відео ТА Анімація = = так ТОДІ  
Рекомендована вартість/h роботи = 31 - 70\$*

Крім асоціативних правил, поширеним способом наочного представлення закономірностей є дерева рішень, що містять в вузлах питання, а в листах – висновки. У більшості ситуацій, пов'язаних з навчанням, набір прикладів, наведених як вхідні дані, є в найкращому випадку таким, який дозволяє робити узагальнюючі висновки, але не містить всіх можливих ситуацій, тобто в цьому сенсі є неповним, і тому аналіз процесу, який досліджується, полягає у створенні його моделі на підставі наявних прикладів (в теорії штучного інтелекту такий підхід називається навчанням з вчителем, тобто іншими словами, побудовою моделей на підставі екземплярів даних, які доступні для дослідника). В реальних

ситуаціях задача побудови моделей на підставі навчання ускладнюється тим, що в наборах даних можуть бути приклади, в яких значення деяких атрибутів невідомі – наприклад, дані про вимірювання не були записані або були втрачені. Також часто трапляються некоректні класифікації через помилки або похибки в даних, необхідних для побудови моделі класифікації. [4]

### 2.3.2 Класифікація

Метод класифікації – це метод аналізу даних, який дозволяє оцінити ймовірність приналежності екземплярів даних до деякого класу залежно від значень їх атрибутів. Як модель класифікації рекомендується використовувати структуру даних «дерево» (Рисунок 2.1), в якій кожен вузол являє собою точку прийняття рішення на підставі значень атрибутів даних, що класифікуються.

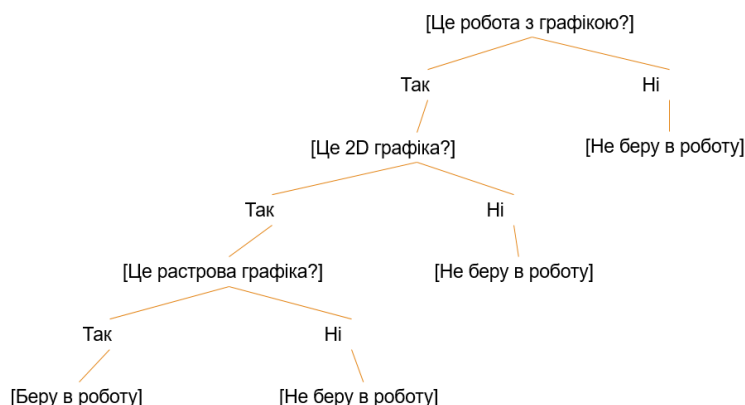


Рисунок 2.2 – Приклад дерева класифікації

На рисунок 2.2 наведено спрощене дерево класифікації, яке дає відповідь на питання «Ви візьмете завдання в роботу?». У кожному вузлі ставиться уточнююче питання (до атрибуту) з відповідями (значення атрибуту) у гілках, відповідаючи ви переходите до наступного вузла (питання, атрибуту) до тих пір, поки не дійдете до листа зі значенням класу, у прикладі це відповіді «Беру в роботу» чи «Не беру в роботу». Перевага класифікаційних дерев полягає у тому, що вони не вимагають надмірної кількості інформації для побудови досить

точного та інформативного дерева рішень. Метод класифікації використовує відомі значення атрибутів екземплярів даних та зв'язки між їх значеннями при побудові моделі класифікації. При наявності нових екземплярів даних невідомого класу, до даних застосовується раніше побудована модель класифікації і визначається відповідний клас.

### 2.3.3 Метод кластеризації

Метод кластеризації – це метод аналізу даних, який дозволяє розділити екземпляри даних за значеннями їх атрибутів на класи, кожен з яких має певні ознаки. Кластерний аналіз використовується в тих випадках, коли необхідно автоматично виділити деякі правила, взаємозв'язки або тенденції у сукупності даних. Як модель кластеризації можна використовувати двовимірний простір Евкліда, в якому відносне скупчення екземплярів даних являє собою певний клас. Просторовий підхід використовує некласифіковані екземпляри для аналізу зв'язків між значеннями їх атрибутів. Побудовану модель кластеризації можна використовувати для оцінювання приналежності нових екземплярів даних до вже відомих кластерів.

Набір даних, який буде застосований для виконання кластерного аналізу, містить інформацію про всі замовлення, про фрілансерів, які ними зацікавилися, про те, наскільки часто фрілансери брали цікаві їм замовлення в роботу. Тепер треба проаналізувати ці дані, щоб виділити різні групи замовлень і зрозуміти, чи можна визначити деякі тенденції в їх розміщенні.

Метод кластеризації використовує відомі дані для аналізу зв'язків значень атрибутів. Коли з'являється новий екземпляр даних, потрібно лише оцінити дані за допомогою побудованої раніше моделі кластеризації та визначити приналежність даних до певного кластеру. Кластеризація відрізняється від класифікації тим, що самі групи заздалегідь не задані. За допомогою кластеризації засобів Data Mining самостійно виділяють різні однорідні групи даних

### 2.3.4 Модель регресійного аналізу

Модель регресійного аналізу використовується для прогнозування значення однієї залежної змінної, виходячи з відомих значень незалежних змінних. Наприклад ціна виконання замовлення (залежна змінна) визначається наступними незалежними змінними (значеннями атрибутів екземплярів даних): наявність анімації, необхідним для виконання ПЗ, рівень володіння ним і т.д. Для побудови регресійної моделі оцінюються параметри схожих замовлень та їх ціна, за якою ці замовлення були виконані (тобто формується навчальна вибірка). Після побудови моделі вказуються значення атрибутів замовлення та розраховується вартість.

### 2.3.5 Аналіз послідовності та шляху

Дані також можуть бути отримані для пошуку шаблонів, у яких певний набір подій або значень призводить до пізніших.

Ідея систем case based reasoning – CBR – на перший погляд вкрай проста. Для того, щоб зробити прогноз на майбутнє чи вибрати правильне рішення ці системи знаходять у минулому близькі аналоги наявної ситуації і вибирають ту ж відповідь, який був для них правильним. Тому цей метод ще називають методом «найближчого сусіда» (nearest neighbour). Останнім часом поширення отримав також термін memory based reasoning, який акцентує увагу, що рішення приймається на підставі всієї інформації, накопиченої в пам'яті.

Системи CBR показують непогані результати в найрізноманітніших задачах. Головним їх мінусом вважають те, що вони взагалі не створюють будь-яких моделей або правил, узагальнюючих попередній досвід у виборі рішення вони ґрунтуються на всьому масиві доступних історичних даних, тому неможливо сказати, на основі яких конкретно факторів CBR системи будують свої відповіді. Інший мінус полягає в свавіллі, який допускають системи CBR

при виборі міри «близькості». Від цієї міри найрішучішим чином залежить обсяг безлічі прецедентів, які потрібно зберігати в пам'яті для досягнення задовільною класифікації або прогнозу. Приклади систем, що використовують CBR – KATE tools (Acknosoft, Франція), Pattern Recognition Workbench (Unica, США).

Дерева рішень (decision trees) є одним з найбільш популярних підходів до вирішення завдань Data Mining. Вони створюють ієрархічну структуру правил типу «ЯКЩО... ТО...» (if – then), що має вигляд дерева. Для прийняття рішення, до якого класу віднести деякий об'єкт або ситуацію, потрібно відповісти на питання, що стоять у вузлах цього дерева, починаючи з його кореня. Запитання мають вигляд «значення параметра А більше х?». Якщо відповідь позитивна, здійснюється перехід до правого вузла наступного рівня, якщо негативна – то до лівого вузла; потім знову слід питання, пов'язане з відповідним вузлом. Популярність підходу пов'язана як би з наочністю і зрозумілістю. Але дерева рішень принципово не здатні знаходити «кращі» (найбільш повні і точні) правила в даних. Вони реалізують наївний принцип послідовного перегляду ознак, створюючи лише ілюзію логічного висновку.

### 2.3.6 Нейронні мережі

Нейронні мережі – це великий клас систем, архітектура яких має аналогію (як тепер відомо, досить слабку) з побудовою нервової тканини з нейронів. В одній із найбільш поширених архітектурі зі зворотним поширенням помилки імітується робота нейронів у складі ієрархічної мережі, де кожен нейрон більш високого рівня з'єднаний своїми входами з виходами нейронів нижчого шару. На нейрони самого нижнього шару подаються значення вхідних параметрів, на основі яких потрібно приймати якісь рішення, прогнозувати розвиток ситуації тощо. Ці значення розглядаються як сигнали, що передаються в наступний шар, ослабляючись або посилюючись в залежності від числових значень (ваг), приписуваних дугами між нейронними зв'язками. У результаті на виході нейрона верхнього шару виробляється деяке значення, яке розглядається як

відповідь – реакція всієї мережі на введені значення вхідних параметрів. Для того щоб мережу можна було застосовувати надалі, її треба «натренувати» на отриманих раніше даних, для яких відомі і значення вхідних параметрів, і правильні відповіді на них. Тренування полягає в підборі ваг, що забезпечують найбільшу близькість відповідей мережі до відомих правильних відповідей.

Основним недоліком нейронно-мережевої парадигми є необхідність мати дуже великий обсяг навчальної вибірки. Інший суттєвий недолік полягає в тому, що навіть натренована нейронна мережа являє собою чорний ящик. Знання, зафіксовані як ваги, абсолютно не піддаються аналізу та інтерпретації людиною (відомі спроби дати інтерпретацію структурі налаштованої нейронної мережі виглядають непереконливими – система «KINOsuite – PR»).

### 2.3.6 Генетичні алгоритми

Окрім нейромереж існують генетичні алгоритми, їх потрібно розглядати скоріше як потужний засіб вирішення різноманітних комбінаторних завдань і завдань оптимізації. Проте генетичні алгоритми увійшли наразі в стандартний інструментарій методів Data Mining.

Перший крок при побудові генетичних алгоритмів – це кодування вихідних логічних закономірностей у базі даних, які іменують хромосомами, а весь набір таких закономірностей називають популяцією хромосом. Далі для реалізації концепції відбору вводиться спосіб зіставлення різних хромосом, який здійснюється за допомогою процедур репродукції, мінливості (мутацій), генетичної композиції. Ці процедури імітують біологічні процеси. Найбільш важливі серед них: випадкові мутації даних в індивідуальних хромосомах, переходи (кросинговер) і рекомбінація генетичного матеріалу, що міститься в індивідуальних батьківських, та міграції генів. У ході роботи процедур на кожній стадії еволюції виходять популяції з усе більш досконалішими індивідуумами.

Генетичні алгоритми зручні тим, що їх легко розпаралелювати. Наприклад, можна розбити покоління на кілька груп і працювати з кожною з них незалежно, обмінюючись час від часу кількома хромосомами. Існують також і інші методи розпаралелювання генетичних алгоритмів. Генетичні алгоритми мають ряд недоліків. Критерій відбору хромосом і використовувані процедури є евристичними і далеко не гарантують знаходження «кращого» рішення. Як і в реальному житті, еволюцію може «заклинити» на яку-небудь непродуктивну гілку. І, навпаки, можна навести приклади, як два неперспективних батька, які будуть виключені з еволюції генетичним алгоритмом, виявляються здатними призвести високоефективного нащадка. Це особливо стає помітно при вирішенні високо розмірних завдань зі складними внутрішніми зв'язками.

Одним з прикладів використання методів інтелектуального аналізу даних у системі фріланс-біржі для 2D/3D Generalis є виявлення навмисного викривлення вартості та часу виконання замовлення. Така методика може бути корисною, наприклад, як підстава для оцінки складності замовлення та рівня фахівця, який може його виконати. По-перше, створюється профіль виконавця та його hard skills, який описує набір ПЗ, з яким має бути досвід для виконання замовлення. По-друге, відбувається кластеризація даних, які є необхідними для опису завдання та рівня фахівця. За вибраними критеріями встановлюються випадки шахрайства (в даному випадку – заниження вартості виконання завдання або ж неприйнятні терміни виконання завдання), які виходять за межі погрішності більше, ніж на стандартне відхилення. На підставі вищезазначеного можна побудувати алгоритм аналізу даних у системі фріланс-біржі для 2D/3D Generalis, що поєднає традиційні форми аналізу та сучасні методи інтелектуального аналізу даних.

### 3 СТРУКТУРНЕ ПРОЄКТУВАННЯ ФІЛЬТРА ДЛЯ ВИЗНАЧЕННЯ ВАРТОСТІ ВИКОНАННЯ ПРОЕКТУ ФРІЛАНСЕРОМ

Для інтелектуального аналізу даних для різних програм обробки даних можна використовувати різні методи. В поточній роботі використовується метод Кластеризації (у цьому випадку елементи даних, які мають спільні характеристики, об'єднуються в кластери – типи завдань) та методи пошуку асоціативних правил. З урахуванням експертної оцінки для всіх видів робіт визначається інтервал з максимальною та мінімальною вартістю робіт, до якого має належати вартість конкретної роботи.

#### 3.1 Кластерний аналіз в контексті задачі прогнозування вартості виконання проекту фрілансером

У випадку фріланс-біржі для художників, кластерний аналіз може допомогти зрозуміти, які характеристики впливають на вартість виконання завдань.

В контексті задачі прогнозування вартості завдань на фріланс-біржі для художників, кластер – це група схожих або подібних завдань або проєктів. Метод кластеризації допомагає групувати завдання, які мають спільні характеристики або особливості, у певні категорії або "кластери".

Кластери у цьому випадку можуть вказувати на різні сегменти ринку або типи завдань, які можуть мати схожу цінову динаміку чи інші спільні риси.

Кроки застосування алгоритму кластеризації:

1) збір даних: зберіть дані про проєкти на фріланс-біржі, включаючи такі параметри, як:

- термін виконання;
- тип графіки, з яким працюють;

2) підготовка даних: обробіть та очистіть дані;

3) вибір алгоритму: виберіть алгоритм кластеризації, який найкраще відповідає вашій задачі;

4) застосування алгоритму: застосуйте алгоритм до підготовлених даних, кластеризуйте проекти на групи на основі подібності їхніх характеристик;

5) оцінка результатів: оцініть, наскільки ефективно алгоритм виокремив схожі групи, використовуйте метрики кластеризації, такі як силует, для визначення якості розділення;

6) інтерпретація кластерів: проаналізуйте отримані кластери.

Розглянемо конкретний приклад застосування алгоритму кластеризації для прогнозування вартості завдань на фріланс-біржі для художників.

**Сценарій:** на фріланс-біржі для художників є безліч проектів з малюнків, графіки, ілюстрацій та інших творчих завдань. Ми хочемо вивчити характеристики цих завдань і спробувати виділити схожі групи проектів на основі їхньої вартості та інших параметрів.

Результат розділення проектів на 3 кластери на основі Видів цифрових замовлень та термінів виконання може виглядати так:

- кластер 1: 2D проекти з коротким терміном виконання;
- кластер 2: 3D проекти з середнім терміном виконання;
- кластер 3: Відео проекти з тривалим терміном виконання.

Цей підхід допомагає зрозуміти, які фактори впливають на вартість та терміни виконання проектів та виявити групи подібних проектів на основі цих факторів.

Застосування кластерного аналізу для фріланс-біржі може допомогти розмежувати художників за певними характеристиками та робити більш точні прогнози вартості їхніх завдань.

Наприклад, можливі такі кластери для видів цифрових замовлень:

1) вид графіки:

- 2D;
- 3D;
- відео;

## 2) терміни виконання:

- термінові завдання;
- проекти із гнучкими термінами;
- довготривалі проекти.

У таблиці 3.1 показано застосування кластерного аналізу для замовлень із 2D графікою.

Таблиця 3.1 – Застосування кластерного аналізу для замовлень із 2D графікою

Напрямок	Ретуш (робота із фотографією)
	Цифровий Художник
	Графічний дизайнер
	Web-дизайнер
Тип графіки	Векторна
	Растрова
Рівень експерта	Базовий (вміє робити базові речі редагування з використанням одного стандартного інструмента або задачі, які складаються з декількох простих ітерацій)
	Середній (вміє комбінувати стандартні прості ітерації, працювати за наявним прикладом або алгоритмом, має знання з плагінів)
	Експерт (вміє комбінувати складні ітерації, працювати із сторонніми плагінами, створювати свої моделі, розробляти з нуля)
Розмір замовлення	Малі проекти (до 3 одиниць)
	Середні проекти (від 4 до 10 одиниць)
	Великі проекти (від 10 одиниць)
Тип мистецтва	Схематичне, спрощене, набросок
	Стилізація

Продовження таблиці 3.1

	Реалізм
	Манга, комікси
	Акварельна техніка
	Імітація олійних фарб
Види елементів дизайну	Статичні
	Використання анімації та ефектів

Для замовлень із 2D графікою за напрямком Ретуш (робота із фотографією) можна виділити такі типи робіт (див.рисунок 3.1):

- обробка портретних фотознімків;
- обробка каталожних фото;
- журнальна обробка фото;
- обробка весільних фото;
- колажування.

Для замовлень із 2D графікою за напрямком Цифровий Художник можна виділити такі типи робіт (див.рисунок 3.2):

- концепт-арт (Concept Artist) – (візуалізує ідеї персонажів, будівель, предметів, локацій у вигляді нарисів);
- маркетинговий арт – створення арту обраного розміру для соціальних мереж, банерів ігор, прилоудерів, також сюди можна віднести і арти під друк на холсті по фотографії;
- оточення та фони – створення елементів оточення, наприклад, пропсів та окремих деталей ігрового середовища, об'єктів, іконок.

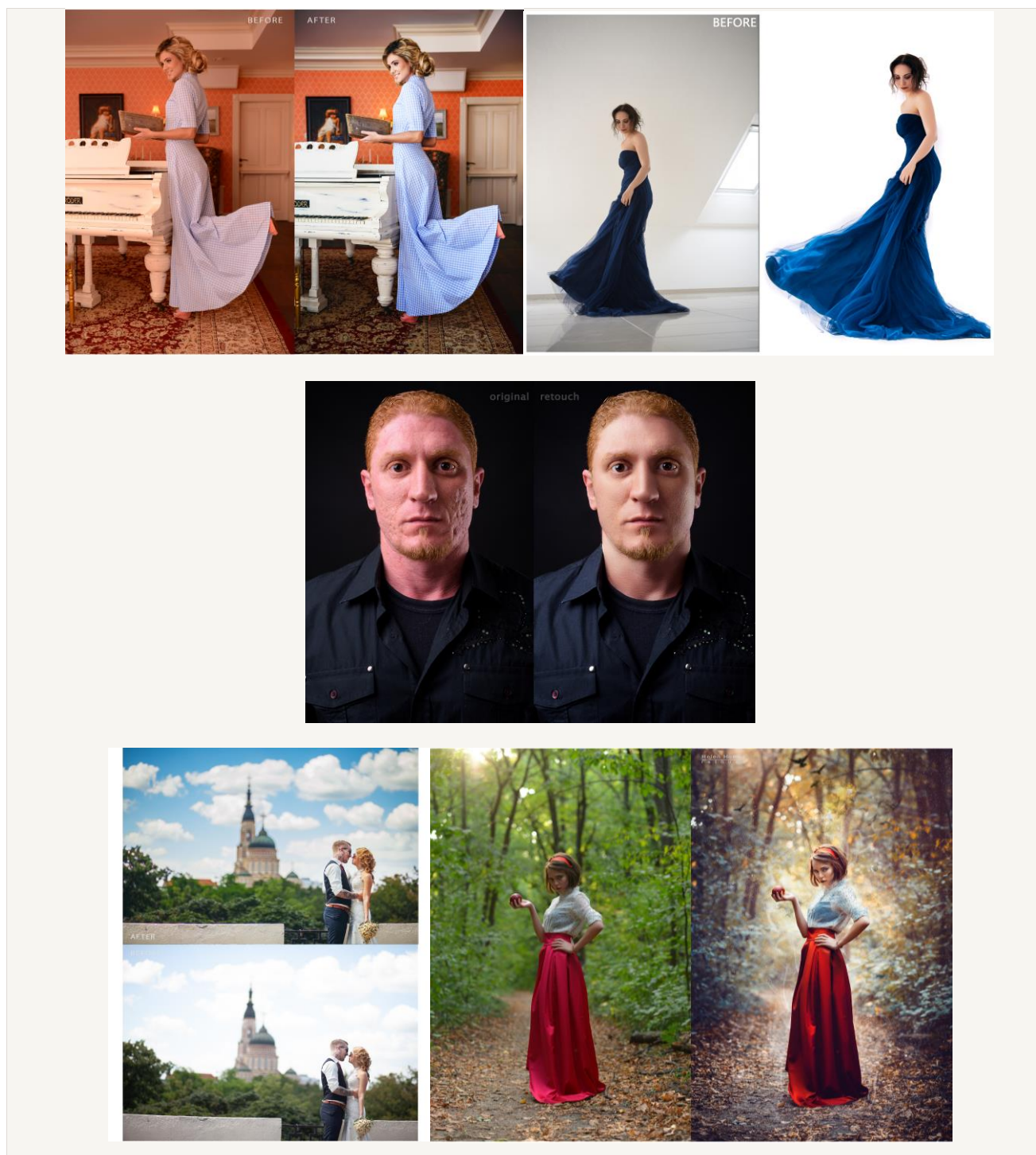


Рисунок 3.1 – Приклади типів робіт для замовлень із 2D графікою за напрямком Ретуш

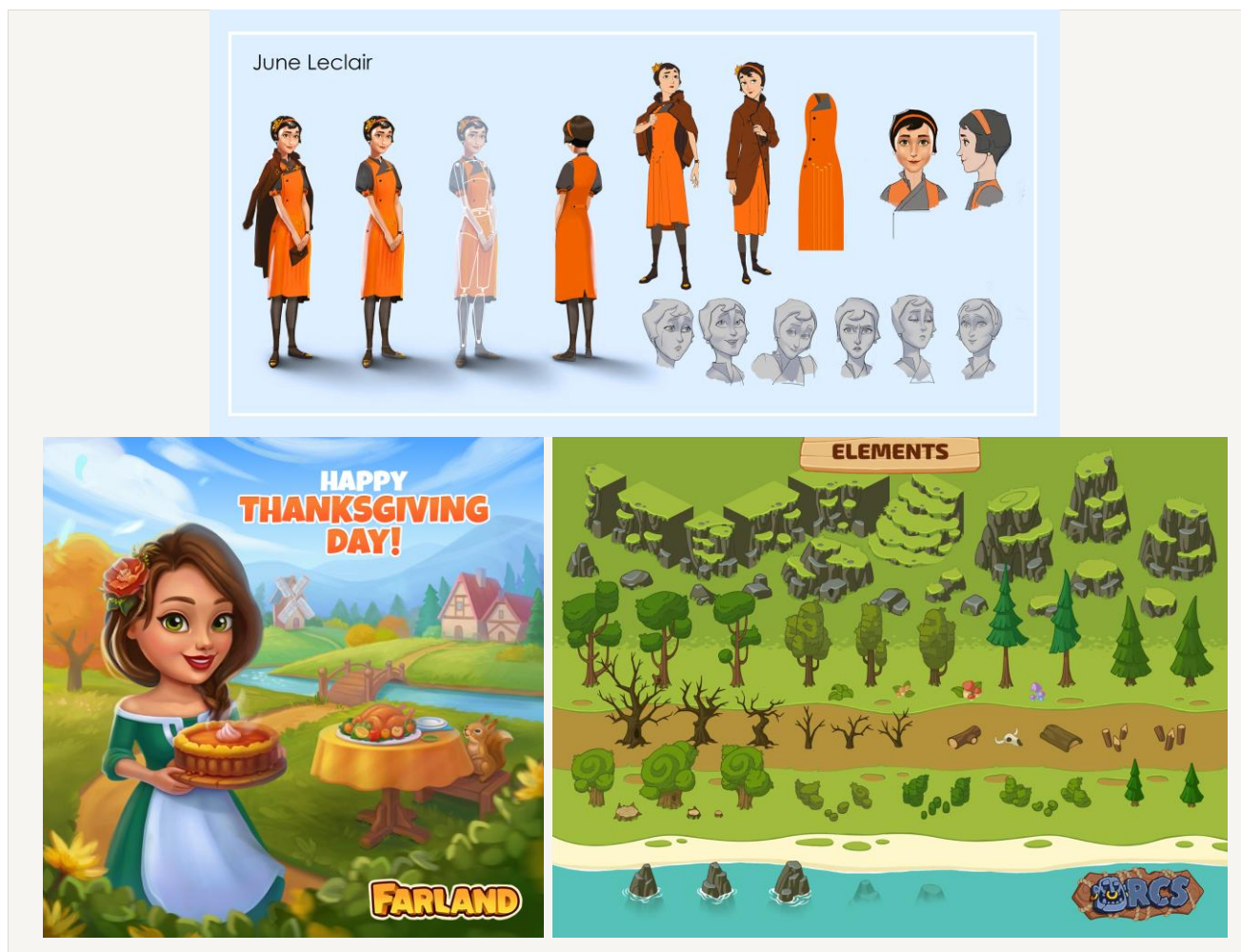


Рисунок 3.2 – Приклади типів робіт для замовлень із 2D графікою за напрямком Цифровий Художник

Для замовлень із 2D графікою за напрямком Графічний дизайнер можна виділити такі типи робіт (див.рисунок 3.3):

- розробка логотипів та брендингу;
- створення рекламних матеріалів (банери, плакати, флаєри):
- графічні елементи для соціальних мереж та медіа-контент.

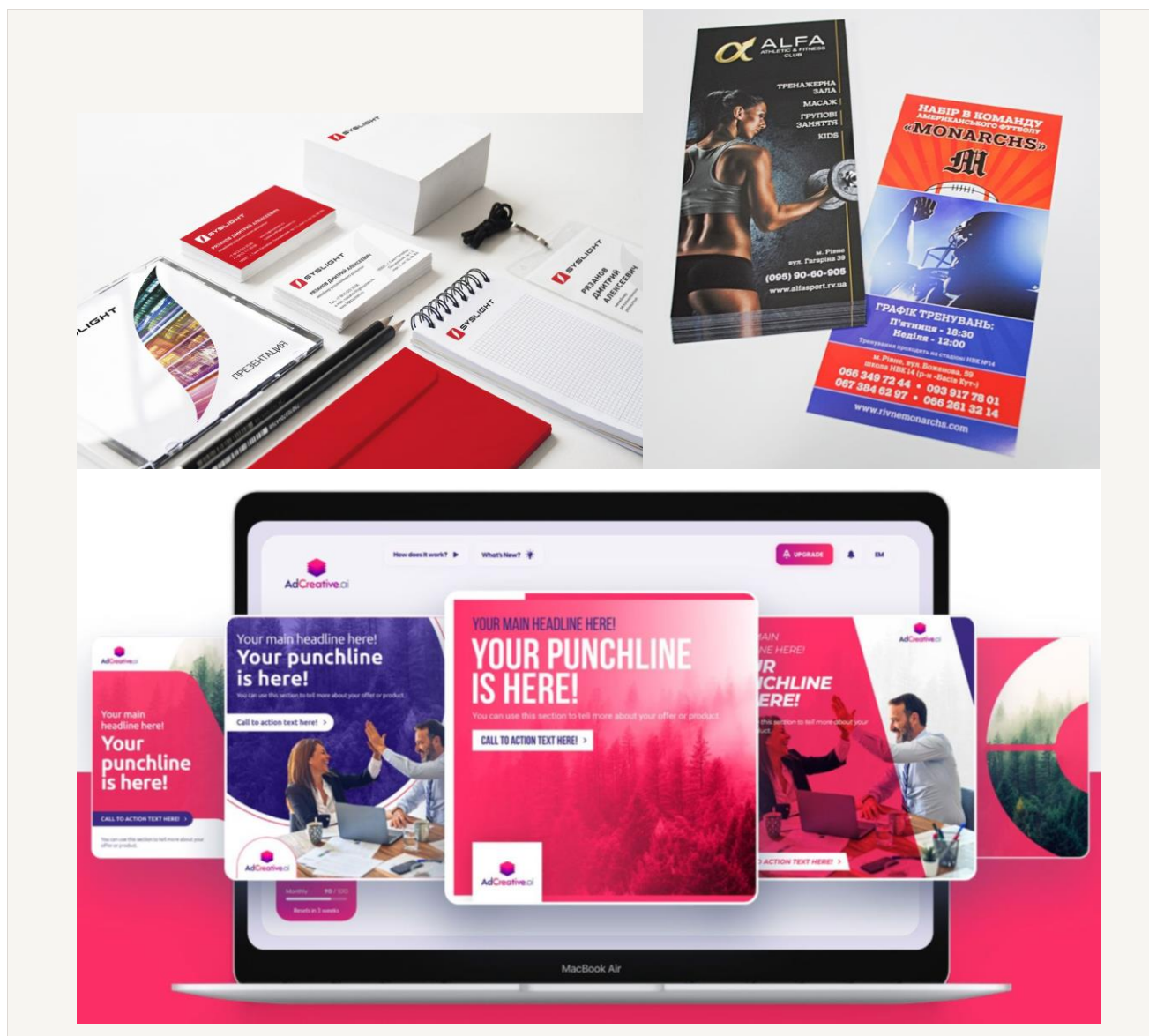


Рисунок 3.3 – Приклади типів робіт для замовлень із 2D графікою за напрямком Графічний дизайнер

Для замовлень із 2D графікою за напрямком Web-дизайнер можна виділити такі типи робіт (див.рисунок 3.4):

- створення статичного (фіксований) дизайну (Веб-сайт має фіксовану ширину і не змінюється відповідно до розміру екрану користувача);
- створення адаптивного (резиновий) дизайну (Веб-сайт адаптується до різних розмірів екрану, щоб забезпечити оптимальний перегляд на різних пристроях);

- створення мобільного дизайну (Спеціально оптимізований для використання на мобільних пристроях, зокрема на смартфонах і планшетах);
- створення односторінкового дизайну, Single Page Design (Створення веб-сайту, який складається з одного документу, розгорнутого по вертикалі чи горизонталі).

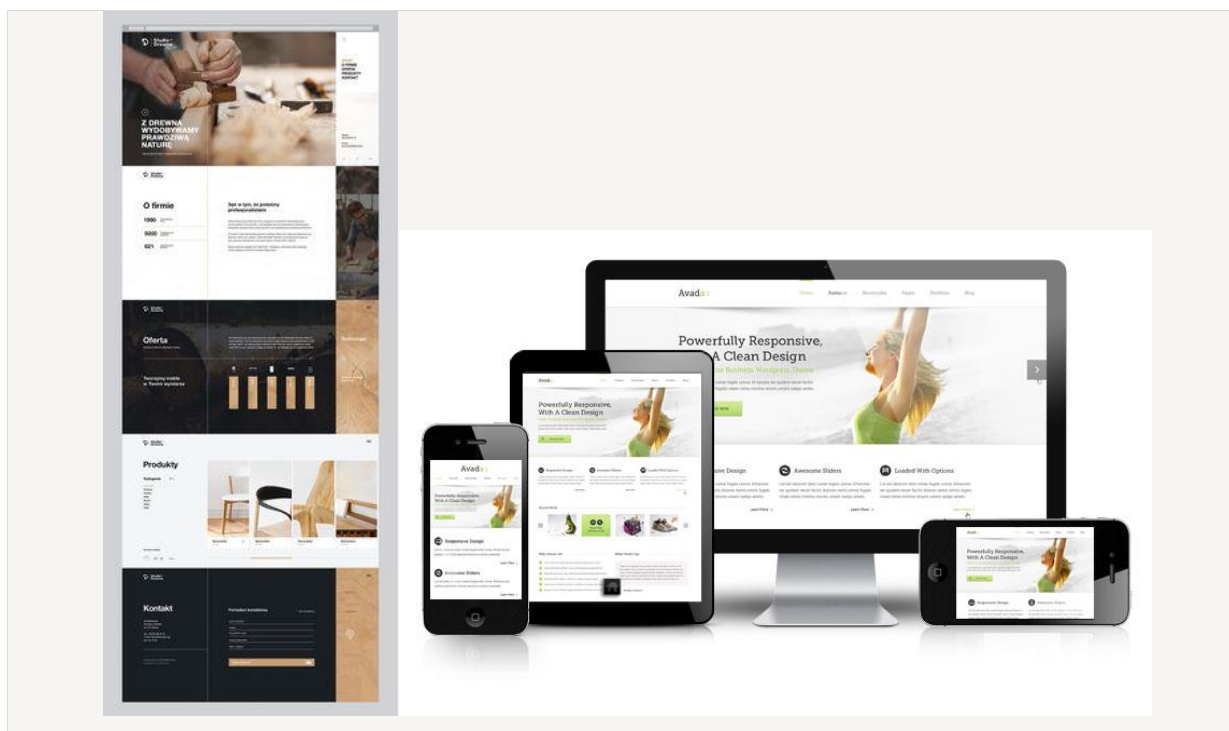


Рисунок 3.4 – Приклади типів робіт для замовлень із 2D графікою за напрямком Web-дизайнер

Опис застосування кластерного аналізу для замовлень із 3D графікою надано у таблиці 3.2.

Таблиця 3.2 – Застосування кластерного аналізу для замовлень із 3D графікою

Напрямок	3D-модельовання об'єктів та сцен (Modeller)
	Розробка 3D-графіки для VR&AR
	Візуалізація архітектурних об'єктів
	Створення текстур (Texture Artist)

Продовження таблиці 3.2

Стиль	Стилізовані об'єкти
	Схематичні або спрощені об'єкти
	Реалізм
	Гіперреалізм
Деталізація	Low-poly
	High-poly
Робота з моделлю	Створення з нуля
	Редагування існуючої основи

Для замовлень із 3D графікою можна виділити такі напрямки:

– 3D-модельовання об'єктів та сцен (Modeller) – промальовує в 3D моделі персонажів, їх одягу, зброї, техніки, будівель, ландшафту. Моделери можуть спеціалізуватися більше на техніці чи органіці. Для цієї спеціальності потрібно опанувати актуальні пакети тривимірної графіки (наприклад, Maya, Blender, 3D Max) і хоча б один двомірний пакет (Photoshop, Painter).

– розробка 3D-графіки для VR&AR;

– візуалізація архітектурних об'єктів;

– створення текстур (Texture Artist) - створює текстури. Зазвичай художник текстурами отримує готову тривимірну модель і накладає неї потрібні текстури в потрібних місцях.

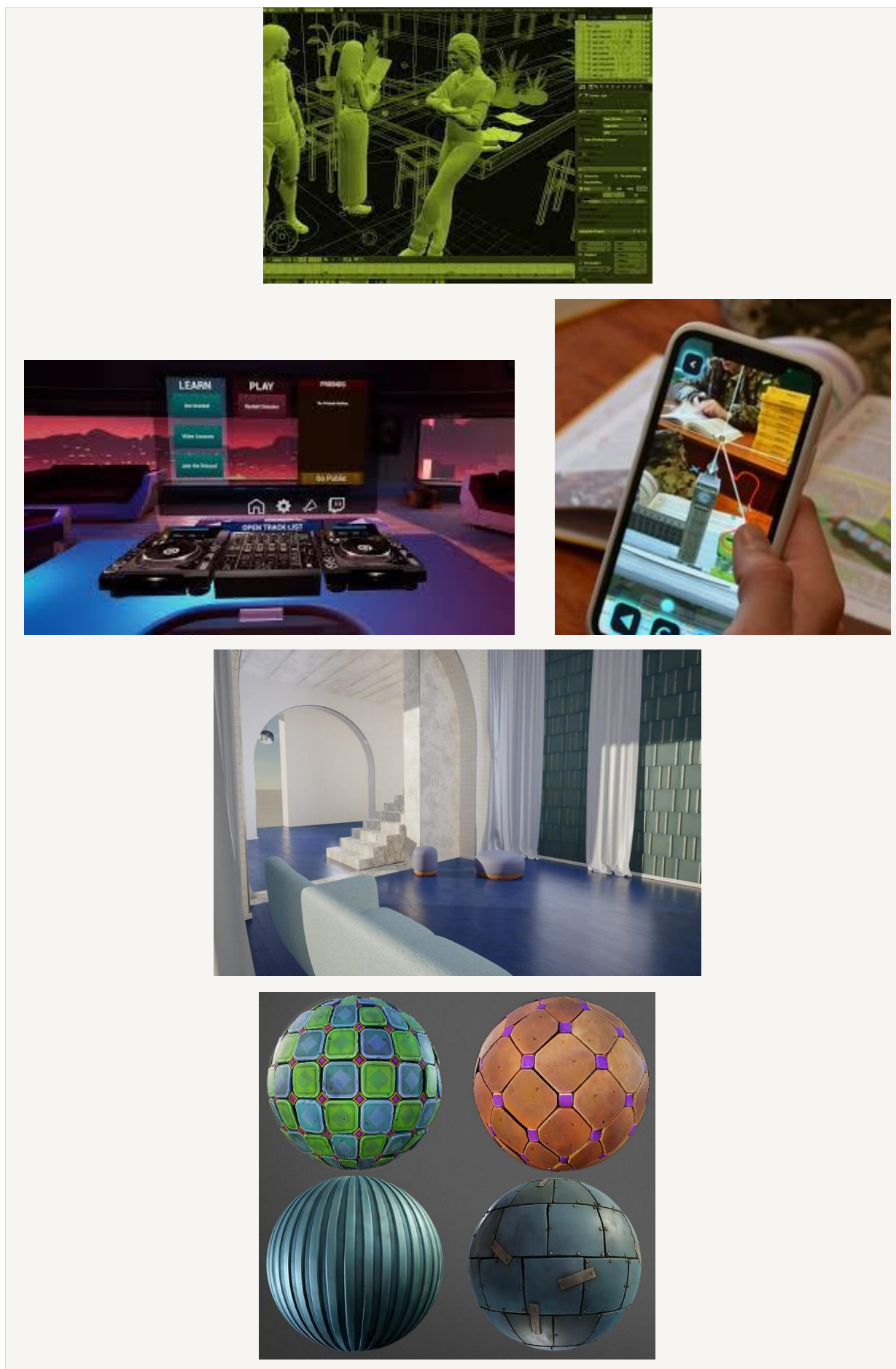


Рисунок 3.4 – Приклади типів робіт для замовлень із 3D графікою

Для замовлень із Відео можна виділити такі напрямки (див.табл.3.3):

- редагування відеопотоку (обрізка, комбінація, додавання музики або звуків)
- анімація (Animator) – анімація 3D-моделей, промальовка руху: ходьбу, біг, бій, а ще – вирази осіб персонажів;
- створення VFX ефектів (FX Artist, FX - від Effects) - створення ефектів на кшталт вибухів, іскор та ін. І оптимізує їх під різні пристрої. Фахівець з ефектів має вміти моделювати, працювати з текстурами, розуміти, як працюють шейдери, розумітися на анімації;
- створення Motion design.

Таблиця 3.3 – Застосування кластерного аналізу для замовлень із Відео

Тип роботи	Редагування відеопотоку (обрізка, комбінація, додавання музики або звуків)
	Анімація
	VFX ефекти
	Motion design
Тривалість ролика	Короткий (до 15 сек)
	Середній (від 15 сек до 2 хв)
	Довгий (більше 2 хв)
Тип вихідного файлу	Мр3, VMI, mov
	gif

Кластеризація може допомогти в ліпшому розумінні внутрішньої структури ринку та підготовці моделей прогнозування, які можуть бути більш точними для конкретних кластерів. Наприклад, модель для прогнозування ціни проекту для створення ілюстрацій для дитячих книг може враховувати інші фактори порівняно з моделлю для прогнозування вартості написання муралу.

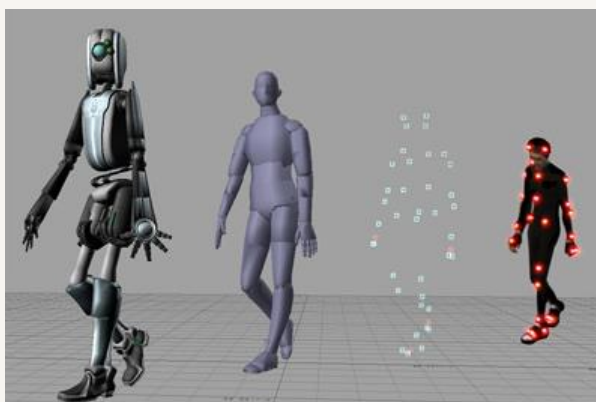


Рисунок 3.4 – Приклади типів робіт для замовлень із Відео

### 3.2 Метод асоціативних правил для розрахунку вартості проекту

Метод асоціативних правил – це техніка аналізу даних, яка дозволяє виявляти цікаві зв'язки та залежності між елементами в наборі даних. Використовується для виявлення асоціацій між різними елементами в транзакційних даних або базах даних.

Основні поняття та етапи методу асоціативних правил:

1) правило асоціацій: логічне висновок або зв'язок між одними чи декількома елементами даних. Приклад: Якщо клієнт купує хліб, то ймовірно, він також купить масло;

2) Ітем (елемент): кожен унікальний об'єкт або подія, який може бути включений в аналіз. У контексті асоціативних правил ітемами можуть бути товари, послуги або будь-які інші елементи;

3) підтримка (Support): вказує на частку вибірки, яка містить певний набір елементів. Висока підтримка означає, що ці елементи часто зустрічаються разом.

4) достовірність (Confidence): ймовірність того, що правило асоціацій буде виконане в конкретній ситуації;

Підтримка або довжина правила (Support or Rule Length): Кількість елементів, які включаються в правило.

Етапи роботи з методом асоціативних правил:

1) визначення підтримки та достовірності: встановлення порогових значень для підтримки та достовірності правил.

2) виявлення всіх наборів елементів: знаходження всіх можливих наборів елементів, які задовольняють встановлені критерії підтримки та достовірності;

3) генерація правил асоціацій: створення правил, які відображають асоціації між елементами;

4) оцінка та фільтрація правил: оцінка і фільтрація правил згідно з встановленими метриками;

5) використання отриманих правил: Використання знайдених асоціацій для прийняття рішень, наприклад, для персоналізації пропозицій клієнтам,

рекомендацій товарів і т. д.

Метод асоціативних правил застосовується у багатьох галузях, таких як роздрібна торгівля, маркетинг, медицина, аналіз великих даних та інші.

Метод асоціативних правил, який в основному використовується для аналізу зв'язків між елементами в транзакційних даних, не є типовим інструментом для визначення вартості проекту. Зазвичай вартість проекту оцінюється на основі різноманітних факторів, таких як трудомісткість, складність завдань, рівень експертизи, строк виконання і так далі.

Проте можливість використання асоціативних правил для визначення вартості проекту може бути реалізована у наступний спосіб:

1) аналіз схожих проектів: використання методу асоціативних правил для аналізу схожих проектів та виявлення зв'язків між їх елементами. Наприклад, можна аналізувати, які аспекти проектів часто взаємодіють або взаємодіють між собою;

2) оцінка впливу факторів: визначення, які фактори чи елементи впливають на вартість проекту. Це може допомогти виявити ключові аспекти, які можуть вплинути на загальну вартість;

3) прогнозування вартості на основі зв'язків: якщо вдало виявлено асоціації між різними елементами проекту, можна використовувати ці зв'язки для прогнозування вартості проекту на основі наявних даних.

Хоча цей підхід може бути цікавим для деяких аспектів проектного управління, важливо пам'ятати, що вартість проекту є складним і багатоаспектним питанням, і розглядати його виключно через призму асоціативних правил може бути недостатнім.

Ми маємо зведену таблицю напрямків арт замовлень та їх характеристик, та оцінки вартості таких замовлень за допомогою експертів (див.табл.3.4).

Таблиця 3.4 – Експертні оцінки для різних замовлень із 2D графікою

Напрямок	Рівень виконавця	Тип мистецтва	Анімація	\$	
Ретуш	Базовий	Реалізм	ні	2	5
Ретуш	Середній	Реалізм	ні	6	15

Продовження таблиці 3.4

Ретуш	Середній	Стилізація	ні	8	17
Ретуш	Експерт	Реалізм	ні	15	35
Ретуш	Експерт	Стилізація	ні	30	50
Цифровий Художник	Базовий	Реалізм	ні	6	15
Цифровий Художник	Базовий	Стилізація	ні	7	17
Цифровий Художник	Базовий	Акварельна техніка	ні	5	17
Цифровий Художник	Базовий	Імітація олійних фарб	ні	6	18
Цифровий Художник	Середній	Реалізм	ні	10	25
Цифровий Художник	Середній	Стилізація	ні	11	17
Цифровий Художник	Середній	Акварельна техніка	ні	9	22
Цифровий Художник	Середній	Імітація олійних фарб	ні	11	17
Цифровий Художник	Експерт	Реалізм	ні	17	30
Цифровий Художник	Експерт	Стилізація	ні	15	30
Цифровий Художник	Експерт	Акварельна техніка	ні	15	27
Цифровий Художник	Експерт	Імітація олійних фарб	ні	45	70
Графічний дизайнер	Базовий	Реалізм	ні	6	15
Графічний дизайнер	Базовий	Реалізм	так	7	17
Графічний дизайнер	Базовий	Стилізація	ні	5	17
Графічний дизайнер	Базовий	Стилізація	так	6	18
Графічний дизайнер	Середній	Реалізм	ні	10	25
Графічний дизайнер	Середній	Реалізм	так	11	17
Графічний дизайнер	Середній	Стилізація	ні	9	22
Графічний дизайнер	Середній	Стилізація	так	11	17
Графічний дизайнер	Експерт	Реалізм	ні	17	30
Графічний дизайнер	Експерт	Реалізм	так	15	30
Графічний дизайнер	Експерт	Стилізація	ні	15	27
Графічний дизайнер	Експерт	Стилізація	так	30	56
Web-дизайнер	Базовий	Реалізм	ні	60	75
Web-дизайнер	Базовий	Реалізм	так	90	112
Web-дизайнер	Базовий	Стилізація	ні	120	160
Web-дизайнер	Базовий	Стилізація	так	150	196
Web-дизайнер	Середній	Реалізм	ні	75	95
Web-дизайнер	Середній	Реалізм	так	100	126
Web-дизайнер	Середній	Стилізація	ні	154	173
Web-дизайнер	Середній	Стилізація	так	165	210
Web-дизайнер	Експерт	Реалізм	ні	90	115
Web-дизайнер	Експерт	Реалізм	так	125	145
Web-дизайнер	Експерт	Стилізація	ні	200	220
Web-дизайнер	Експерт	Стилізація	так	240	270

Для спрощення формування вартості проекту, оберемо основні ознаки, які точно може розуміти замовник, не являючись експертом, та які мають вплив на формування вартості:

- напрям;
- розмір замовлення;
- тип мистецтва;
- наявність анімації (актуальне для графічних дизайнерів та Web-дизайнер).

Окремо слід зазначити наступні відмінності:

а) для Web-дизайнера під стилізацією Реалізм розуміється стандартне рішення веб-дизайну, а Стилізація – створення креативного рішення

б) напрям роботи обумовлює зазвичай тип графіки, тому Растрова або векторна графіка не є ключовим параметром.

Сформуємо для них асоціативні правила діапазону вартості проекту, де мінімальний .

*ЯКЩО Напря́м = = Ретуш ТА Рівень виконавця = = Середній ТА Тип мистецтва = = Реалізм ТОДІ Рекомендована вартість роботи = 6- 70\$*

*ЯКЩО Напря́м = = Ретуш ТА Рівень виконавця = = Середній ТА Тип мистецтва = = Стилізація ТОДІ Рекомендована вартість роботи = 8- 75\$*

*ЯКЩО Напря́м = = Ретуш ТА Рівень виконавця = = Експерт ТА Тип мистецтва = = Реалізм ТОДІ Рекомендована вартість роботи = 15- 300\$*

*ЯКЩО Напря́м = = Цифровий Художник ТА Рівень виконавця = = Базовий ТОДІ Рекомендована вартість роботи = 5- 18\$*

*ЯКЩО Напря́м = = Цифровий Художник ТА Рівень виконавця = = Середній ТОДІ Рекомендована вартість роботи = 9- 25\$*

*ЯКЩО Напря́м = = Цифровий Художник ТА Рівень виконавця = = Експерт ТОДІ Рекомендована вартість роботи = 15- 70\$*

*ЯКЩО Напря́м = = Графічний дизайнер ТА Рівень виконавця = = Базовий ТОДІ Рекомендована вартість роботи = 6- 18\$*

*ЯКЩО Напря́м = = Графічний дизайнер ТА Рівень виконавця = = Середній ТОДІ Рекомендована вартість роботи = 9- 25\$*

*ЯКЩО Напря́м = = Графічний дизайнер ТА Рівень виконавця = = Експерт ТОДІ Рекомендована вартість роботи = 15- 56\$*

*ЯКЩО Напря́м = = Web-дизайнер ТА Рівень виконавця = = Базовий ТОДІ Рекомендована вартість роботи = 60- 196\$*

*ЯКЩО Напря́м = = Web-дизайнер ТА Рівень виконавця = = Середній ТОДІ Рекомендована вартість роботи = 75- 210\$*

*ЯКЩО Напря́м = = Web-дизайнер ТА Рівень виконавця = = Експерт ТОДІ Рекомендована вартість роботи = 90- 270\$*

Також слід зазначити, що вартість для різного розміру проекту визначається за формулою:

$$\text{Вартість} = (\text{Рекомендована вартість проекту} \times \text{Кількість ітенів в проекті}),$$

де *Рекомендована вартість проекту* – діапазон значень від мінімальної вартості проекту до максимальної, яку ми визначили за допомогою експертної оцінки та визначення асоціативних правил;

*Кількість ітенів в проекті* – кількість зображень/відео/макетів, тощо, які необхідно виконати в завданні.

Таблиця 3.5 – Експертні оцінки для різних замовлень із 3D графікою

Напря́м	Стиль	Деталізація	Робота з моделлю	\$	
3D-модельювання об'єктів та сцен	Стилізовані об'єкти	Low-poly	Створення з нуля	25	35
			Редагування існуючої основи	18	27
		High-poly	Створення з нуля	35	45
			Редагування існуючої основи	29	38
	Схематичні або спрощені об'єкти	Low-poly	Створення з нуля	12	17

Продовження таблиці 3.5

			Редагування існуючої основи	5	9	
	Реалізм	High-poly	Створення з нуля	35	57	
			Редагування існуючої основи	20	37	
	Гіперреалізм	High-poly	Створення з нуля	90	160	
			Редагування існуючої основи	52	76	
Розробка 3D-графіки для VR&AR	Стилізовані об'єкти	Low-poly	Створення з нуля	20	30	
			Редагування існуючої основи	12	22	
		High-poly	Створення з нуля	30	40	
			Редагування існуючої основи	22	33	
	Схематичні або спрощені об'єкти	Low-poly	Створення з нуля	7	11	
			Редагування існуючої основи	5	9	
	Реалізм	High-poly	Створення з нуля	35	57	
			Редагування існуючої основи	20	37	
	Гіперреалізм	High-poly	Створення з нуля	90	160	
			Редагування існуючої основи	52	76	
	Візуалізація архітектурних об'єктів	Стилізовані об'єкти	Low-poly	Створення з нуля	28	38
				Редагування існуючої основи	21	30
High-poly			Створення з нуля	38	47	
			Редагування існуючої основи	31	40	
Схематичні або спрощені об'єкти		Low-poly	Створення з нуля	15	20	
			Редагування існуючої основи	9	12	
Реалізм		High-poly	Створення з нуля	39	60	
			Редагування існуючої основи	20	37	

Продовження таблиці 3.5

	Гіперреалізм	High-poly	Створення з нуля	100	170
			Редагування існуючої основи	62	81
Створення текстур	Стилізовані об'єкти	Low-poly	Створення з нуля	17	26
			Редагування існуючої основи	5	10
		High-poly	Створення з нуля	20	29
			Редагування існуючої основи	12	17
	Схематичні або спрощені об'єкти	Low-poly	Створення з нуля	9	13
			Редагування існуючої основи	3	12
	Реалізм	High-poly	Створення з нуля	29	47
			Редагування існуючої основи	17	22
	Гіперреалізм	High-poly	Створення з нуля	50	65
			Редагування існуючої основи	37	45

Через те, що кожний напрям у 3D графіці приблизно однаковий за обсягом роботи та відрізняється лише виглядом готової моделі, вартість послуг дуже схожа, тому різницею можна знехтувати та об'єднати асоціативні правила для визначення діапазону вартості.

Сформуємо для них асоціативні правила діапазону вартості проекту:

*ЯКЩО Стиль = = Стилізовані об'єкти ТОДІ Рекомендована вартість роботи = 12- 47\$*

*ЯКЩО Стиль = = Схематичні або спрощені об'єкти ТОДІ Рекомендована вартість роботи = 3- 20\$*

*ЯКЩО Стиль = = Реалізм ТОДІ Рекомендована вартість роботи = 17- 60\$*

*ЯКЩО Стиль = = Гіперреалізм ТОДІ Рекомендована вартість роботи = 37- 170\$*

Таблиця 3.6 – Експертні оцінки для різних замовлень із Відео

Тип роботи	Тривалість ролика	\$	
Редагування відеопотоку	Короткий	15	22
	Середній	20	37
	Довгий	35	100
Анімація	Короткий	15	22
	Середній	20	37
	Довгий	35	100
VFX ефекти	Короткий	45	100
	Середній	60	130
	Довгий	110	700
Motion design	Короткий	45	100
	Середній	60	130
	Довгий	110	700

Сформуємо для них асоціативні правила діапазону вартості проекту:

*ЯКЩО Тривалість ролика = = Короткий ТА Тип роботи = = Редагування відеопотоку АБО Анімація ТОДІ Рекомендована вартість роботи = 15- 22\$*

*ЯКЩО Тривалість ролика = = Середній ТА Тип роботи = = Редагування відеопотоку АБО Анімація ТОДІ Рекомендована вартість роботи = 20- 37\$*

*ЯКЩО Тривалість ролика = = Довгий ТА Тип роботи = = Редагування відеопотоку АБО Анімація ТОДІ Рекомендована вартість роботи = 35- 100\$*

*ЯКЩО* Тривалість ролика = = Короткий *ТА* Тип роботи = = VFX ефекти  
*АБО* Motion design *ТОДІ* Рекомендована вартість роботи = 45- 100\$

*ЯКЩО* Тривалість ролика = = Короткий *ТА* Тип роботи = = VFX ефекти  
*АБО* Motion design *ТОДІ* Рекомендована вартість роботи = 60- 130\$

*ЯКЩО* Тривалість ролика = = Короткий *ТА* Тип роботи = = VFX ефекти  
*АБО* Motion design *ТОДІ* Рекомендована вартість роботи = 110- 700\$

Переваги методу асоціативних правил:

1) виявлення зв'язків: метод асоціативних правил може виявити кореляції та зв'язки між різними елементами оцінок проектів. Це може допомогти розуміти, як певні фактори взаємодіють між собою;

2) прогнозування на основі історії: Якщо у вас є достатньо даних про експертні оцінки та результати проектів, метод асоціативних правил може допомогти прогнозувати можливі результати нових проектів на основі зв'язків, виявлених в історії.

Недоліки:

1) спрощення реальності: оцінки проектів експертами можуть бути суб'єктивними та визначатися багатьма чинниками. Метод асоціативних правил може спростовувати реальну складність та унікальність кожного проекту;

2) обмеженість врахування контексту: Метод асоціативних правил не завжди дозволяє врахувати контекст або особливості конкретного проекту, що може бути важливим для визначення його вартості;

3) неадекватність для складних проектів: Якщо проекти мають багато складних факторів, їх оцінка може вимагати глибшого аналізу, який метод асоціативних правил може не забезпечити.

Враховуючи ці переваги та недоліки, важливо визначити, чи метод асоціативних правил є найкращим інструментом для визначення вартості проекту на основі експертних оцінок в конкретному випадку.

## 4 ПРОГРАМНА РЕАЛІЗАЦІЯ ПРОГРАМНОГО ФІЛЬТРУ ДЛЯ ВИЗНАЧЕННЯ ВАРТОСТІ ВИКОНАННЯ ПРОЕКТУ ФРІЛАНСЕРОМ

### 4.1 Вибір мови програмування

Для написання сайту для фріланс-біржі можна використовувати різні мови програмування та технології залежно від ваших уподобань, вимог проекту та ресурсів, які ви плануєте використовувати. Ось деякі популярні варіанти:

#### 1. JavaScript (Node.js) + React/Angular/Vue.js:

- **Backend:** Node.js може бути використаний для написання серверної частини. Ви можете використовувати фреймворки, такі як Express або Nest.js.
- **Frontend:** React, Angular або Vue.js можуть бути використані для створення динамічного та інтерактивного інтерфейсу користувача.

#### 2. Python + Django/Flask:

- **Backend:** Python є популярним вибором для серверної розробки. Django та Flask є популярними фреймворками для побудови веб-застосунків на Python.

- **Frontend:** Можна використовувати шаблонізатори, такі як Jinja2, або включити JavaScript-бібліотеки для інтерактивності.

#### 3. Ruby + Ruby on Rails:

- **Backend:** Ruby on Rails є потужним фреймворком для швидкої розробки веб-застосунків на Ruby.

- **Frontend:** Використовуйте ERB або включіть JavaScript-фреймворки для фронтенду.

#### 4. PHP + Laravel/Symfony:

- **Backend:** PHP залишається однією з найпопулярніших мов для веб-розробки. Laravel та Symfony - це потужні фреймворки для PHP.

- **Frontend:** Використовуйте шаблонізатори, такі як Blade (для Laravel), або включіть JavaScript-бібліотеки.

#### 5. Java + Spring Boot:

- **Backend:** Java з Spring Boot може бути використано для розробки надійного та масштабованого серверного застосунку.
- **Frontend:** Використовуйте JavaServer Pages (JSP), Thymeleaf або включіть JavaScript-фреймворки.

Важливо враховувати, що вибір мови та фреймворку також може залежати від навичок, вимог проекту, командної спроможності та інших факторів.

Фільтр необхідно додати на вже існуючий веб-сайт, тому необхідно взяти до уваги технології, які використовуються на цьому сайті. Зазвичай фільтри веб-сайтів можна реалізувати за допомогою JavaScript, але це залежить від того, як саме веб-сайт створено.

Розглянутий сайт фріоанс-біржі використовує статичні HTML-сторінки без великої кількості динамічної взаємодії, тому можна написати фільтр, який працює на боці клієнта (в браузері) за допомогою JavaScript для маніпулювання DOM-елементами та взаємодії з подіями користувача або бібліотеки, такої як jQuery.

JavaScript – це мова програмування, яка використовується для реалізації веб-сценаріїв та взаємодії з користувачем на стороні клієнта (у браузері). Фільтри можна використовувати для обробки та відображення обраних даних на веб-сайті.

Проте, важливо враховувати, що JavaScript працює на стороні клієнта, і дані можуть бути легко модифіковані або обхіднуті, тому для важливих операцій та фільтрів також слід використовувати серверні механізми та перевірки безпеки.

Переваги використання JavaScript для написання фільтрів включають:

1) миттєвий зворотній зв'язок: фільтри на стороні клієнта дають користувачам миттєвий зворотній зв'язок без перезавантаження сторінки;

2) використання ресурсів клієнта: Всі обчислення відбуваються на стороні клієнта, зменшуючи навантаження на сервер.

3) зручність в реалізації: JavaScript легко інтегрується з HTML та CSS, що полегшує реалізацію і підтримку фільтрів.

Важливо враховувати, що обробка даних на стороні клієнта може бути менш безпечною, тому обробка фільтрів також повинна відбуватися на стороні сервера для забезпечення коректності та безпеки даних.

#### 4.2 Програмна розробка фільтру для визначення рекомендацій вартості роботи фрілансера

Написаний на JavaScript фільтр для створення замовлення на сайті, приймає від користувача такі дані: Напря́м, Рі́вень виконавця, Тип мистецтва, Анімація і в залежності від введених значень за правилами визначає рекомендовану вартість проекту. Принцип роботи фільтра представлений на рисунку 4.1.

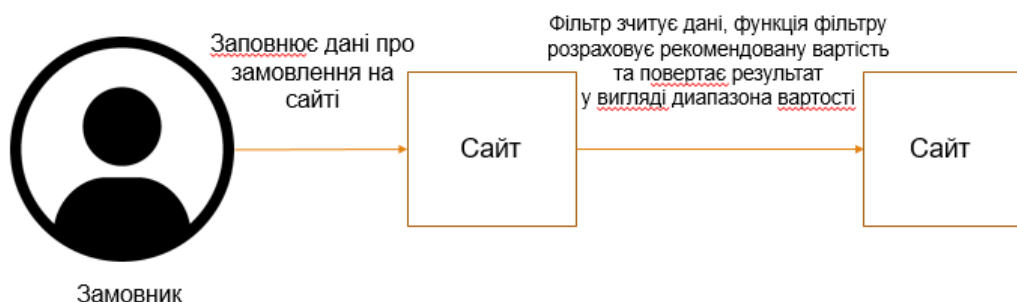


Рисунок 4.1 – Принцип роботи фільтра

Для реалізації знадобиться функція на JavaScript для обробки введених користувачем даних та визначення рекомендованої вартості проекту згідно з вказаними правилами.

Загальний алгоритм для цієї функції (див.рисунок 4.2):

##### 1. Визначення вхідних параметрів:

- напрям - напрям роботи.
- рівень - рівень виконавця.
- типМистецтва - тип мистецтва.

- `анімація` - наявність анімації (true/false).
- `розмірЗамовлення` - розмір замовлення (може бути `null`).
- `кількістьЗавдань` - кількість завдань, яку вводить користувач.

## 2. Визначення правил:

- Створення масиву об'єктів, кожен з яких представляє собою правило з відповідними умовами та вартістю.

## 3. Пошук відповідного правила:

- Використання методу `find` для знаходження правила, яке відповідає введеним користувачем параметрам.

## 4. Врахування анімації:

- Якщо анімація включена і існує відповідне правило, додати до мінімальної та максимальної вартості відповідні суми.

## 5. Розрахунок вартості з урахуванням кількості завдань:

- Помноження мінімальної та максимальної вартості на кількість завдань.

## 6. Повернення результату:

- Повернення діапазону від мінімальної до максимальної вартості як рекомендації.



Рисунок 4.2 – Блок-схема алгоритма функції на JavaScript для обробки введених користувачем даних та визначення рекомендованої вартості проекту згідно з вказаними правилами

Реалізація цієї функції на JavaScript буде виглядати наступним чином:

```

function визначитиВартістьПроекту(напря́м, рівень, типМистецтва, анімація,
розмірЗамовлення, кількістьЗавдань) {
  // Задаємо правила
  const правила = [
    // ... (залишається без змін)
  ];

  // Знаходимо відповідне правило
  const відповідне = правила.find(правило =>
    правило.напря́м === напря́м &&
  
```

```

    правило.рівень === рівень &&
    (!правило.типМистецтва || правило.типМистецтва === типМистецтва)
    &&
    (!правило.розмірЗамовлення || правило.розмірЗамовлення ===
розмірЗамовлення)
    );

    // Перевірка на наявність анімації
    if (анімація && відповідне) {
        відповідне.вартість[0] += 5; // Додати 5 до мінімальної вартості за
наявності анімації
        відповідне.вартість[1] += 20; // Додати 20 до максимальної вартості за
наявності анімації
    }

    if (!відповідне) {
        return null;
    }

    // Розрахунок вартості враховуючи кількість завдань
    const мінімальнаВартість = відповідне.вартість[0] * кількістьЗавдань;
    const максимальнаВартість = відповідне.вартість[1] * кількістьЗавдань;

    return [мінімальнаВартість, максимальнаВартість];
}

// Приклад використання
const напрям = 'Ретуш';
const рівень = 'Середній';
const типМистецтва = 'Реалізм';
const анімація = false;
const розмірЗамовлення = null;
const кількістьЗавдань = 3; // Кількість завдань, яку вводить користувач

const рекомендованаВартість = визначитиВартістьПроекту(напря́м, рівень,
типМистецтва, анімація, розмірЗамовлення, кількістьЗавдань);

if (рекомендованаВартість) {
    const [мінімальна, максимальна] = рекомендованаВартість;
    console.log(`Рекомендована вартість проекту за ${кількістьЗавдань}
завдань: від $$${мінімальна} до $$${максимальна}`);
} else {
    console.log('Рекомендована вартість не знайдена');
}

```

В даному кодї створюється функція визначитиВартістьПроекту, яка приймає введені користувачем дані і повертає рекомендовану вартість проекту відповідно до заданих правил. Також враховується наявність анімації та розміру замовлення.

Додамо в в функцію розрахунку вартості проекту формулу, яка рекомендованаВартість помножує на кількість завдань, які необхідно виконати (цей параметр задається користувачем в окремому полі на сайті) та видає діапазон від мінімальної вартості проекту до максимальної в якості рекомендації у текстовому вигляді на сайті:

```
function визначитиВартістьПроекту(напря́м, рівень, типМистецтва, анімація,
розмірЗамовлення, кількістьЗавдань) {
    // Задаємо правила
    const правила = [
        // ... (залишається без змін)
    ];

    // Знаходимо відповідне правило
    const відповідне = правила.find(правило =>
        правило.напря́м === напря́м &&
        правило.рівень === рівень &&
        (!правило.типМистецтва || правило.типМистецтва === типМистецтва)
        &&
        (!правило.розмірЗамовлення || правило.розмірЗамовлення ===
розмірЗамовлення)
    );

    // Перевірка на наявність анімації
    if (анімація && відповідне) {
        відповідне.вартість[0] += 5; // Додати 5 до мінімальної вартості за
наявності анімації
        відповідне.вартість[1] += 20; // Додати 20 до максимальної вартості за
наявності анімації
    }

    if (!відповідне) {
        return null;
    }
}
```

```

}

// Розрахунок вартості враховуючи кількість завдань
const мінімальнаВартість = відповідне.вартість[0] * кількістьЗавдань;
const максимальнаВартість = відповідне.вартість[1] * кількістьЗавдань;

return [мінімальнаВартість, максимальнаВартість];
}

// Приклад використання
const напрям = 'Ретуш';
const рівень = 'Середній';
const типМистецтва = 'Реалізм';
const анімація = false;
const розмірЗамовлення = null;
const кількістьЗавдань = 3; // Кількість завдань, яку вводить користувач

const рекомендованаВартість = визначитиВартістьПроекту(напря́м, рівень,
типМистецтва, анімація, розмірЗамовлення, кількістьЗавдань);

if (рекомендованаВартість) {
  const [мінімальна, максимальна] = рекомендованаВартість;
  console.log(`Рекомендована вартість проекту за ${кількістьЗавдань}
завдань: від $$${мінімальна} до $$${максимальна}`);
} else {
  console.log('Рекомендована вартість не знайдена');
}

```

Ця версія функції враховує введену користувачем кількість завдань та повертає рекомендовану вартість проекту за цю кількість завдань.

Результат роботи фільтра представлений на рисунок 4.3.

Таким чином, була прописана певна логіка фільтра, що дозволяє формувати рекомендовану вартість виконання проекту для цифрових митців, з урахуванням типу замовлення та його характеристик. Для поліпшення логіки можна збільшити кількість характеристик. Це суттєво підвищить ефективність фільтра.

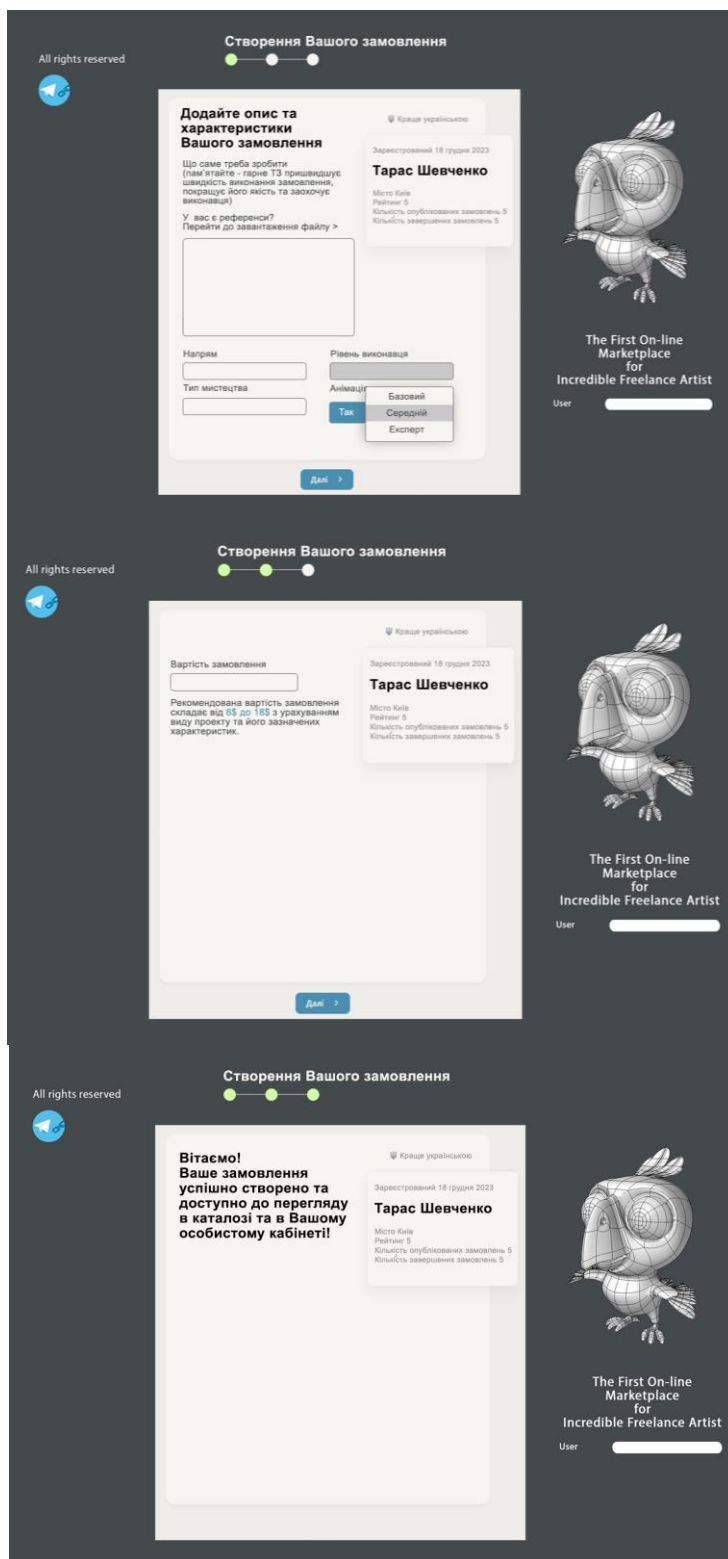


Рисунок 4.3 – Результат роботи фільтра

## ВИСНОВКИ

Згідно завданню необхідно провести дослідження застосування методів інтелектуального аналізу даних у системах фриланс-бірж для 2D/3D Generalist з метою об'єктивного встановлення вартості поточних робіт цифрових митців.

Для вирішення поставленої задачі проведено аналіз предметної області, на основі якого були визначені вигоди та переваги на фриланс-біржі та поза нею, вимоги до сучасної біржі фрилансу та проведено аналіз існуючих фриланс-бірж та виявлені їх недоліки. Так, в результаті аналізу виявлено проблему існуючих фриланс-бірж – на біржі замовник замовляє роботу і, на свій розсуд, сам визначає її вартість, через що виникає дуже великий дисбаланс вартості робіт цифрових митців.

В рамках дослідження методів інтелектуального аналізу даних та застосування їх в системах фриланс-бірж проаналізовані KDD та модель інтелектуального аналізу даних, інтелектуальні методи отримання інформації для аналізу, методи інтелектуального аналізу даних.

В рамках структурного проектування програмного фільтру для визначення вартості виконання проекту фрилансером проведено кластерний аналіз в контексті задачі прогнозування вартості виконання проекту фрилансером та розглянуто метод асоціативних правил для розрахунку виконання проекту фрилансером.

В рамках програмної реалізації на основі вибраної мови програмування та виконана програмна розробка фільтру для визначення рекомендацій вартості роботи фрилансера.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Фриланс біржа «UpWork»: веб-сайт. URL: <https://www.upwork.com/> (дата звернення: 10.10.2023).
2. Фриланс біржа «Кабанчик»: веб-сайт. URL: <https://kabanchik.ua/ua> (дата звернення: 13.10.2023).
3. Фриланс сервіс «Freelancehunt»: веб-сайт. URL: <https://freelancehunt.ua/> (дата звернення: 15.10.2023).
4. Ілляшенко К.В. Роль інтелектуального аналізу даних у сучасному бухгалтерському обліку // ПРОБЛЕМИ СИСТЕМОГО ПІДХОДУ В ЕКОНОМІЦІ. ВИПУСК № 6(74), 2019. URL: [http://www.psae-jrnl.nau.in.ua/journal/6\\_74\\_3\\_2019\\_ukr/15.pdf](http://www.psae-jrnl.nau.in.ua/journal/6_74_3_2019_ukr/15.pdf) (дата звернення: 19.10.2023).
5. Методи інтелектуального аналізу даних. Телекомунікаційний бізнес. URL: [https://stud.com.ua/10028/ekonomika/telekomunikatsiyiny\\_biznes](https://stud.com.ua/10028/ekonomika/telekomunikatsiyiny_biznes) (дата звернення: 21.10.2023).
6. Що таке data mining (аналіз даних). URL: <https://futurenow.com.ua/shho-take-data-mining-analiz-danyh/> (дата звернення: 25.10.2023).
7. Data Mining – добыча данных. URL: <https://basegroup.ru/community/articles/data-mining> (дата звернення: 05.12.2024).
8. Сергеев-Горчинський О. О., Іщенко Г. В. Інтелектуальний аналіз даних
9. Вуйцик В., Смаилова С., Увалиева І. Применение технологии «Data Mining» для анализа данных учебного процесса. Актуальні проблеми економіки. 2014. № 1 (151). С. 487–495.
10. Michael Abernethy. Data mining with WEKA / IBM developerWorks, 2010.
11. Christopher Pal, Mark Hall, Eibe Frank, Ian Witten. Data Mining: Practical Machine Learning Tools and Techniques, 4rd ed. / Morgan Kaufmann, 2016.
12. Jason Bell. Machine Learning: Hands-On for Developers and Technical Professionals / John Wiley & Sons, 2014.
13. Ерік Фрімен, Елізабет Робсон «Head First. Програмування на JavaScript».