

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
Західноукраїнський національний університет  
Факультет комп'ютерних інформаційних технологій  
Кафедра інформаційно-обчислювальних систем і управління

АНІСІМОВА Інна Юріївна

Метод формування навчальної вибірки для сегментації  
організаторів тендерів на основі машинного навчання /  
Method of Forming a Training Sample for Segmentation of  
Tender Organizers on Machine Learning Basis

спеціальність: 122 - Комп'ютерні науки  
освітньо-професійна програма - Комп'ютерні науки

Кваліфікаційна робота

Виконав студент групи  
КНМ-21  
І.Ю. Анісімова

---

Науковий керівник:  
к.т.н., доцент Х.В. Лип'яніна-  
Гончаренко

---

Кваліфікаційну роботу  
допущено до захисту:  
«\_\_» \_\_\_\_\_ 20\_\_ р.  
Завідувач кафедри  
\_\_\_\_\_ М.П. Комар

ТЕРНОПІЛЬ - 2021

## ЗМІСТ

Вступ.....	3
1 Аналіз предметної області та відомих рішень у сфері машинного навчання.....	5
1.1 Суть та особливості організації тендерів.....	5
1.2 Дослідження методів аналізу тендерних закупівель.....	12
1.3 Аналіз існуючих підходів на основі машинного навчання.....	22
1.4 Постановка задачі.....	24
Висновки до розділу 1:.....	26
2 Метод сегментації організаторів тендерів на основі машинного навчання.....	27
2.1 Метод сегментації.....	27
2.2 Метод пониження розмірності.....	35
2.3 Методи кластеризації на основі машинного навчання.....	41
2.4 Метод формування навчальної вибірки для сегментації організаторів тендерів на основі машинного навчання.....	45
Висновки до розділу 2:.....	49
3 Реалізація методу сегментації організаторів тендерів на основі машинного навчання.....	50
3.1 Інформаційно-аналітичне забезпечення публічних закупівель.....	50
3.2 Реалізація сегментації тендерів.....	52
3.3 Реалізація кластерного аналізу.....	55
3.4 Опис отриманих результатів.....	58
Висновки до розділу 3:.....	61
Висновки.....	62
Список використаної літератури:.....	64
ДОДАТОК А Код реалізації.....	72
ДОДАТОК Б Апробація отриманих результатів.....	83

## ВСТУП

Тендери є важливим інструментом у сучасному механізмі ринкової економіки: вони сприяють розвитку внутрішньої торгівлі, що, в свою чергу, є засобом забезпечення більш високих темпів зростання національної економіки. Тендер є індикатором економічної цивілізації країни, оскільки ступінь розвитку механізму державних закупівель може визначати рівень розвитку всієї економічної системи. Tender [20] — конкурентна форма розміщення замовлення на закупівлю товарів, надання послуг чи виконання робіт відповідно до наперед визначених у документації умов в узгоджені терміни на принципах загальності, справедливості й ефективності.

Учаснику тендерів, важливо знати, про організатора ґрунтовну інформацію, скільки організатор проводив аукціонів, скільки з них успішних та на яку суму. Тому для цього потрібна автоматизована система для вибору організаторів тендерів на основі машинного навчання, що також дасть можливість автоматизувати процес участі у тендері. Для цього потрібно сформувати навчальну вибірку, яка зможе навчити систему розпізнавати організаторів тендерів, без додаткових обрахунків.

У зв'язку з цим можна вважати, що розробка методу формування навчальної вибірки для сегментації організаторів тендерів на основі машинного навчання є одним із найбільш перспективних напрямків у електронних тендерах.

**Метою** роботи є розробка методу сегментації організаторів тендерів на основі машинного навчання.

Досягнення цієї мети зумовило потребу теоретичних розробок, визначення та послідовного вирішення таких завдань:

1. визначити суть та особливості організації тендерів;
2. дослідити методи аналізу тендерних закупівель;
3. провести аналіз існуючих підходів на основі машинного навчання;
4. дослідити методи сегментації на основі машинного навчання;
5. дослідити методи кластеризації на основі машинного навчання;

6. розробити метод формування навчальної вибірки для сегментації організаторів тендерів на основі машинного навчання;
7. провести реалізацію методу сегментації організаторів тендерів на основі машинного навчання.

**Об'єктом дослідження** є процес впровадження машинного навчання в сегментацію організаторів тендерів.

**Предметом дослідження** є сукупність теоретико-методологічних та прикладних засад по сегментації організаторів тендерів.

**Методи дослідження.** У роботі використані методи математичного програмування; системного аналізу; машинного навчання.

**Наукова новизна отриманих результатів.** У процесі досягнення мети роботи автором було отримано ряд результатів, які мають наукову новизну. Новими є такі наукові результати: розроблено метод формування навчальної вибірки для сегментації організаторів тендерів на основі машинного навчання.

**Практичне значення** одержаних результатів полягає у формуванні навчальної вибірки для сегментації організаторів тендерів на основі машинного навчання.

**Структура та обсяг роботи.** Дипломна робота складається із вступу, трьох розділів, висновків, списку використаних джерел. Повний обсяг роботи становить 73 ст. комп'ютерного тексту, який включає 17 рисунки та 7 таблиць. Список використаних джерел із 44 найменувань викладено на 5 сторінках.

**Апробація результатів дослідження.** Основні теоретичні положення роботи й практичні результати дослідження доповідалися й обговорювалися на III Всеукраїнської конференції студентів та викладачів закладів освіти «АКТУАЛЬНІ ПРОБЛЕМИ СУЧАСНОЇ ОСВІТИ: РЕАЛІЇ ТА ПЕРСПЕКТИВИ», Маріуполь (04 травень 2021 р.) та XXII Всеукраїнська наукова конференція молодих учених «Актуальні проблеми природничих і гуманітарних наук у дослідженнях молодих учених «Родзинка – 2021»», Черкаси (22 – 23 квітня 2021 року).

# 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ВІДОМИХ РІШЕНЬ У СФЕРІ МАШИННОГО НАВЧАННЯ

## 1.1 Суть та особливості організації тендерів

В умовах ринкової економіки особливу значимість набуває оцінка ефективності організації та проведення тендерів, оголошених підприємствами при пошуку контрагентів на ринку товарів, робіт, послуг, виявлення недоліків та їх причин для прийняття адекватних управлінських заходів. Одним із способів вирішення виниклої проблеми є проведення управлінського аудиту. Ефективність діяльності організації будується на основі оцінки економічного, організаційного та соціального аспектів.

Наш час можна назвати економічно вільним, епохою ринку і конкуренції. Сучасна конкуренція характеризується високою активністю і жорсткістю. В таких умовах компанії змушені бути гнучкими і винахідливими в методах просування своїх товарів і послуг, залучення нових клієнтів і замовлень.

В якості традиційних методів залучення клієнтів і замовлень застосовуються реклама, PR, особисті продажі, організація та проведення зустрічей, презентацій, робота на виставках, управління брендом та інші. Останнім часом фахівці в області маркетингу ведуть дискусії про те, що ефективність традиційних методів знижується, що необхідно приділяти увагу новим методам просування і новим технікам продажів. Крім інших до нових методів у сучасних ринкових умовах можна віднести і участь в тендерах.

Тендер (від англ. Tender - пропозиція) - це конкурентна форма відбору пропозицій на поставку товарів або надання послуг на принципах змагальності, справедливості та ефективності. Контракт укладається з переможцем тендера - учасником, які подали заявку, відповідну вимогам документації, в якій запропоновані найкращі умови [1].

Варто відзначити, що сама структура тендеру була придумана після Другої світової війни в США для того, щоб боротися з корупцією.

Відомо, що в 1946 році на відновлення німецьких міст були виділені величезні суми, проте левова частка цих грошей осіла в кишенях американських чиновників. Щоб уникнути подібних ситуацій надалі, американці постаралися зробити процес вибору постачальника якомога більш прозорим і об'єктивним, розписавши в деталях покрокові дії цієї процедури [2].

Термін «тендер» легко увійшов в повсякденну мову і часто використовується як аналог термінів конкурс, аукціон. Але в законодавстві РФ відсутнє поняття «тендер». У законі № 94-ФЗ замість нього використовується термін «відкритий конкурс». Також закон визначає такі типи закупівель як запит котирувань, відкритий аукціон, відкритий аукціон в електронній формі, попередній відбір, біржові торги [3].

Тендери розрізняються по складу учасників і за допустимим змінам конкурсних вимог. Порівняльна характеристика сучасних типів закупівель представлена в таблиці 1.1.

Таблиця 1.1. -Порівняльна характеристика сучасних типів закупівель

Характеристики закупівлі	Тип закупки		
	Відкритий тендер	Закритий тендер	Запит котирувань
Склад виконавців	Будь-які компанії в будь-якій кількості	Вибрані організаторами компанії	Будь-які компанії, не менше 3-х
Тип товару (послуги)	Будь-який товар або послуга	Технічно складні товари; товари для задоволення потреб безпеки	Серійна продукція
Обсяги закупівель	Як правило, 100 тис грн і вище	Як правило, 100 тис грн і вище	Існує ліміт, вище якого закупівлі проводяться на відкритому тендері
Інформаційне	На електронних ресурсах та ЗМІ	Не публікується	На електронних ресурсах та ЗМІ
Розміщення	Висока конкуренція	Заощаджують час, менш витратні щодо відкритих; заявки можуть подати тільки запрошені компанії	Висока оперативність; вважається що відбувся навіть якщо після запиту встановлено, що лише один постачальник здатний виконати замовлення

Всі тендери можуть проводитися як в один, так і в два етапи. Двохетапні тендери проводяться за відсутності чітко сформульованих вимог до предмета тендеру, або при необхідності провести переговори з можливими виконавцями для оцінки їх можливостей.

Специфікою проведення тендеру в два етапи є те, що на першому етапі учасники подають свої заявки без зазначення ціни. І тільки на другому етапі, після уточнення вимог до предмету тендеру і внесення змін в документацію, постачальник зазначає ціну. Як правило, до двохетапним тендерами вдаються, коли потрібно постачання складної або унікальної продукції або послуги [4].

На отримання тендера в Україні можуть претендувати як вітчизняні, так і зарубіжні претенденти. Але, як правило, перевага отримує все таки вітчизняний претендент.

Якщо ж говорити про галузі, в яких використовуються тендерні закупівлі, то майже 20% всіх тендерів проводяться на замовлення підприємств з галузі медицини та фармакології, 17% тендерів проводить будівельна галузь і 14% - харчова промисловість. Частка інших галузей в загальній структурі закупівель 5% і менше [5]. Структура тендерних закупівель в розрізі галузей представлена на рис. 1.1.

Інформація про проведені тендери розміщується на офіційному сайті держзакупівель ([tender.me.gov.ua](http://tender.me.gov.ua)), регіональному сайті, сайтах компаній-замовників, а також на спеціальних ресурсах. Там же відображаються і підсумки відкритих конкурсів.

Останнім часом широкого поширення набувають відкриті аукціони в електронній формі, що проводяться на сайті електронні торгові майданчики (ЕТМ), комерційних майданчиках ЯТ5-тендер і В25-Сетег.

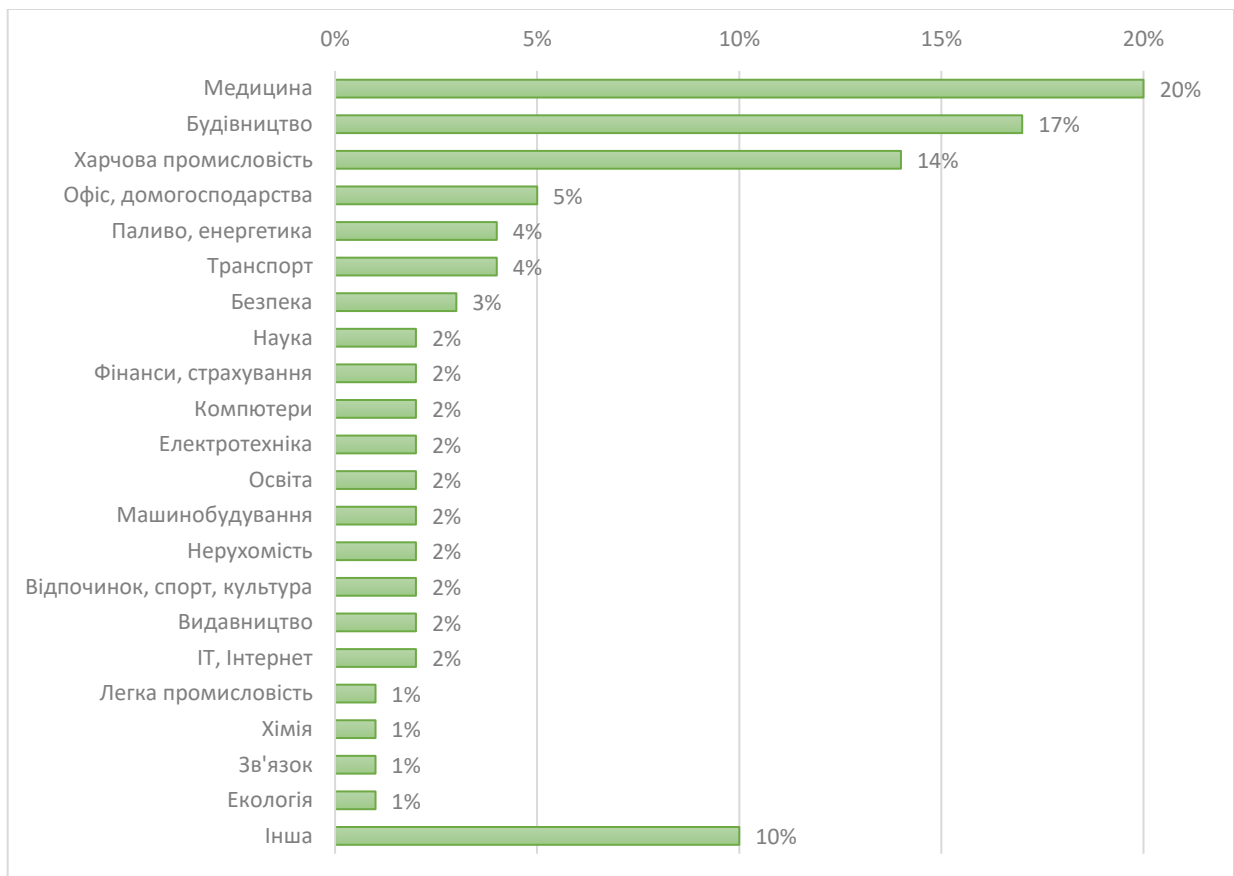


Рис. 1.1 - Структура тендерних закупівель в розрізі галузей

Перевагами відкритого аукціону на ЕТМ є:

- Можливість проведення «швидких» торгів у режимі он-лайн з автоматичним формуванням протягом декількох хвилин протоколу з підсумками торгів;
- Для замовника: можливість оперативного пошуку постачальника, який надає кращу ціну;
- Для постачальника: можливість зробити кілька цінових пропозицій за тендером і змінити свою початкову пропозицію.

Постачальники, які бажають взяти участь у тендері, зобов'язані в зазначені терміни подати заявку. Як правило, вимоги до заявок і документів зрозумілі і доступні, часто на сайтах розміщуються готові форми для заповнення. Переможцем визнається учасник, що запропонував найнижчу ціну контракту. Також виявляється учасник, який зробив кращу пропозицію за ціною, наступне за пропозицією переможця [3].



У тендера, як і у будь-якого методу отримання замовлень, є свої особливості, переваги і недоліки.

В якості сильних сторін для постачальника послуг можна виділити:

1. Збільшення потоку фінансових коштів. Причому даний аспект можна розглядати як збільшення прибутку, так і збільшення оборотних коштів. Згідно зі статистикою економічний ефект від участі в тендері складає 25-30% і виражається в отриманій постачальником прибутку [4].

2. Розширення ринків збуту. Тендер дає можливість розширити, як галузеві ринки збуту, так і географічні.

3. Збільшення обсягу продажів послуг і, як наслідок, зростання закупівель матеріалів, техніки у виробника (постачальника), що позначається на розмірі знижки, бонусів та умови взаємодії. Збільшення зацікавленості співробітників при відрядній оплаті праці.

4. Економія на просуванні, пошуку нових клієнтів і замовлень. Для просування послуг традиційними методами, як правило, необхідні суттєві фінансові та часові витрати. У той час, як для участі в тендері потрібні мінімальні грошові вкладення, участь не забирає багато часу і не вимагає спеціальної підготовки [1];

5. Можливість виявити сильних конкурентів, придбати корисні зв'язки.

На практиці у тендерів, крім сильних сторін, є і слабкі. По-перше, недоліком тендеру є «гонка на пониження ціни». Механізм проведення тендерів та аукціонів такий, що перемагає постачальник, що пропонує найбільш низьку ціну. Ця умова впливає і на баланс «ціна - якість»: знижується ціна і, як правило, знижується якість (використовується низькоякісна сировина, наймається некваліфікований персонал, зменшується контроль тощо). В результаті страждає і замовник, і кінцевий споживач, і постачальник. Негативні наслідки для останнього проявляються у втраті репутації, погіршення партнерських відносин; можливі збитки і судові розгляди.

Компанії, які працюють на принципах високої якості надання послуги і дорожать репутацією, часто змушені відмовитися від участі в тендерах, або

працювати з дуже низькою або навіть нульовим прибутком з метою отримання іншої не фінансової вигоди.

Другим недоліком є жорсткі вимоги до правильності заповнення документації. У випадку, якщо в заявці або прикладеної документації допущена хоч якась помилка, учасник може бути знятий без повідомлення.

Для участі в тендері необхідно грамотно оформляти документи, усувати неточності вимог тендерної документації з боку замовника шляхом направлення запитів на роз'яснення положень документації, враховувати останні зміни в законодавстві. При цьому за даними проведеного дослідження журналом «Радник», 44% респондентів в якості основної проблеми проведених тендерів виділяють неоперативність організаторів у роз'ясненні виникаючих питань (6).

І, по-третє, не завжди боротьба за тендер проходить чесно. У мережі Інтернет давно вже з'явилося поняття «нечесний тендер». Тобто тендер проводиться формально для дотримання регламентів і процедур. Також існує практика підкупів і фальсифікації результатів тендерів, порушення та обмеження прав учасників [6].

Останнім часом відзначається збільшення кількості скарг на порушення в проведенні тендерів в контролюючі органи. Це є хорошим показником, оскільки вказує на зростання юридичної грамотності учасників, а також на те, що постачальники готові не тільки брати участь і перемагати, але й відстоювати свої права.

Тим не менш, незважаючи на всі недоліки тендерної системи, можна говорити про те, що для постачальника, зацікавленого у зменшенні кадрових, матеріальних і тимчасових витрат, тендер - ефективний метод отримання замовлень.

Організація тендерів сприяє виявленню на ринку сумлінних організацій-партнерів, які виконують роботу відповідально і якісно. а також укладення договорів за реальними (незавишеними) цінами. Для виконавця укладення договору за підсумками тендерів є гарантією забезпечення довгострокових

замовлень, стабільності у підприємницькій діяльності, позитивного іміджу на ринку.

Характеристика існуючих на сьогоднішній день різновидів тендерів представлена в таблиці 1.2.

Таблиця 1.2. - Класифікація тендерів

Найменування кваліфікаційної ознаки	Вид конкурсу	Коротка характеристика
1. Склад учасників тендеру	1.1. Відкриті. 1.2. Закриті	1.1. У конкурсі можуть брати участь будь-які організації. Про проведення відкритих тендерів публікуються оголошення в періодичній пресі. Велике число учасників загострює конкуренцію, що дозволяє розмістити замовлення на більш вигідних умовах. 1.2. Тендери оголошуються тільки для спеціально запрошених осіб. Відібраних організаторами торгів. Зазвичай в закритих тендерах беруть участь найбільші фірми або їх консорціуми
2. Допустимість конкурсних вимог	2.1. Одноповерхові. 2.2. Двоетапний	2.1. Особа бажає брати участь у тендері, подає одночасно документи для визначення учасників тендеру та переможця тендера. В даному виді тендера переговори заборонені. 2.2. На першому етапі організатор розробляє первісний (приблизний) варіант технічного завдання для підготовки постачальниками заявок без зазначення ціни та інших комерційних умов. За пропонованими заявками проводять переговори для визначення різних варіантів рішення задачі і встановлення остаточного варіанту технічного завдання. На другому етапі учасники подають остаточні конкурсні заявки з технічним і комерційною пропозицією, підготовленим згідно з уточненим варіанту конкурсної документації
3. Кваліфікаційний відбір	3.1.3 попереднім кваліфікаційним відбором. 3.2. Без попереднього кваліфікаційного відбору	3.1. Експертним шляхом визначається ступінь відповідності учасників конкурсу пропонованим вимогам. Даний відбір проводиться у вигляді окремої процедури до подачі заявок з техніко-комерційними пропозиціями [7]. 3.2. Попередня кваліфікація не виробляється

Представлена класифікація дозволяє сформулювати уявлення про те, що тендерам як особливому способу укладання договорів початку властива вільна конкуренція. Саме наявністю елемента змагальності укладення договору за допомогою оголошення тендерів відрізняється від звичайного способу виникнення договірних зобов'язань шляхом акцепту оферти.

Однак на практиці часто виникають ситуації, коли організатору надходить тільки одна конкурсна заявка або коли формально в тендерах беруть участь кілька осіб, але вони не змагаються між собою (наприклад, подають свідомо неприйнятні для організатора тендерів заявки). Такі ситуації можна об'єднати поняттям «неконкурентні тендери» [8].

## 1.2 Дослідження методів аналізу тендерних закупівель

Успіх соціально-економічної політики в Україні залежить від багатьох факторів, одним з яких є сфера державних закупівель (закупівля товарів, робіт і послуг за державні кошти), ефективність якої визначається дотриманням законодавства та розумною диверсифікаційною політикою. Тому пріоритетним завданням для органів управління державними фінансами є забезпечення контролю за витрачанням державних коштів як в частині забезпечення виконання функцій держави, так і в частині виконання зобов'язань перед громадськістю.

Оцінка стану державних закупівель в Україні показала, що за період 2007-2013 років їх обсяг збільшився з 142 млрд 285 млн грн до 428 млрд 063 млн грн, що майже на 200% [9]. Обсяг виділених коштів свідчить про те, що такі закупівлі є потужним інструментом у системі державного регулювання національної економіки, а їх зростання збільшує рівноважний обсяг ВВП, що, в свою чергу, стимулює соціально-економічне зростання в країні [10]. Однак наукові оцінки сучасного стану нормативно-правової бази системи державних закупівель в Україні дозволяють зробити висновок, що в правовому полі такі закупівлі

залишається значна частка невизначеності як на макро- та мікрорівні. Це стосується, в першу чергу, процедур внутрішнього контролю та аудиту державних закупівель, визначення місця та ролі громадського контролю, вимог до змісту звернення (оскарження) закупівель. Низка питань, починаючи від законодавства до питань, пов'язаних з відсутністю висококваліфікованих кадрів, які здатні працювати в тендерних комітетах і надавати технічну підтримку за всіма договорами, перешкоджають розвитку державних закупівель товарів, робіт і послуг.

Водночас спостерігається відсутність контролю за державними закупівлями, що обґрунтовується тенденцією до збільшення порушень та зловживань у цій сфері. Так, у 2013 році частка порушень у сфері публічних закупівель до загальної кількості порушень бюджету становила 24,9%, що на 5,7% більше, ніж у 2012 році [11]. Слід зазначити, що з року в рік підвищується ефективність контрольних заходів, але ситуація з усуненням порушень і реагуванням на їх виявлення погіршується. Цей факт свідчить про те, що численні зміни до законодавства про закупівлі не вирішують проблемних питань, які виникають під час тендеру, а кількість контрольних заходів не зменшує кількість порушень.

Часто учасники торгів не обмежуються завданням організації та проведення торгів, вони змушені поєднувати свої функціональні обов'язки з іншими робочими навантаженнями, що є наслідком низького інтересу до економічної ефективності та ефективності торгів. Відсутність чіткого уявлення про характер і види фінансових порушень, ознаки різних видів фінансових порушень і санкції, які відповідають кожному виду порушення, призводять до прецедентів оскарження результатів аукціону. За даними Антимонопольного комітету України, який є органом оскарження державних закупівель, його постійно діюча адміністративна колегія з розгляду скарг на порушення законодавства про публічні закупівлі протягом 2012-2013 років прийняла 2517 рішень за скаргами, поданими суб'єктами оскарження (табл. 1.3) [12, 11].

Ефективність даних свідчить про тенденцію до збільшення кількості скарг, поданих до органу апеляційної скарги, які часто є результатом різних тлумачень замовниками та учасниками процедур закупівель Закону України «Про публічні закупівлі». Це свідчить про необхідність посилення ролі фінансового контролю за державними закупівлями та викладенні чітких методологічних засад проведення контрольної та аналітичної роботи.

Виходячи з цього, визначаються фактори, що впливають на ефективність державних закупівель, незалежно від особливостей замовників або сфер розміщення замовлень (рис. 1.2).



Рис. 1.2. Чинники, що впливають на ефективність державних закупівель

Складовими будь-якої системи є елементи, їх особливості, взаємозв'язки з іншими компонентами. Розмежування елементів системи контролю державних закупівель ґрунтується на симбіозі складових фінансового контролю як функції управління та безпосередньо державних закупівель як об'єкта контролю, що передбачає конвергенцію та формування відносин між цими складовими (табл. 1.3).

Ключовою групою елементів системи фінансового контролю та контролю закупівель є її теоретична основа. Адже без належної теоретичної основи не

будуть функціонувати інші елементи цієї системи, не буде єдиної, інтегрованої системи фінансового контролю.

Таблиця 1.3 - Інформація органу оскарження щодо кількості скарг та прийнятих рішень протягом 2019 - 2020 рр.

Інформація щодо скарг	Рік		Відхилення	
	2019	2020	абсолютне	відносне
Усього надійшло скарг, шт.	665	787	122	18,4
Прийнято рішень, у тому числі, шт.:	1077	1440	363	33,7
- прийняття до розгляду	557	712	155	27,8
- відмовлено у задоволенні скарги	159	171	12	7,5
- задоволено скарг (повністю та частково)	234	338	104	44,4
Сума коштів, що була сплачена до Державного бюджету України за подання скарг, тис. грн	3310	4080	770	23,3

Таблиця 1.4 - Елементи системи контролю закупівель за державні кошти

ЕЛЕМЕНТИ					
Теоретична база	Нормативна база і методичне забезпечення	Організаційні структури	Науково-дослідна і навчальна бази	Матеріально-технічне і фінансове забезпечення	Інформаційно-комунікативна інфраструктура
Визначення поняття, змісту і сутності, цілей і задач, принципів, суб'єкта і об'єкта, форм і методів контролю державних закупівель. Побудова класифікації видів фінансового контролю, методологію організації і проведення фінансового контролю, визначити критерії ефективності фінансового контролю	Система нормативно-правових актів, стандарти фінансового контролю, контрольні процедури і тести, блок-схеми, анкети, періодичні видання, методологічні посібники	Суб'єкти, котрі здійснюють фінансовий контроль: органи, ЦОВВ, департаменти, підрозділи та відповідальні особи, їх статус, взаємодія, розподіл обов'язків	Науково-дослідні та навчальні заклади (центри, частини, ВНЗ тощо), програми, курси	Фінансові ресурси, техніка, оргтехніка, технології, приміщення тощо	Інформаційне забезпечення та системи комунікації

На основі таблиці. 1.4 Встановлено, що основним елементом теоретичної основи системи контролю закупівель є визначення її концепції. Ю. О. Голуб визначає контроль державних закупівель як систематичне спостереження за об'єктом контролю та процедурою державних закупівель [13]. Однак, на нашу думку, систематичний моніторинг не повною мірою розкриває зміст і функції контролю за державними закупівлями, оскільки контроль повинен бути на всіх етапах процесу закупівель, а оскільки одним з таких етапів є виконання договору, необхідно контролювати, чи виконуються умови договору, чи ефективно використовуються придбані товари або отримані відповідні послуги, виконаних робіт. Адже процесуально-документальні закупівлі можуть проводитися правильно, а закуплені предмети не використовуються за призначенням, або взагалі не використовуються, як це часто буває, наприклад, з придбаним медичним обладнанням, коли немає фахівців, які могли б експлуатувати придбане обладнання або неправильно підготовлені умови використання обладнання.

Наступною умовою досягнення мети контролю за державними закупівлями, побудови його інтегрованої та ефективної системи є дотримання принципів, що охоплюють теорію контролю в цілому та сферу закупівель. Основними серед них є принципи:

- законність - суб'єкти контролю повинні неухильно дотримуватися норм і правил, встановлених законом;
- публічність - відкритість діяльності органів фінансового контролю, наявність доступу до інформації про результати контрольної діяльності та зміст виявлених порушень;
- незалежність - суб'єкти контролю повинні мати організаційну, матеріальну незалежність від суб'єктів, які купують товари, роботи або послуги;
- об'єктивність - діяльність суб'єктів контролю повинна виключати упередженість і схильність впливати на прийняття рішень;



- обов'язки - суб'єкти контролю повинні нести економічну, дисциплінарну та адміністративну відповідальність за неналежне виконання покладених на нього контрольних функцій і покладених на нього завдань;
- систематичний - передбачає існування системи суб'єктів фінансового контролю за державними закупівлями для розподілу їх функціональних обов'язків.

Ступінь доступності та дотримання цих домінант може бути використана як орієнтир якості системи прийняття рішень та перевірки наявності перешкод для державних закупівель.

Система фінансового контролю державних закупівель неможлива без таких взаємопов'язаних елементів, як суб'єкт і об'єкт контролю. Суб'єктами фінансового контролю, як правило, вважаються органи, що здійснюють цей контроль: спеціальні органи державного контролю (відділи, де розташовані парламенти, комітети), підрозділи та служби підприємства (бухгалтерський облік, контрольні послуги, внутрішній аудит, планово-економічний відділ тощо), громадські організації. Система суб'єктів фінансового контролю визнає їх підзвітність і контроль, а також механізм взаємодії між цими органами (табл. 1.5). Склад об'єкта контролю державних закупівель характеризується різноманітністю в поглядах вчених. Так, Шевчук В.В. В.[14] зазначає, що об'єктом контролю є «діяльність конкретного суб'єкта, на який спрямовані контрольні дії». Виговська Н.Г. [15] до складових об'єкта контролю відносяться процеси господарської діяльності, майно підприємства та джерела його формування. На нашу думку, об'єктами фінансового контролю закупівель за державний кошт є контрольовані суб'єкти, які беруть участь у процесі закупівель (є замовниками товарів, робіт або послуг), організують та проводять процедури закупівель, а також інформаційно-аналітичне забезпечення закупівель за державні кошти (планові документи; господарські договори; тендерна документація).

Не менш важливою особливістю контролю за державними закупівлями є форма його здійснення, яку слід розуміти як спосіб специфічного вираження та

організації контрольних дій, спрямованих на виконання контрольних функцій, тобто методів контролю та конфігурації контрольних заходів [16].

Таблиця 1.5 - Форми та прийоми контролю закупівель за державні кошти

Форми та прийоми	Застосування підчас контролю
1. Моніторинг	Забезпечення додержання замовниками законодавства у сфері державних закупівель та ефективного і раціонального використання державних коштів, максимальної їх економії при здійсненні державних закупівель
2. Інспектування	Подальший контроль за дотриманням законодавства при здійсненні замовниками державних закупівель, здійснюється у формі ревізій і перевірок
3. Перевірка	спосіб документального виявлення незаконного та нецільового використання коштів, у т. ч. бюджетних, матеріальних цінностей, упередження фінансових зловживань, інших порушень законодавства у сфері державних закупівель
3.1. Арифметична перевірка документів	Контроль проведених у документах обчислень, правильність алгоритму визначення переможця торгів
3.2. Зустрічна перевірка	Підтвердження у суб'єктів господарювання, які мали правові відносини з підконтрольною установою, виду, обсягу та якості операцій і розрахунків, що здійснювалися між ними, для з'ясування їх реальності та повноти відображення в обліку підконтрольної установи
4. Ревізія	Повне, комплексне, системне та об'єктивне обстеженням фінансово-господарської діяльності підконтрольного об'єкта, вивчення первинних фінансових документів щодо правильності відображення в них усіх фінансово-економічних операцій з метою встановлення відповідності прийнятих управлінських рішень та здійснених фінансових операцій вимогам законодавства, у т. ч. вимогам державних закупівель
5. Анкетування	Отримання необхідної для проведення контролю закупівель інформації від посадових осіб суб'єкта господарювання або осіб, які є користувачами товарів, робіт і послуг суб'єкта господарювання, шляхом застосування анкет
6. Інвентаризація	Підтвердження наявних обсягів, кількості придбаних товарів зазначених у звітах
7. Контрольний обмір робіт	Встановлення фактичних обсягів і вартості виконаних будівельних, ремонтних робіт, перевірка їх правильності відображення в актах приймання робіт, правильність списання матеріалів на проведення цих робіт
8. Нагляд	Пов'язаний із спостереженням за дотриманням підконтрольним об'єктом встановлених чинним законодавством закупівельних правил і нормативів
9. Фінансовий аудит	Перевірка та аналіз фактичного стану справ щодо законного та ефективного використання коштів виділених на закупівлю з метою надання висновку з відповідними рекомендаціями

Зазначено в таблиці. 1.5 Кількість форм і методів контролю за державними закупівлями обумовлена тим, що верифікація державних закупівель може

здійснюватися в процесі інспекції, державного фінансового аудиту, моніторингу та окремо. Моніторинг державних закупівель може аналізувати ефективність такої закупівлі та визначати вплив порушень (за наявності) на об'єктивність, правомірність визначення переможця процедури закупівлі та може встановити обґрунтовані підстави вважати, що закупівля була здійснена з такими порушеннями та/або відповідно до умов пропозиції, визнаний замовником найбільш економічно вигідним, завдав або може завдати шкоди державі [13]. Однак Порядок не містить чітких критеріїв віднесення порушень до тих, які вплинули або не вплинули на результати закупівлі. Конкретизація в цьому питанні дозволила б забезпечити однаковість в роботі контролюючих органів, усунути можливі розбіжності з організаціями, що проходять аудит.

За допомогою аналітичних даних Державної фінансової інспекції України за останні п'ять років виявлено негативні тенденції контролю за державними закупівлями товарів, робіт і послуг головними розпорядниками бюджетних коштів (керівниками міністерств та інших центральних органів виконавчої влади, державних адміністрацій), що дозволяє виявити внутрішні чинники, що відображають наявність кризової ситуації в системі державних закупівель. До них відносяться:

- ненадання оптимальної кількості внутрішніх аудиторів та/або необґрунтованих скорочень;
- відсутність ефективної системи внутрішнього контролю для запобігання та запобігання порушенням законодавства про державні закупівлі;
- виїзні підрозділи внутрішнього контролю не проводять і безпосередньо не інформують вище керівництво установи про результати кожної перевірки;
- відсутній внутрішній контроль за повнотою надходжень, отриманих розпорядчими бюджетними керівниками та одержувачами бюджетних коштів нижчого рівня, та їх витрачанням бюджетних коштів.

Відсутність належного внутрішнього контролю за державними закупівлями дозволила виявити типові порушення, які мали місце в різних органах державної влади протягом 2020 року (табл. 1.6). Таким чином, суттєві

порушення та недоліки, виявлені у державних закупівлях, певною мірою є результатом відсутності контролю вищого керівництва, що призводить до необґрунтованої закупівлі товарів і робіт, заниження або завищення їх вартості, надлишків та неефективного використання або невикористання нового обладнання.

Таблиця 1.6 - Типові порушення у сфері державних закупівель протягом 2020 р. [6]

Структура	Сума, млн грн
<b>Міністерство охорони здоров'я України</b>	
Неврахування реальної потреби у медичних препаратах (закуплено більше потреби та списано)	2,6
Не забезпечено належного управління коштами субвенцій з державного бюджету місцевим бюджетам зазначених регіонів на закупівлю витратних матеріалів і медичного обладнання для комунальних закладів охорони здоров'я	35,8
<b>Вищий адміністративний суд України</b>	
Придбане комп'ютерне устаткування не використовується	1,5
<b>Державна податкова служба</b>	
Придбане за завищеними цінами апаратне забезпечення	3,5
<b>Міністерство праці та соціальної політики України</b>	
Придбане устаткування для обробки даних не використовується	39,7
<b>Міністерство внутрішніх справ України</b>	
Недосконала системи управління матеріально-технічним забезпеченням та необґрунтовані управлінські рішення	6,9
Оплата послуг за завищеною вартістю та придбання товарно-матеріальних цінностей, які надалі не використовувалися у миротворчій діяльності	7

Це свідчить про послаблення функції внутрішнього контролю в умовах певної інституційної підтримки. Тому важливо посилити внутрішній контроль над підприємствами, установами чи організаціями, що належать до сфери державного управління, а, отже, посилити щоденний контроль за станом справ на місцях. Адже недоліки в організації контролю значно знижують якість всієї інституційної системи та системи державних закупівель. Більшість порушень у сфері публічних закупівель відбуваються на етапах підготовки тендерної документації та процедур закупівель, коли свідомо встановлюються дискримінаційні умови. Тому для ефективності внутрішнього контролю необхідно встановити інформаційні зв'язки між суб'єктами внутрішнього і

зовнішнього контролю і контрагентами. Виходячи з цього, інформаційним аспектом контролю закупівель є система обробки інформації, що складається з наступних частин: підсистема збору та сортування інформації, підсистема зберігання інформації, підсистема аналізу інформації (рис. 1.3).



Рис. 1.3. Структура процесу обробки інформації державних закупівель

Система обробки інформації є основою для створення інформаційного механізму закупівель та дозволяє оперативно реагувати на зміни в нормативно-правовому та фіскальному середовищі, прогнозувати та планувати необхідний обсяг закупівель, створювати реєстр учасників торгів тощо.

Таким чином, моніторинг та оцінка сучасного стану нормативно-правової бази системи державних закупівель в Україні дозволяє зробити висновок, що правове поле залишається значною часткою невизначеностей, які знижують ефективність процесу закупівель, спричиняють необґрунтоване витрачання бюджетних коштів, знижують якість послуг. Ключовими факторами, які негативно впливають на розвиток системи державних закупівель, є недостатній

рівень контролю та аналітичного забезпечення організації конкурсних торгів, невизначеність та нестабільність нормативно-правової бази у цій сфері. Основними факторами, що впливають на ефективність державних закупівель, є: інформаційне забезпечення, прозорість процедури, розвиток конкурентного середовища, законодавче регулювання, система контролю, компетентність замовника та його відповідальність.

Встановлено, що контроль повинен здійснюватися на всіх етапах процесу закупівлі. Контроль закупівель за державний кошт повинен також покривати ефективність використання закуплених об'єктів, їх цільове використання та зберігання. Суттєві порушення та відхилення у державних закупівлях свідчать про послаблення функції внутрішнього контролю в умовах певної інституційної підтримки. Встановлені інформаційні зв'язки, які пов'язують три напрямки процесу закупівель (самовизначення потреб, які необхідні в першу чергу для планування; прямих закупівель та виконання контрактів), відповідатимуть договірним, правовим та нормативним вимогам закупівельного середовища.

### 1.3 Аналіз існуючих підходів на основі машинного навчання

Багато авторів проводять дослідження державних тендерних закупок [26] в сферах: охорони здоров'я [21]; відновлюваних джерел енергії [22]; харчовій промисловості [23]; будівельній галузі [24]; гірничої промисловості [27] та по окремих країнах [25, 28-30].

У статті [2] розглядаються методи формування класифікованої навчальної вибірки, що генерується лише за допомогою активних перешкод, для адаптації вагових коефіцієнтів просторових фільтрів за умов поєднаної присутності перешкод.

У статті [3] запропоновано адаптивний метод формування класифікованої навчальної вибірки на основі використання порогової оцінки коефіцієнта кореляції міжканалів комбінованих перешкод.

У статті [4] розглянуто проблему ефективного формування репрезентативної вибірки для навчання нейронної мережі багат шарового перцептрона Типу (MLP).

У статті [5] запропоновано алгоритм розробки навчального набору, який найкраще описує об'єкти розпізнавання.

У дослідженні [6] оцінюється ефективність різних підходів кластеризації даних для пошуку вигідних сегментів споживачів у галузі гостинності у Великобританії.

Дослідження [7] засноване на моделі RFM (Recency, Frequency and Monetary) та використовує принципи сегментації набору даних за допомогою алгоритму K-Means. Отримані таким чином результати щодо операцій з продажу порівнюються з різними параметрами, такими як нещодавній продаж, частота продажів та обсяг продажів.

Дослідження [8] має на меті розробити новий методологічний підхід, інтегруючи "Recency, Frequency and Monetary" із рідкісним  $K$ -означає алгоритм кластеризації Вітгена та Тібшірані. Запропонований підхід придатний для обробки великих, великих розмірів та розріджених споживчих даних.

У статті [9] наведено приклад використання методів науки про дані для класифікації покупців інтернет-магазину за їх купівельною діяльністю.

У статті [10] поєднує радіаційні значення взаємовідносин спільноти з моделлю RFM та вдосконалює алгоритм індексу  $M$  для формування моделі RFMC, роблячи її більш придатною для підприємств електронної комерції з характером просування громади.

У роботі [11] досліджено на вирішенні реальних проблем на підприємстві, моделями RFM та алгоритм кластеризації  $K$ -засобів використовуються для проведення сегментації споживачів та аналізу вартості за допомогою даних онлайн-продажів. На цій основі висуваються різні стратегії CRM для досягнення високого рівня задоволеності клієнтів.

У статті [12] за допомогою аналізу індексу оцінки алгоритму кластеризації та експерименту аналізу візуалізації результати показують, що модель та

алгоритм класифікації споживачів ефективніше і повно відображають цінність споживача.

У роботі [13] розглядаються особливості поведінки клієнтів при кластеризації клієнтів, а також метод пошуку оптимальної кількості кластерів та початкових значень центрів кластерів для отримання більш точних результатів.

Дослідження [14] вдосконалює дослідження щодо розробки стратегій цифрового маркетингу на основі рекомендацій, надаючи компаніям прогнозу модель. Ця модель використовує науку даних, включаючи методи машинного навчання та великі дані, для персоналізації фінансових стимулів для користувачів на основі якості нових клієнтів, яких вони посилають на веб-сайт з поверненням готівки.

Дослідження сегментації клієнтів, є надзвичайно популярною темою, що і підтверджено аналізом вище. Проте жоден з них не розглядає сегментацію організаторів тендерних аукціонів, адже це дасть учасникам аукціонів проаналізувати: нові ринки збуту своєї продукції; прозорість та чесність системи електронних торгів – виграє кращий; вибір найбільш привабливого організатора, що дасть можливість слідкувати за найвигіднішими пропозиціями. Також на основі отриманих даних, можна розробити систему для вибору організаторів тендерів на основі машинного навчання, що дасть можливість автоматизувати процес участі у тендері.

На відміну від аналогів [7,8] розроблений метод формування навчальної вибірки для сегментації організаторів тендерів на основі машинного навчання, дозволить сформуванню навчальну вибірку, на основі якої можна буде розробити систему для вибору організаторів тендерів на основі машинного навчання.

#### 1.4 Постановка задачі

Учаснику важливо детально дізнатися про організатора, скільки аукціонів провів організатор, скільки з них були успішними і на яку суму. Тому для цього



потрібна автоматизована система відбору організаторів тендерів на основі машинного навчання, що також дасть можливість автоматизувати процес участі в тендері. Для цього потрібно сформувавши навчальний зразок, який може навчити систему розпізнавати організаторів тендерів, без додаткових розрахунків.

У зв'язку з цим можна вважати, що розробка методу формування навчальної вибірки для сегментації організаторів тендерів на основі машинного навчання є одним із найбільш перспективних напрямків у електронних тендерах.

**Метою** дипломної роботи є розробка методу сегментації організаторів тендерів на основі машинного навчання.

Досягнення цієї мети зумовило потребу теоретичних розробок, визначення та послідовного вирішення таких завдань:

1. визначити суть та особливості організації тендерів;
2. дослідити методи аналізу тендерних закупівель;
3. провести аналіз існуючих підходів на основі машинного навчання;
4. дослідити методи сегментації на основі машинного навчання;
5. дослідити методи кластеризації на основі машинного навчання;
6. розробити метод формування навчальної вибірки для сегментації організаторів тендерів на основі машинного навчання;
7. провести реалізацію методу сегментації організаторів тендерів на основі машинного навчання.

Завдання 1-3 розглянуто в параграфах 1.1, 1.2 та 1.3 відповідно, а виконання завдань 4-7 представлено у параграфах 2.1-3.3.

## Висновки до розділу 1:

Визначено класифікацію, що дозволяє сформувавши уявлення про те, що тендерам як особливому способу укладання договорів початку властива вільна конкуренція. Саме наявністю елемента змагальності укладення договору за допомогою оголошення тендерів відрізняється від звичайного способу виникнення договірних зобов'язань шляхом акцепту оферти.

Також, встановлено, що контроль повинен здійснюватися на всіх етапах процесу закупівлі. Контроль закупівель за державний коштом повинен також покривати ефективність використання закуплених об'єктів, їх цільове використання та зберігання. Суттєві порушення та відхилення у державних закупівлях свідчать про послаблення функції внутрішнього контролю в умовах певної інституційної підтримки. Встановлені інформаційні зв'язки, які пов'язують три напрямки процесу закупівель (самовизначення потреб, які необхідні в першу чергу для планування; прямих закупівель та виконання контрактів), відповідатимуть договірним, правовим та нормативним вимогам закупівельного середовища.

Досліджено, що на відміну від аналогів розроблений метод формування навчальної вибірки для сегментації організаторів тендерів на основі машинного навчання, дозволить сформувавши навчальну вибірку, на основі якої можна буде розробити систему для вибору організаторів тендерів на основі машинного навчання.

## 2 МЕТОД СЕГМЕНТАЦІЇ ОРГАНІЗАТОРІВ ТЕНДЕРІВ НА ОСНОВІ МАШИННОГО НАВЧАННЯ

### 2.1 Метод сегментації

У зв'язку з поставленою задачею сегментування організаторів тендеру за станом або подією, аналізу сегментів тендерної бази, прогнозування поведінки клієнтів і розробку конкретних рекомендацій щодо «цільового» впливу на сегменти. Для сегментації було обрано метод Recency Frequency Monetary (RFM, «рецепт, частота, гроші»).

Концепція *RFM*-аналізу була введена в роботі [58] і виявилася дуже ефективною в застосуванні до бази даних в маркетингу [59]. Даний вид аналізу ґрунтується на трьох змінних.

1. Recency - давність покупки (період, який пройшов з моменту останньої покупки до сьогодні). Чим менше часу минуло з моменту останньої покупки, тим ймовірніше, що споживач повернеться знову.

Слід зазначити, що з кожної сфери бізнесу поняття «купівля» своє.

Це може бути як покупка телевізора (якщо йдеться про звичайний магазин, інтернет-магазин аудіо-, відеотехніки), так і перехід на новий тарифний план (якщо йдеться про інтернет-провайдер, веб-студію тощо) і взагалі будь-яка зміна клієнтом попередньої моделі співробітництва.

2. Frequency - частота покупок (скільки всього покупок зробив споживач за історію спілкування з компанією). Вважається, що чим більше робить покупок той чи інший споживач, тим вища ймовірність того, що в наступному періоді він знову зробить покупку.

3. M-netary – грошова цінність покупок (загальна сума грошей, яку було здійснено покупки). Чим більше коштів витратив споживач, тим із вищою ймовірністю він їх витратить знову.

***R-змінна.*** Ймовірність приналежності до позитивного класу негативно залежить від ресенсу: чим більше часу минуло з моменту останнього замовлення,

тим менша ймовірність майбутнього замовлення. Типовий вид кривої recency представлений на рис. 1.

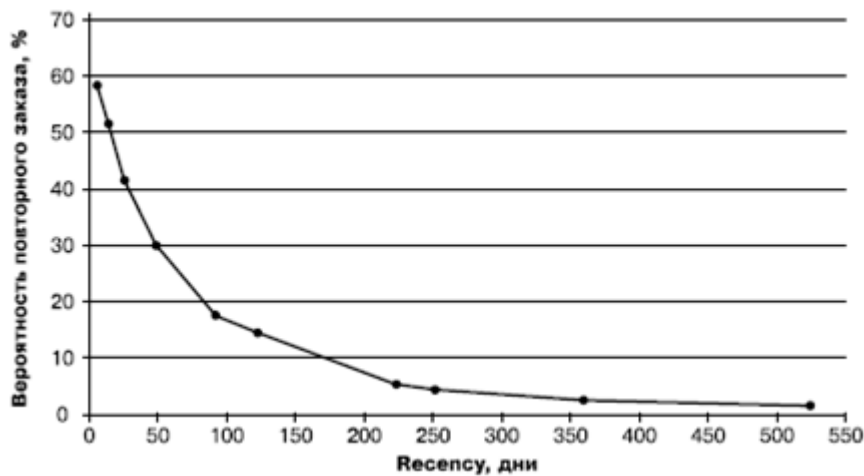


Рис. 2.1. Залежність ймовірності повторного замовлення від давності останнього замовлення

Покупцеві властиво робити замовлення повторно через невеликий проміжок часу, тобто. поки що покупець ще не забув про існування компанії, з якою взаємодіє. Більше того, набуваючи звички, покупець відчуває необхідність робити звичну дію дедалі частіше.

Коли ж покупець втрачає звичку, інтервал між останнім його замовленням та майбутнім, якого, швидше за все, не буде вже ніколи, зростає нескінченно. Ефект частоти та проміжку часу між наступними діями відкриває ще один постулат, який пояснюється нижче.

***F-змінна.*** Ймовірність приналежності до позитивного класу позитивно залежить від frequency: чим більше замовлень зробив клієнт, тим більша ймовірність, що він зробить ще одне замовлення. Типовий вид кривої frequency представлений на рис. 2.2.

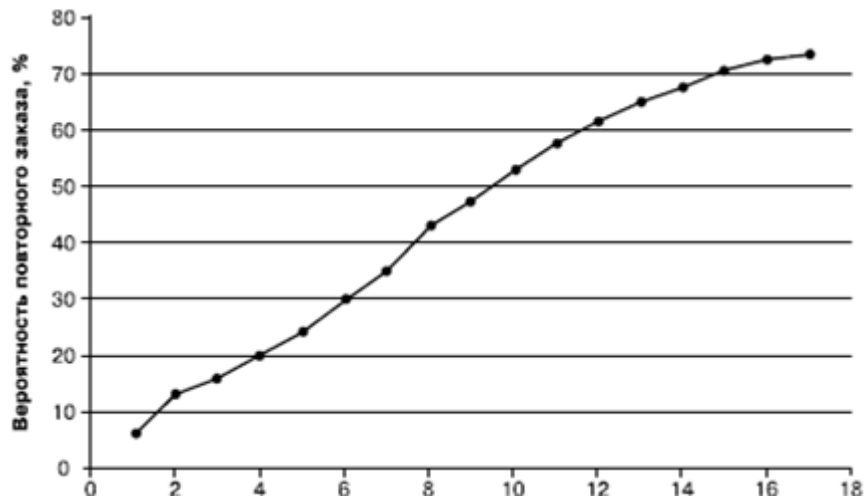


Рис. 2.2. Залежність ймовірності повторного замовлення від кількості зроблених замовлень

***M-змінна*** . Ймовірність приналежності до позитивного класу позитивно залежить від  $m$ -netary: що більше витрачено грошей, то більша ймовірність майбутнього замовлення. Цей постулат також очевидний і вимагає особливих пояснень. Типовий вид кривої показано на рис. 2.3.

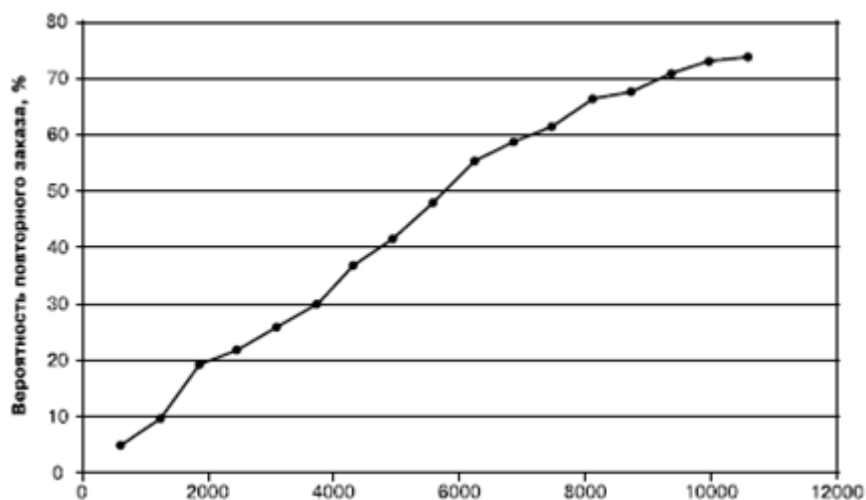


Рис. 2.3. Залежність ймовірності повторного замовлення від кількості грошей, витрачених на всі замовлення

### ***Методика проведення RFM-аналізу***

1 . *Дослідження Resency* . Для початку тут потрібно визначитися з тим, що буде критерієм активності клієнта. Це може бути покупка, відвідування магазину або навіть перехід за посиланням в Інтернеті на сайт компанії. Все залежить від того, які цілі переслідує компанія, проводячи аналіз, та від

специфіки її діяльності. Дане дослідження проводиться у наступній послідовності:

- кожного клієнта визначається дата останньої купівлі;
- для кожного клієнта розраховується давність покупки як різниця між поточною датою та датою останньої покупки;
- отримані дані розбиваються п'ять груп (квантилей). Кожен споживач у своїй отримує ідентифікатор від 1 до 5 залежно з його активності для здійснення покупок.

2. *Визначення Frequency* . Принцип розподілу споживачів на п'ять груп тут такий самий, як і в *Recency*. Тільки замість виділення часових відрізків потрібно визначити, яка кількість покупок буде критерієм входження споживача до тієї чи іншої групи. Наприклад, 20 і більше покупок – п'ята група, 0 – 2 покупки – перша група.

Це дослідження проводиться в наступній послідовності:

- для кожного споживача визначити кількість покупок за певний період;
- розбити отримані дані на 5 груп. Споживачам, які здійснили найбільше покупок, буде присвоєно код  $F=5$ , найменш активні покупці отримають  $F=1$ .

3. *Класифікація за параметром M-netary* . Тут так само, як і в попередніх пунктах. Приміщення споживача у ту чи іншу групу складає основі величини витрачених для купівлі коштів.

Дане дослідження проводиться у наступній послідовності:

- кожному за споживача визначити суму витрачених грошей;
- розбити отримані дані на 5 груп. Споживачам, які витратили найбільші суми, надається код  $M = 5$ ; споживачам, які витратили менші суми, –  $M=1$ .

4. *Об'єднання результатів* , отриманих на попередніх етапах аналізу.

Кожен клієнт при цьому отримає код *RFM* , що складається із трьох цифр.

Можливе виникнення збігаються спостережень – 2 і більше однакових значень нещодавності, частоти, грошової суми. Бажано, щоб приблизно однакове число клієнтів потрапляло в кожен категорію, але велика кількість

значень, що збігаються, може вплинути на розподіл клієнтів за категоріями [12]. Існує два методи обробки збігаються спостережень.

Перший метод - відносити збігаються спостереження до однієї групи. Цей метод відносить збігаються значення однієї категорії незалежно від цього, як вони впливають розподіл категорій. Тим самим забезпечується послідовне угруповання: якщо у двох клієнтів – те саме значення нещодавності, то їм завжди буде присвоєна одна і та ж оцінка нещодавності. Однак у крайньому випадку у нас може бути ситуація, коли є 1000 клієнтів і 500 з них зробили свою останню покупку в один і той же день. За 5 категорій 50% клієнтів мають отримати оцінку нещодавності 5 замість бажаних 20%. Зазначимо, що разом із викладеним способом угруповання метод послідовного угруповання трохи складніший для інтерпретації оцінок частоти та грошової суми, оскільки оцінки частот задаються всередині груп, сформованих за оцінками нещодавності, а оцінки грошових сум – усередині груп, сформованих за оцінками частот. Таким чином, два клієнти з одним і тим же значенням частоти не можуть мати однакову оцінку частоти, якщо у них різні оцінки нещодавності, незалежно від обробки збігаються спостережень.

Розносити збігаються спостереження за сусідніми групами слід випадковим чином. Метод гарантує рівномірний розподіл категорій. Метод впливає на початкові спостереження. Він лише використовується, щоб усунути можливість різного тлумачення спостережень, що збігаються. Поряд з тим, що він забезпечує рівномірний розподіл категорій (приблизно однакову кількість клієнтів у кожній категорії), метод дозволяє отримати різні оцінки для клієнтів, які, мабуть, схожі, або ідентичні, за значеннями нещодавності, частоти та/або грошової суми – особливо коли загальна кількість клієнтів відносно невелика та/або кількість збігаються спостережень велика.

Рознесення споживачів з різних груп (категоризація) може здійснюватися з допомогою кількох методів [23].

У вкладеному методі категоризації спочатку проводиться угруповання (ранжування) за показником нещодавності (цей показник вважається найбільш

важливим), потім – за показником частоти всередині кожної групи по нещодавності, потім – за показником грошової суми всередині кожного поєднання груп за нещодавністю і частотою. Наприклад, якщо для клієнта ми маємо значення  $R = 1$ ,  $F = 5$ ,  $M = 5$ , це означає, що сума покупок даного клієнта входить до числа 20% найбільших покупок у підгрупі клієнтів, що належать до 20% клієнтів у вибірці, які здійснили покупки давно, і одночасно до 20% клієнтів з найбільшою частотою покупок усередині цієї групи давно нічого не купували клієнтів. Цей метод характеризується більш рівномірним розподілом *RFM* оцінки. Його недолік полягає в тому, що оцінки частоти, що проранжують, і сум покупок складно коректно інтерпретувати.

**Незалежний метод** розбиває кожен із трьох показників на 5 груп по 20% спостережень у кожній незалежно один від одного, таким чином він надає однакову «важливість» кожному з трьох показників. Наприклад, якщо для клієнта ми маємо значення  $M = 5$ , то сума покупок для даного клієнта входить до 20% найбільших покупок у всій вибірці клієнтів. Інтерпретація кожного з трьох показників є однозначною і не становить складності.

Оцінка частоти 5 одного клієнта позначає те саме, як і оцінка частоти 5 іншого клієнта, незалежно від своїх оцінок нещодавності. На невеликих вибірках проявляється недолік методу - розподіл *RFM* оцінок стає менш рівномірним. У табл. 1 наводиться порівняння розглянутих методів категоризації.

У табл. 2.1 зарахування збігаються спостережень до однієї групи дає нерівномірний розподіл категорій: 5 (10%), 4 (40%), 3 (20%), 2 (20%), 1 (10%).

Другий спосіб дає 20% спостережень у кожній категорії, але досягається це за рахунок того, що 4 спостереження з датою 10/28/2006 розносяться за 3 різними категоріями, а 2 спостереження з датою 8/13/2006 віднесені до різних категорій.



Таблиця 2.1 - Порівняння методів

ID	Остання покупка (давність)	Ранжування значень нещодавності	
		Відносити співпадають спостереження до однієї групи	Розносити збігаються спостереження по сусіднім групам випадковим чином
1	10/29/2020	5	5
2	10/28/2020	4	4
3	10/28/2020	4	4
4	10/28/2020	4	5
5	10/28/2020	4	3
6	9/21/2020	3	3
7	9/21/2020	3	2
8	8/13/2020	2	2
9	8/13/2020	2	1
10	6/20/2020	1	1

**Визначення змінної R.** Від поточної дати нам потрібно відібрати дату останньої покупки клієнта. Отримаємо кількість днів. Умовно, якщо ми беремо період 60 днів, можемо поділити їх на 5 груп. Вийде кожна група по 12 днів.

Перша група включає клієнтів, які зробили покупку протягом найближчих 12 днів. Їм надається 5 балів. Наступна група з покупкою від 13 до 24 днів отримує 4 бали і т.д.

**Визначення змінної F.** Нам знадобляться дані про кількість покупок. Для того щоб розділити клієнтів на 5 груп у класі  $F$ , потрібно вибрати максимальну та мінімальну кількість чеків/накладних. Зверніть увагу на накладні, які були зроблені одного дня на одного клієнта. Вам не потрібно враховувати їх для змінної  $F$ . Один день – максимум одна накладна. Якщо клієнту «добирали» товар, був довіз або через годину після замовлення він згадав, що щось не купив, ці «зайві» накладні негативно вплинуть на точність статистики. Отже, ми маємо максимальну кількість чеків та мінімальну. Припустимо це 27 та 2. Визначаємо межі для класів –  $(27+2)/5 = 5,8$ , округляємо до 6; 5 балів одержують клієнти з кількістю чеків від 21 до 27; 4 - від 15 до 20 і т.д.

**Визначення змінної  $M$ .** Визначається із загальної суми покупок. Підхід такий самий, як і при визначенні змінної  $F$ . Зверніть увагу на явні відриви загального масиву даних. Наприклад, якщо 27 чеків - це єдиний показник, а після нього основна маса покупок починається, скажімо, з 21, то, можливо, варто зрушити межі груп. Це стосується кожного із трьох показників. Період аналізу індивідуальний кожної компанії.

*RFM*-аналіз може послужити корисну службу маркетологу, якщо стоїть мета - вивчити клієнтську базу і зрозуміти, що, кому, коли, в яких обсягах та за якою ціною пропонувати. Зрештою, це один із способів заробити більше, використовуючи при цьому колишні ресурси [25].

*RFM*-аналіз - метод аналізу, класифікації та сегментації клієнтів за ступенем їхньої лояльності. З іншого боку, його застосовують визначення прибутковості клієнтів, виявлення ключових клієнтів, вивчення продуктового асортименту, оцінки ймовірності догляду клієнтів.

Цей аналіз може бути корисним для розподілу коштів, включаючи людські ресурси, але також може застосовуватися при плануванні комунікаційної політики, спрямованої на клієнтів. *RFM*-аналіз може бути корисним під час виконання адресної розсилки з пропозиціями товарів чи послуг за обраною групою зацікавлених клієнтів; при відстеженні результатів маркетингових акцій шляхом виміру відсотка відгуків клієнтів пропозиції.

Плануючи поштову розсилку, необхідно розуміти, що і кому має бути відправлено, – тільки тоді цей метод просування дасть позитивний результат.

Можливе використання *RFM*-аналізу для організації ефективної рекламної діяльності.

Можна визначити, наскільки доцільно витратити рекламний бюджет залучення тієї чи іншої клієнта. По історії покупок можна вирішити, чи заслуговує клієнт особливого маркетингового підходу.

Сьогодні більшість комерційних організацій веде свою базу даних клієнта. Це інтернет-магазини, стільникові оператори, директ-маркетингові агенції, мережі супермаркетів, мережі ресторанів, аптечні мережі, авіакомпанії,

туристичні агенції, банки тощо. Використання *RFM*-аналізу дозволяє аналізувати дані про клієнтів та їх споживчі переваги та ефективно користуватися отриманою інформацією при плануванні маркетингової діяльності.

*RFM*-аналіз є інструментом, що дозволяє проводити оцінку споживачів на основі їх минулих дій. Модель *RFM* простежує споживача через історії придбання ними певних продуктів та групує споживачів у сегменти з однаковими розмірами покупок, частотою покупок та часом їх здійснення. *RFM*-аналіз застосовний тільки стосовно споживачам, що відбулися, оскільки у потенційних споживачів ще немає історії покупок.

Цей метод сегментації клієнтської бази однаково підходить як споживчого ринку, так корпоративного.

Особливо питання аналізу бази даних клієнтів є актуальним у контексті стрімкого розвитку інтернет-торгівлі. Інтернет-магазинів стає все більше, але лише деякі отримують достатню вигоду з онлайн-торгівлі. Зрозуміти переваги своїх клієнтів допоможе їхня сегментація за певними параметрами, у тому числі за допомогою *RFM*-аналізу.

*RFM*-аналіз можна використовувати при прогнозуванні поведінки споживачів.

Передбачувальні можливості трьох змінних *RFM*-аналізу ранжуються аналогічно послідовності букв у назві *RFM*: *recency* має найкращу можливість прогнозування, другим йде *frequency* і останнім за силою є *m-netary* [20].

## 2.2 Метод пониження розмірності

У багатовимірному статистичному аналізі кожен об'єкт описується вектором, розмірність якого довільна (але однакова для всіх об'єктів). Однак людина може безпосередньо сприймати тільки числові дані або точки на площині. Набагато складніше проаналізувати скупчення точок в тривимірному

просторі. Пряме сприйняття більш високовимірних даних неможливе. Тому цілком природно хотіти перейти від багатовимірної вибірки до дрібних даних, щоб «на них можна було подивитися».

Крім прагнення до ясності, існують і інші мотиви для зменшення виміру. Фактори, від яких змінна не залежить від дослідника, лише перешкоджають статистичному аналізу. По-перше, ресурси витрачаються на збір інформації про них. По-друге, можна довести, що їх включення в аналіз погіршує властивості статистичних процедур (зокрема, збільшує дисперсію оцінок параметрів і характеристик розподілів). Тому від таких факторів бажано позбутися.

При аналізі багатовимірних даних зазвичай розглядається не одне, а багато завдань, зокрема шляхом вибору незалежних і залежних змінних по-різному. Тому розглянемо проблему зменшення виміру в наступному формулюванні. Дається багатовимірний зразок. Необхідно перейти від нього до набору векторів меншого виміру, максимально зберігаючи структуру вихідних даних, по можливості не втрачаючи інформації, що міститься в даних. Завдання вказується в кожному конкретному способі зменшення виміру.

Основний метод компонентів є одним з найбільш часто використовуваних методів зменшення виміру. Його основна ідея полягає в тому, щоб послідовно визначати області, в яких дані мають найбільшу розбіжність. Нехай вибірка складається з векторів, які порівну розподілені з вектором  $X = (x(1), x(2), \dots, x(n))$ . Розглянемо лінійні комбінації

$$Y(\lambda(1), \lambda(2), \dots, \lambda(n)) = \lambda(1)x(1) + \lambda(2)x(2) + \dots + \lambda(n)x(n),$$

де

$$\lambda^2(1) + \lambda^2(2) + \dots + \lambda^2(n) = 1.$$

тут вектор  $\lambda = (\lambda(1), \lambda(2), \dots, \lambda(n))$  лежить на одиничній сфері в  $n$ -мірному просторі.

У методі основних компонент передусім знаходять напрям максимального розкиду, тобто. таке  $\lambda$ , при якому досягає максимуму дисперсія випадкової

величини  $Y(\lambda) = Y(\lambda(1), \lambda(2), \dots, \lambda(n))$ . Тоді вектор задає першу головну компоненту, а величина  $Y(\lambda)$  є проекцією випадкового вектора  $X$  на вісь першої головної компоненти.

Потім, з точки зору лінійної алгебри, розглянемо гіперплан в  $n$ -мірному просторі, перпендикулярно першому головному компоненту, і спроектуйте всі елементи зразка на цю гіперплощину. Розмірність гіперплощини на 1 менше, ніж розмірність вихідного простору.

В аналізованому гіперплощині процедуру повторюють. У ньому знайти напрямок найбільшого розсіювання, тобто. Другий основний компонент. Потім виберіть гіперплощину перпендикулярно першим двом основним компонентам. Його розмірність на 2 менше, ніж розмірність вихідного простору. Потім - наступна ітерація.

З точки зору лінійної алгебри мова йде про побудову нової основи в  $n$ -мірному просторі, орти якого є основними компонентами.

Дисперсія, що відповідає кожному новому головному компоненту, менша, ніж для попереднього. Зазвичай зупиняються, коли він менше зазначеного порогу. Якщо вибрати  $k$  основних компонентів, це означає, що з  $n$ -мірного простору можна було перейти до  $k$ -розмірного, тобто. зменшити розмірність від  $n$ - до  $k$ , практично не спотворюючи структуру вихідних даних.

Для візуального аналізу даних часто використовують проекції вихідних векторів на площині перших двох основних компонентів. Зазвичай добре видно структуру даних, виділяються компактні скупчення об'єктів і векторів, які виділяються.

Багатовимірне масштабування. Використовуючи відстань (міри близькості, відмінності)  $d(X, Y)$  між ознаками  $X$  і встановлено великий клас методів багатовимірного масштабування [15, 16]. Основна ідея цього класу методів полягає в представленні кожного об'єкта точкою геометричного простору (зазвичай розміри 1, 2 або 3), координатами яких є значення прихованих (латентних) факторів, які разом адекватно описують об'єкт. При цьому взаємозв'язок між об'єктами замінюється взаємозв'язком між точками - їх

представниками. Таким чином, дані про схожість об'єктів - відстані між точками, дані про перевагу - взаємне розташування точок [17].

На практиці використовується ряд різних моделей багатовимірного масштабування. Всі вони мають проблему оцінки істинного виміру факторного простору. Розглянемо цю проблему на прикладі обробки даних про схожість об'єктів за допомогою метричного масштабування.

Нехай буде  $n$  об'єктів  $O(1), O(2), \dots, O(n)$ , кожна пара об'єктів  $O(i), O(j)$  з урахуванням міри їх подібності  $s(i, j)$ . Ми припускаємо, що завжди  $s(i, j) = s(j, i)$ . Походження чисел  $s(i, j)$  не має значення для опису роботи алгоритму. Вони можуть бути отримані або шляхом прямого вимірювання, або з використанням експертів, або шляхом розрахунку набору описових характеристик, або іншим чином.

У Евклідовому просторі розглянуті  $n$  об'єкти повинні бути представлені конфігурацією  $n$  точок, а мірою близькості репрезентативних точок є Евклідова відстань  $d(i, j)$  між відповідними точками. Ступінь відповідності між множиною об'єктів і набором точок, що їх представляють, визначається шляхом порівняння подібності матриць  $\|s(i, j)\|$  відстані  $\|d(i, j)\|$ . Функціональна метрична схожість має форму

$$S = \sum_{i,j} |s(i, j) - d(i, j)|^2$$

Геометричну конфігурацію слід вибирати так, щоб функціонал  $S$  досягав найнижчого значення [17].

Зауваження. У неметричному масштабуванні замість близькості самих мір близькості і самих відстаней розглядається близькість порядків для набору заходів близькості і набір відповідних відстаней. Замість функціонального  $S$  використовуються аналоги коефіцієнтів кореляції Спірмена і Кендалл рангу. Іншими словами, неметричне масштабування засноване на припущенні, що міри близькості вимірюються за порядковою шкалою.

Нехай Евклідовий простір має розмірність  $m$ . Розглянемо мінімальну середню квадратну помилку

$$\alpha_m = \frac{2}{n(n-1)} \min S,$$

де мінімум береться для всіх можливих конфігурацій  $n$  точок в  $m$ -вимірному Евклідовому просторі. Можна показати, що аналізується мінімум досягається на певній конфігурації. Зрозуміло, що зі збільшенням  $m$  значення  $\alpha_m$  зменшується монотонно (точніше, не збільшується). Ми можемо показати, що з  $m > n - 1$  дорівнює 0 (якщо  $s(i, j)$  є метрикою). Щоб збільшити можливості осмисленого тлумачення, бажано діяти в просторі якомога менше. Однак розмірність повинна бути обрана таким чином, щоб точки представляли об'єкти без особливих спотворень. Виникає питання: як оптимально вибрати розмірність, тобто. натуральне число  $m$ ?

В рамках детермінованого аналізу даних обґрунтованої відповіді на це питання, ймовірно, немає. Тому необхідно вивчити поведінку  $\alpha_m$  в певних імовірнісних моделях. Якщо міри близькості  $s(i, j)$  є випадковими змінними, розподіл яких залежить від «істинного виміру»  $m_0$  (і, можливо, від будь-яких параметрів), то в класичному математичному і статистичному стилі ми можемо встановити задачу оцінки  $m_0$ , пошуку багатих оцінок і т.д.

Почнемо будувати імовірнісні моделі. Прийmemo, що об'єкти є крапками в евклідовому просторі розмірності  $k$ , де  $k$  досить велике. Те, що «справжня розмірність» дорівнює  $m_0$  означає, що всі ці точки лежать на гіперплощині розмірності  $m_0$ . Прийmemo для визначеності, що сукупність точок, що розглядаються, являє собою вибірку з кругового нормального розподілу з дисперсією  $\sigma^2(0)$ . Це означає, що об'єкти  $O(1), O(2), \dots, O(n)$  є незалежними в сукупності випадковими векторами, кожен з яких будується як  $\zeta(1)e(1) + \zeta(2)e(2) + \dots + \zeta(m_0)e(m_0)$ , де  $e(1), e(2), \dots, e(m_0)$  – ортонормальний базис у підпросторі розмірності  $m_0$ , в якому лежать розглянуті точки, а  $\zeta(1), \zeta(2), \dots, \zeta(m_0)$  – незалежні в сукупності одновимірні нормальні випадкові величини з математичним очікуванням і дисперсією  $\sigma^2(0)$ .

Розглянемо дві моделі отримання мір близькості  $s(i, j)$ . У першій з них  $s(i, j)$  відрізняються від евклідової відстані між відповідними точками

через те, що точки відомі з спотвореннями. Нехай  $z(1), z(2), \dots, z(n)$  - Розглянуті точки. Тоді

$$s(i, j) = d(c(i) + \varepsilon(i), c(j) + \varepsilon(j)), \quad i, j = 1, 2, \dots, n,$$

де  $d$  – евклідова відстань між точками в  $k$ - вимірному просторі, вектора  $\varepsilon(1), \varepsilon(2), \dots, \varepsilon(n)$  являють собою вибірку з кругового нормального розподілу в  $k$ - вимірному просторі з нульовим математичним очікуванням та ковараційною матрицею  $\sigma^2(1)I$  де  $I$  - одинична матриця. Інакше кажучи,  $\varepsilon(i) = \eta(1)e(1) + \eta(2)e(2) + \dots + \eta(k)e(k)$ , де  $e(1), e(2), \dots, e(k)$  – ортонормальний базис у  $k$ - мірному просторі, а  $\{\eta(i, t), i = 1, 2, \dots, n, t = 1, 2, \dots, k\}$  – сукупність незалежних у сукупності одновимірних випадкових величин з нульовим математичним очікуванням та дисперсією  $\sigma^2(1)$ .

У другій моделі спотворення накладено безпосередньо на самі відстані:

$$s(i, j) = d(c(i), c(j)) + \varepsilon(i, j), \quad i, j = 1, 2, \dots, n, \quad i \neq j,$$

де  $\{\varepsilon(i, j), i, j = 1, 2, \dots, n\}$  – незалежні разом нормальні випадкові величини з математичним очікуванням 0 і дисперсією  $\sigma^2(1)$ .

У роботі [18] показано, що для обох сформульованих моделей мінімум середнього квадрата помилки  $\alpha_m$  при  $n \rightarrow \infty$  схоже на ймовірність до

$$f(m) = f_1(m) + \sigma^2(1)(k - m), \quad m = 1, 2, \dots, k,$$

де

$$f_1(m) = \begin{cases} \sigma^2(1)(m_0 - m), & m < m_0, \\ 0, & m \geq m_0. \end{cases}$$



Таким чином, функція  $f(m)$  є лінійною на інтервалах  $[1, m_0]$  і  $[m_0, k]$ , а на першому інтервалі зменшується швидше, ніж на другому. З цього випливає, що статистика

$$m^* = \underset{m}{\text{Arg min}} (\alpha_{m+1} - 2\alpha_m + \alpha_{m-1})$$

є здатною оцінкою істинної вимірності  $m_0$ .

Таким чином, з імовірнісної теорії випливає рекомендація - як оцінка розмірності факторного простору для використання  $m^*$ . Відзначимо, що така рекомендація була сформульована як евристична одним із засновників багатовимірного масштабування Я. Краскаля [15]. Він спирався на досвід практичного використання багатовимірного масштабування та обчислювальних експериментів. Імовірнісна теорія дозволила обґрунтувати цю евристичні рекомендації.

### 2.3 Методи кластеризації на основі машинного навчання

При вирішенні реальних проблем у більшості випадків дані не матимуть зумовлених міток, тому потрібно розробити моделі машинного навчання, які можуть правильно класифікувати ці дані, знайшовши для себе деяку спільність у функціях, які будуть використовуватися для прогнозування класів за новими даними.

Загальний процес, який при розробці моделі навчання без вчителя, можна узагальнити (рис. 2.4).

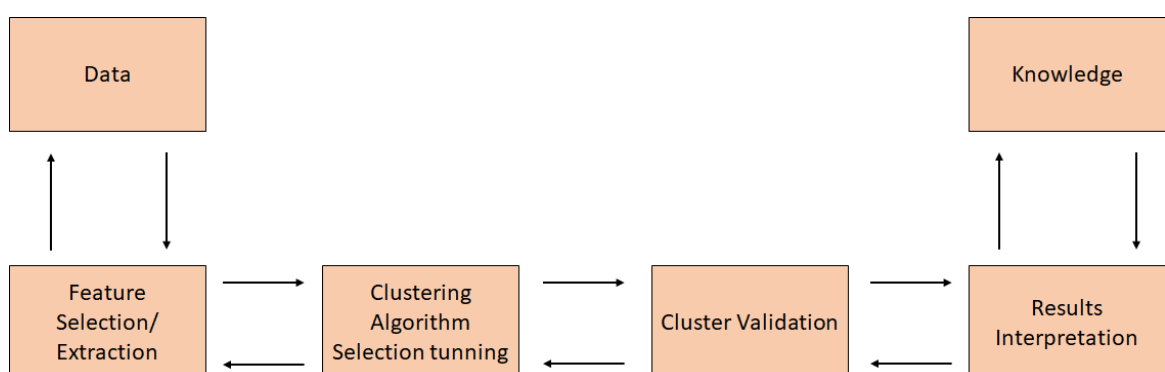


Рис.2.4. Процес навчання без вчителя

Основні програми для навчання без вчителя:

- Сегментування наборів даних щодо деяких загальних атрибутів.
- Виявлення аномалій, які не підходять до жодної групи.
- Спростіть набори даних, агрегуючи змінні зі схожими атрибутами.
- Таким чином, основна мета полягає в тому, щоб вивчити внутрішню (і зазвичай приховану) структуру даних.

Ці методи можуть бути зведені до проблеми кластеризації, які намагаються вирішити неконтрольоване навчання.

Загалом, мета кластеризації - знайти різні групи в елементах даних. Для цього алгоритми кластеризації знаходять структуру даних так, що елементи одного кластера (або групи) більш схожі один на одного, ніж елементи з різних кластерів.

Алгоритми K-Means надзвичайно прості у реалізації та дуже ефективні у обчислювальному відношенні. Це основні причини, що пояснюють їхню популярність. Але вони дуже добре ідентифікують класи, коли мають справу з групами, які мають сферичної форми розподілу.

Алгоритми K-Means спрямовані на пошук та групування в класах точок даних, які мають високу схожість між ними. У термінах алгоритму ця схожість сприймається як протилежність відстані між точками даних. Чим ближче точки даних, тим більше вони схожі і з більшою ймовірністю належать до одного кластера.

Найбільш часто використовувана відстань у K-середніх означає квадрат Евклідова відстані. Приклад цього відстані між двома точками  $x$  і  $y$  в  $m$ -мірному просторі це:

$$d(\mathbf{x}, \mathbf{y})^2 = \sum_{j=1}^m (x_j - y_j)^2 = \|\mathbf{x} - \mathbf{y}\|_2^2$$

Ось  $J$  це  $-я$  вимір (або стовпець елементів) точок вибірки  $x$  і  $y$ .

Інерція кластера - це ім'я, дане сумі квадратичних помилок у контексті кластеризації, і наведено таким чином:

$$SSE = \sum_{i=1}^n \sum_{j=1}^k w^{(i,j)} \left\| \mathbf{x}^{(i)} - \boldsymbol{\mu}^{(j)} \right\|_2^2$$

де  $\mu^{(j)}$  центр тяжкості для кластера  $J$ , а також  $w^{(i,j)}$  1, якщо зразок  $x^{(i)}$  знаходиться в кластері  $J$  і 0 інакше.

K-середні можна розуміти як алгоритм, який намагатиметься мінімізувати коефіцієнт інерції кластера.

Кроки алгоритму:

1. По-перше, нам потрібно вибрати  $k$ , кількість кластерів, які ми хочемо знайти.
2. Потім алгоритм випадково вибере центроїди кожного кластера.
3. Він буде призначений кожному пункту даних найближчому центроїду (з використанням евклідової відстані).
4. Буде обчислено інерцію кластера.
5. Нові центроїди розраховуватимуться як середнє значення точок, які належать центроїду попереднього кроку. Іншими словами, обчислюючи мінімальну квадратичну помилку точок даних до центру кожного кластера, переміщуючи центр до цієї точки
6. Повернутись до кроку 3.

Гіперпараметри K-середніх:

- Кількість кластерів: кількість кластерів та центроїдів для генерації.
- Максимум ітерацій: алгоритм за один прогін.
- Початкове число: число, протягом якого алгоритм працюватиме з різним насінням центроїдів. Кінцевий результат буде найкращим виходом числа, визначеного послідовними прогонами, з погляду інерції.

Вихідні дані для будь-якого фіксованого навчального набору не завжди будуть однаковими, тому що початкові центроїди встановлені випадковим чином, і це вплине на процес алгоритму.

Як зазначалося раніше, через характер евклідової відстані він не підходить для роботи з кластерами, які набувають несферичних форм.

Точки, які слід враховувати при застосуванні K-середніх:

- Елементи повинні бути виміряні в одній і тій же шкалі, тому може знадобитися виконати стандартизацію за z-шкалою або максимальне-мінімальне масштабування.
- При роботі з категоріальними даними ми використовуватимемо функцію `get dummies`.
- Дослідницький аналіз даних (EDA) дуже корисний для огляду даних та визначення того, чи є K-Means найбільш підходящим алгоритмом.
- Метод міні-пакета дуже корисний при великій кількості стовпців, проте він менш точний.

Вибір правильної кількості кластерів одна із ключових моментів алгоритму K-Means. Щоб знайти це число, є кілька методів:

- Польові знання
- Рішення для бізнесу
- Метод локтя

Будучи узгодженим із мотивацією та природою Data Science, метод локтя є кращим варіантом, оскільки для ухвалення рішення він спирається на аналітичний метод, підкріплений даними.

Метод локтя використовується визначення правильної кількості кластерів в наборі даних. Він працює шляхом побудови графіків зростаючих значень K порівняно із загальною помилкою, отриманою під час використання цього K.

$$\% \text{ Variance} = \frac{\text{Variance between groups}}{\text{Total variance}}$$

Ціль полягає в тому, щоб знайти  $k$ , який для кожного кластера не буде значно збільшувати дисперсію (Рис. 2.5).

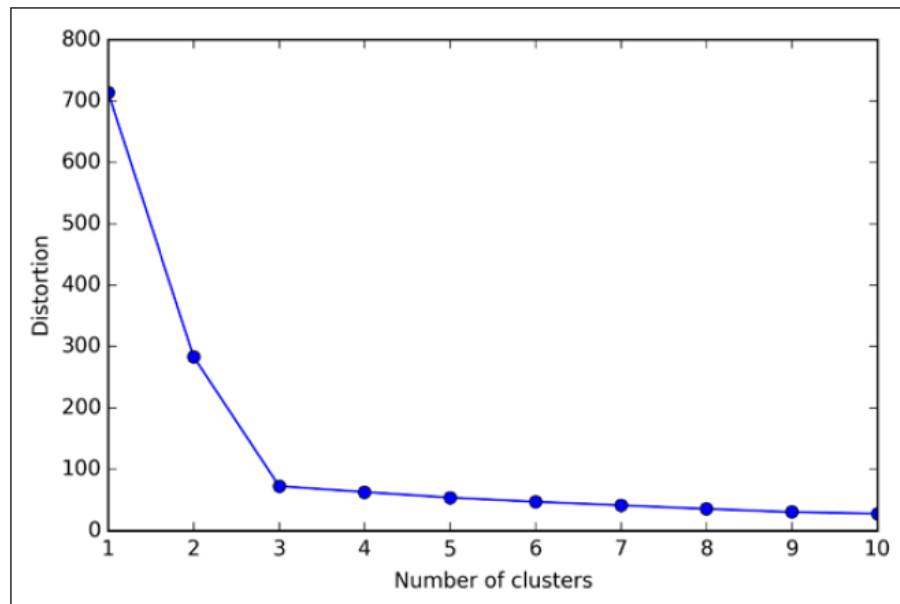


Рис.2.5. Визначення кількості кластерів

Хоча K-Means є відмінним алгоритмом кластеризації, він найбільш корисний, коли ми заздалегідь знаємо точну кількість кластерів і маємо справу з розподілами сферичної форми.

Дуже поширено використовувати алгоритм K-Means як зразок з метою оцінки продуктивності інших методів кластеризації.

Модель машинного навчання зможе укласти, що є два різних класу, не знаючи нічого з даних.

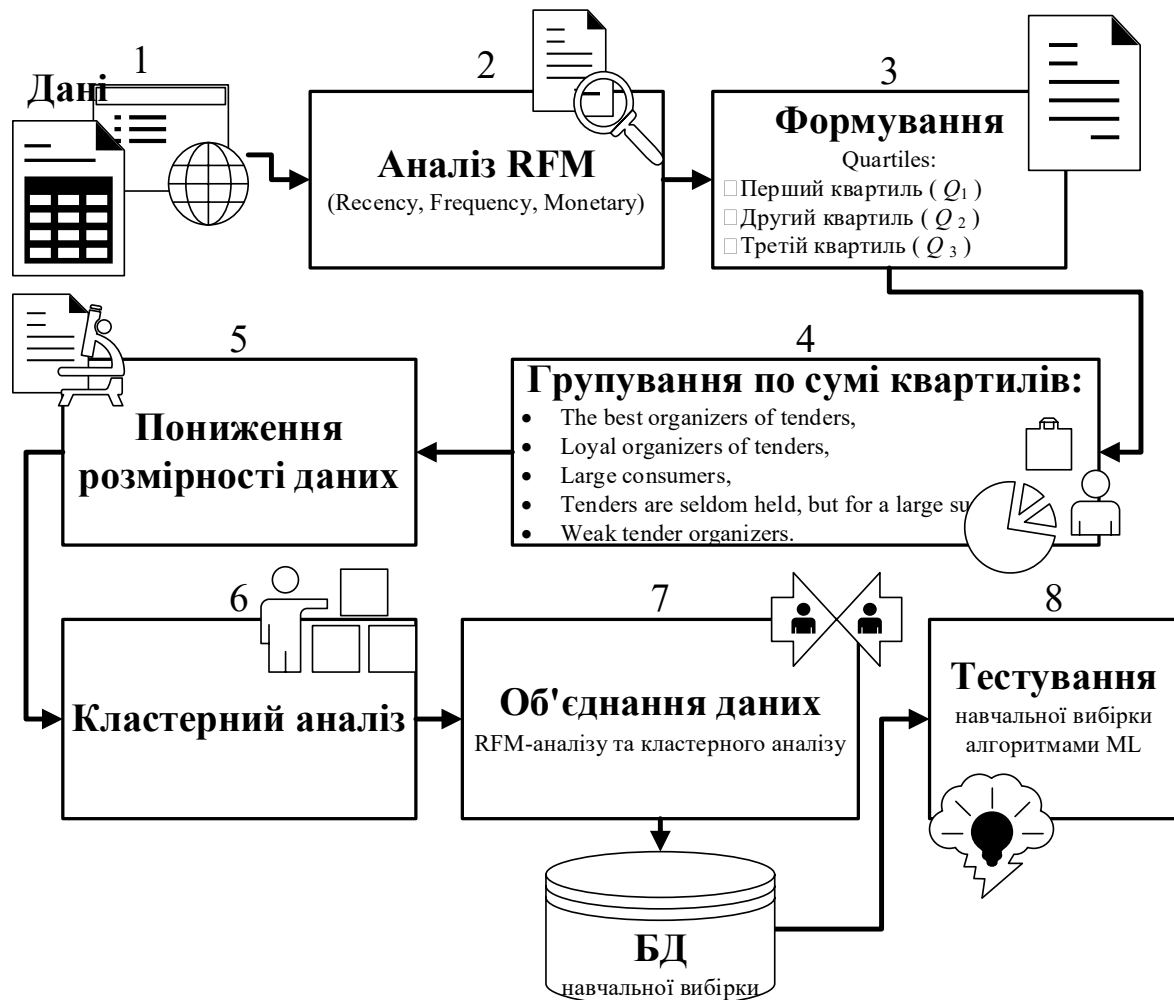
Ці неконтрольовані алгоритми навчання мають неймовірно широкий спектр застосувань і дуже корисні для вирішення реальних проблем, таких як виявлення аномалій, рекомендації щодо систем, групування документів або пошук клієнтів із загальними інтересами на основі їх покупок.

#### 2.4 Метод формування навчальної вибірки для сегментації організаторів тендерів на основі машинного навчання

Для визначення організаторів тендерів, які проводять найбільше за кількістю аукціонів та тих, які витрачають найбільшу кількість коштів по

завершенню угоди, авторами розроблено метод формування навчальної вибірки для сегментації організаторів тендерів на основі машинного навчання.

Запропонований метод ілюструється схематично (Рис.2.6) та представлений наступними кроками:



**Рис. 2.6.** Структура формування навчальної вибірки для сегментації організаторів тендерів на основі машинного навчання

1. Ввід вхідних даних (Блок 1): завершені угоди учасників тендерів.
2. RFM-аналіз (Блок 2). Аналіз RFM (Recency, Frequency, Monetary) - це техніка сегментації споживачів, яка використовує поведінку минулих угод, щоб розділити клієнтів на групи. Цей метод аналізу теж можна добре використати для сегментації організаторів тендерів. RFM-аналіз для сегментації організаторів тендерів буде, відбувається за трьома показниками:

- Resency - давність угоди - проміжок часу з моменту останньої угоди.
- Frequency – кількість угод - кількість угод за відповідний період.
- Monetary - сума завершених угод - сума всіх завершених угод за відповідний період.

3. Формування Quartiles (Блок 3). Quartiles ділять число точок даних на чотири частини, або чверті, на більш-менш однакового розміру. Дані повинні бути впорядковані від найменшого до найбільшого для обчислення кватилів. Три основні кватилі такі:

- Перший кватиль ( $Q_1$ ) визначається як середнє число між найменшим числом ( мінімальним ) та медіаною набору даних. Він також відомий як *нижчий* або *25-й емпіричний* кватиль, оскільки 25% даних знаходиться нижче цієї точки.
- Другий кватиль ( $Q_2$ ) є медіаною набору даних; таким чином, 50% даних лежить нижче цієї точки.
- Третій кватиль ( $Q_3$ ) - це середнє значення між медіаною та найвищим значенням ( максимумом ) набору даних. Він відомий як *верхній* або *75-й емпіричний* кватиль, оскільки 75% даних лежить нижче цієї точки.

4. Групування по сумі кватилів (Блок 4). Групування організаторів тендерів за наступними критеріями: The best organizers of tenders, Loyal organizers of tenders, Large consumers, Tenders are seldom held, but for a large sum і Weak tender organizers.

5. Пониження розмірності даних (Блок 5). Зниження розмірності є процесом скорочення кількості випадкових змінних шляхом отримання множини головних змінних. Виділення ознак та зниження розмірності можна об'єднати в один етап за допомогою методу головних компонент (МГК), лінійного розділювального аналізу (ЛРА), канонічного кореляційного аналізу (ККА) або розкладення невід'ємних матриць (РНМ). Дані по тендерних угодах є дуже розкидами, тому пониження розмірності є важливим кроком для подальшої кластеризації даних.

6. Кластеризація (Блок 6). Кластерний аналіз — задача розбиття заданої вибірки об'єктів (ситуацій) на підмножини, які називаються кластерами, так, щоб кожен кластер складався з схожих об'єктів, а об'єкти різних кластерів істотно відрізнялися. Кластерний аналіз є глибшим аналізом для сегментації організаторів тендерів, тому для навчальної вибірки є важливим. Даний аналіз поділяється на наступні етапи:

- Проведення дослідження.
- Підготовка даних до кластерного аналізу.
- Вибір методу кластерного аналізу.
- Вибір міри відстані між об'єктами та її обчислення.
- Вибір стратегії кластеризації.
- Застосування обраної стратегії для утворення кластерів.
- Перевірка результатів кластерного аналізу на осмисленість і їх інтерпретацію.

7. Об'єднання (Блок 7) даних RFM-аналізу та кластерного аналізу та внесення в БД. На основі цих даних можна провести навчання класифікації сегментації організаторів тендерів.

8. Тестування (Блок 8) навчальної вибірки на основі алгоритмів класифікації машинного навчання.

Для підтвердження розробленого методу формування навчальної вибірки для сегментації організаторів тендерів на основі машинного навчання в наступному розділі проведено його реалізацію.



## Висновки до розділу 2:

У розділі проведено дослідження методу сегментації, пониження розмірності та кластеризації.

Для поставленої мети роботи потрібно провести сегментування організаторів тендеру за станом або подією, аналізу сегментів тендерної бази, прогнозування поведінки клієнтів і розробку конкретних рекомендацій щодо «цільового» впливу на сегменти. Тому для сегментації було обрано метод Resency Frequency Monetary (RFM, «рецепт, частота, гроші»).

При аналізі багатовимірних даних зазвичай розглядається не одне, а багато завдань, зокрема шляхом вибору незалежних і залежних змінних по-різному. Тому розглянемо проблему зменшення виміру в наступному формулюванні. Дається багатовимірний зразок. Необхідно перейти від нього до набору векторів меншого виміру, максимально зберігаючи структуру вихідних даних, по можливості не втрачаючи інформації, що міститься в даних. Завдання вказується в кожному конкретному способі зменшення виміру.

При вирішенні поставленої задачі дані не мають зумовлених міток, тому потрібно розробити моделі машинного навчання, які можуть правильно класифікувати ці дані, знайшовши для себе деяку спільність у функціях, які будуть використовуватися для прогнозування класів за новими даними.

На основі досліджених методів для визначення організаторів тендерів, які проводять найбільше за кількістю аукціонів та тих, які витрачають найбільшу кількість коштів по завершенню угоди, авторами розроблено метод формування навчальної вибірки для сегментації організаторів тендерів на основі машинного навчання.

### 3 РЕАЛІЗАЦІЯ МЕТОДУ СЕГМЕНТАЦІЇ ОРГАНІЗАТОРІВ ТЕНДЕРІВ НА ОСНОВІ МАШИННОГО НАВЧАННЯ

#### 3.1 Інформаційно-аналітичне забезпечення публічних закупівель

Закупівлі є важливою частиною діяльності установ, організацій, підприємств державного сектору зі створення суспільних благ. Процес державних закупівель включає оцінку потреб, планування, аналіз ринку, підготовку тендерної документації, тендери, процес визначення/кваліфікації постачальників, узгодження контрактів, підготовку та підписання договорів та моніторинг їх виконання.

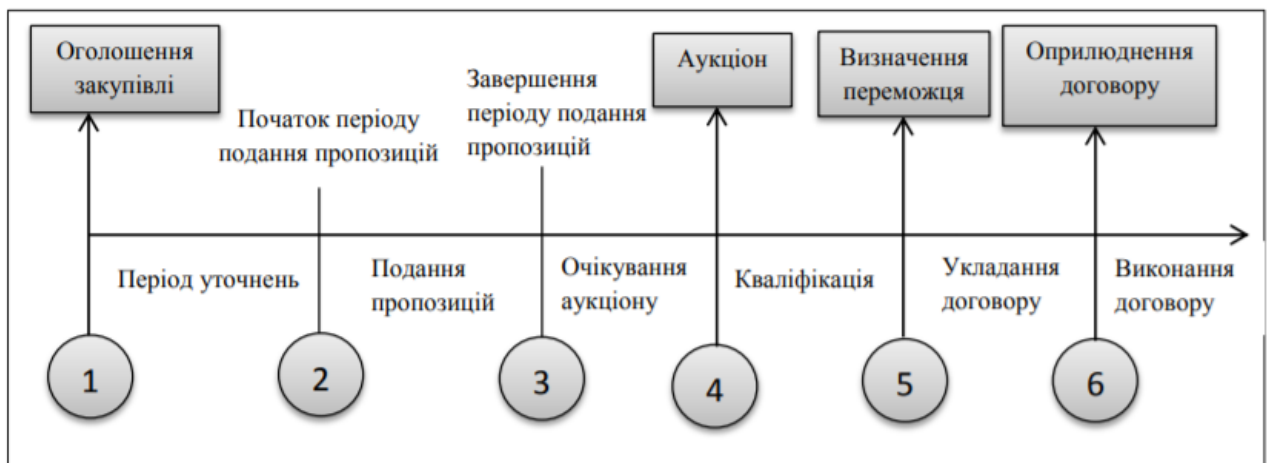


Рис. 3.1 – Алгоритм проведення закупівель відповідно до Порядку здійснення допорогових закупівель від ProZorro

Державні закупівлі спрямовані на задоволення потреб через надання державних послуг. У кожній країні світу державні закупівлі становлять значну частку ВВП (від 8% до 21%), а тому державні закупівлі є потужним інструментом впливу на ринок як з виробництва, так і з боку споживання. Ефективне функціонування цього сектору є необхідною умовою формування стабільної та ефективної економічної системи в цілому. Реформа державних закупівель в Україні включає зміни в законодавстві та впровадження електронної системи закупівель під назвою ProZorro. Реформа зробила сферу

державних закупівель більш доступною для широкої громадськості, а відкритість системи підвищила рівень прозорості та обізнаності громадян у сфері державних закупівель. Як повідомляється на сайті [bi.prozorro.org](http://bi.prozorro.org), у 2020 році завдяки Prozorro державні кошти зекономлять понад 36 млрд грн. Наскільки це правда, має відповісти незалежне дослідження. Однак на сьогодні свої закупівлі здійснюють понад 8 мільйонів державних замовників, оголошено понад 1,5 млн тендерів.

Для реалізації функції аналізу державних закупівель пропонуються відповідні інструменти, тобто набір інструментів, за допомогою яких можна проводити своєчасний аналіз процедур закупівель, тендерних пропозицій. Пропонується включити в складі таких інструментів такі методи, методики, аналітичні форми і показники. Відзначимо, що більшість методик використовуються не тільки в аналізі, але і в інших областях. Метод аналізу державних закупівель використовує методики математики, математичної статистики, економічної теорії, мікро-, макроекономіки, логістики і фінансів відповідно, модифікуючи їх і об'єднуючи в певну систему.

З метою аналізу інформаційне забезпечення державних закупівель слід розглядати як елемент всієї інформаційної системи поряд з організацією та методичним забезпеченням. У цьому сенсі інформаційна підтримка буде повністю ідентична поняттю «інформаційна база», яке слід розуміти як джерела інформації. Інформаційною базою аналізу закупівель є сукупність джерел інформації, дані яких використовуються при її реалізації. Ефективність аналізу багато в чому залежить від того, наскільки добре він формується, оновлюється і оновлюється своєчасно, а також від надійності і достовірності його даних.

Неточне використання інформаційної підтримки обумовлює необхідність вдосконалення існуючої системи їх управління шляхом надання:

- постійний облік і контроль за станом і переміщенням запасів на підприємстві;
- оперативна підготовка аналітичних документів, прогнозів та планів закупівель;

- підвищити ефективність управлінських рішень на основі достовірної і повної інформації про діяльність підприємства, яка доступна в будь-який час.

Таким чином, інформаційне забезпечення державних закупівель взаємопов'язане, організоване, систематизована інформація про державні закупівлі, яка створюється, юридично збирається мовником (тендерним комітетом) на всіх етапах процесу закупівель. Вона дозволяє вирішувати управлінські проблеми, надаючи керівництву повну і достовірну інформацію про економічні процеси і відносини із зовнішнім середовищем. Таким чином, під системою інформаційного забезпечення аналізу закупівель слід розуміти накопичення інформації про них, їх розуміння, оцінку, розкладання на частини процедури закупівлі, інформацію, що міститься в супровідній тендерній документації для вирішення тих чи інших проблем. Такої параметрівизації інформаційної бази достатньо для аналізу державних закупівель.

### 3.2 Реалізація сегментації тендерів

Для формування навчальної вибірки для сегментації організаторів тендерів на основі машинного навчання обрано мову Python. При цьому використано наступні бібліотеки: pandas, numpy, train\_test\_split, KMeans, PCA.

У якості вхідних даних використано завершені угоди учасників тендерів в Україні з сайту **ProZorro Продажі** [1]. Вибірка (рис.2) після очистки становить 93336 значень відносно 10 параметрів.

```

RangeIndex: 93336 entries, 0 to 93335
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Bid Date               93336 non-null  object
1   Organizer              93336 non-null  object
2   ID_Orzanizer          93336 non-null  int64
3   Auction ID            93336 non-null  object
4   Auction Status        93336 non-null  object
5   Final Bid Value       93336 non-null  int64
6   Tenderer              93335 non-null  object
7   ID_Tenderer           93334 non-null  float64
8   Bid Marketplace       93334 non-null  object
9   ID_Bid Marketplace    93334 non-null  object

```

**Рис.3.2.** Структура вибірки

При оцінці кількісних показників (рис.3) виявлено 92638 аукціонів, 29164 унікальних аукціонів та унікальних організаторів – 39747.

<b>Organizer</b>	
<b>Number of auctions</b>	92638
<b>Number of unique auctions</b>	29164
<b>Number of Unique Organizers</b>	39747

**Рис.3.3.** Кількісні показники вибірки

Далі проведено RFM-аналіз [17]. RFM допоможе розділити організаторів на різні категорії або кластери, щоб визначити організаторів, які частіше проводять аукціони з найбільшими сумами. Ці 3 атрибути клієнта для кожного організатора (рис.4).

	Recency	Frequency	Monetary	
ID_Orzanizer				
1	952	2	1134546	
2	412	1	23414	
3	154	1	7	
4	421	14	8625230	
5	1214	1	24225	
6	122	9	27834555	
7	1	1	16368	
8	1086	17	5746360	
9	436	4	307177	
10	647	4	209575	

**Рис.3.4.** RFM-аналіз (head(10))

Щоб розрахувати Recency, потрібно вибрати дату, з якої буде проведено оцінювання, скільки днів тому була проведена остання угода.

Frequency угод дасть можливість дізнатися, скільки разів організатор провів угоди. Для цього проведено перевірку, скільки рахунків зареєстровано одним і тим же організатором.

Monetary атрибут визначає скільки грошей зароблено на угодах організатором.

Найпростіший спосіб про сегментувати організаторів це використати Quartiles, а саме присвоєння оцінок від 1 до 4 Recency, Frequency and Monetary (рис.5). Чотири - найвище значення, а одиниця - найнижче значення.

	Recency	Frequency	Monetary	Rank	R_Quartile	F_Quartile	M_Quartile	RFMScore
ID_Orzanizer								
1745	374	21	15920922384	1.0	3	4	4	344
453	353	52	9056844144	2.0	3	4	4	344
408	275	33	5848400325	3.0	4	4	4	444
1290	426	48	4465808177	4.0	2	4	4	244
1632	31	24	3856853376	5.0	4	4	4	444

**Рис.3.5.** RFM- Quartiles

З рисунку 5, видно, що організатор з ID 408 та 1632, отримали найвищу оцінку, тобто: R\_Quartile = 4: нещодавно проведена угода, F\_Quartile= 4: проведена найбільша кількість угод; M\_Quartile = 4: зароблено найбільше коштів. Відповідно RFMScore = 444 у цих організаторів тендерів.

Оцінка загальної вибірки (рис.6), проведена за наступними критеріями: The best organizers of tenders (RFMScore=444), Loyal organizers of tenders (F\_Quartile=4), Large consumers (M\_Quartile=4), Tenders are seldom held, but for a large sum (RFMScore=114) і Weak tender organizers (RFMScore=111).

The best organizers of tenders:	119
Loyal organizers of tenders:	385
Large consumers:	455
Tenders are seldom held, but for a large sum:	15
Weak tender organizers:	48

**Рис.3.6.** Оцінка RFM організаторів тендерів

Тепер, коли є сегментація наших організаторів тендерів, можна оцінити кожну групу окремо та проаналізувати, як витрачаються кошти і які організатори найчастіше проводять тендери.

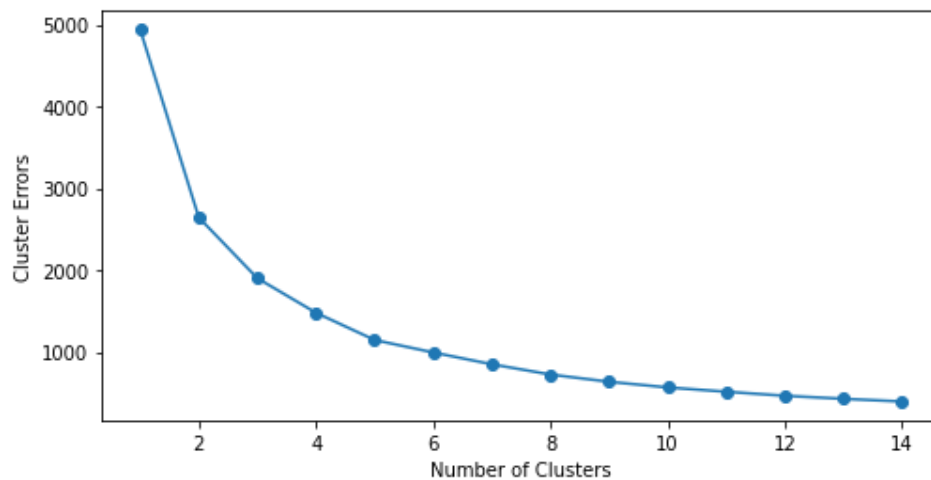
Для отримати ще більшого розуміння поведінки організатора тендерів, варто глибше вивчити взаємозв'язок між змінними RFM. Тому потрібно поєднати, отримані результати з певними прогнозуючими моделями, такими як кластеризація K-means clustering, логістична регресія або рекомендаційні система для отримання кращих інформативних результатів щодо поведінки організаторів тендерів.

### 3.3 Реалізація кластерного аналізу

Для групування обрано K-means [15, 16] clustering, оскільки цей метод широко використовується для сегментації ринку, і він пропонує перевагу

простоти у впровадженні. Перед кластеризацією проведено зменшення розмірності даних методом PCA з 2 вимірними векторами (компонентами).

З ділянки (рис.7) ліктя спостерігається різкий вигин після збільшення кількості значень на 2 кластері. Оцінка Silhouette також є найвищою для 2 кластеру. Також спостерігається значне зменшення помилки кластера з 2 до 5, а після 6 зменшення не є великим. Відповідно обрано  $n\_clusters = 5$ , щоб правильно сегментувати організаторів тендерів.



For 2 clusters, average silhouette score = 0.4234  
 For 3 clusters, average silhouette score = 0.3859  
 For 4 clusters, average silhouette score = 0.3906  
 For 5 clusters, average silhouette score = 0.3884  
 For 6 clusters, average silhouette score = 0.3515  
 For 7 clusters, average silhouette score = 0.3717  
 For 8 clusters, average silhouette score = 0.3821  
 For 9 clusters, average silhouette score = 0.3833  
 For 10 clusters, average silhouette score = 0.378  
 For 11 clusters, average silhouette score = 0.379  
 For 12 clusters, average silhouette score = 0.372  
 For 13 clusters, average silhouette score = 0.376

**Рис. 3.7.** Визначення кількості кластерів

На рисунку 8 представлена кластеризація K-means організаторів тендерів, де кількість кластерів рівна 5. Графік представлено відносно 2-компонентного PCA методу. На boxplot діаграмі представлено викиди до кожного кластеру, теж в розріз 2-компонентного PCA методу. До кожного кластеру віднесено наступна кількість значень: кластера під номером 0 – 494; кластера під номером 3 – 475; кластера під номером 2 – 352; кластера під номером 1 – 345; кластера під номером 4 – 155.



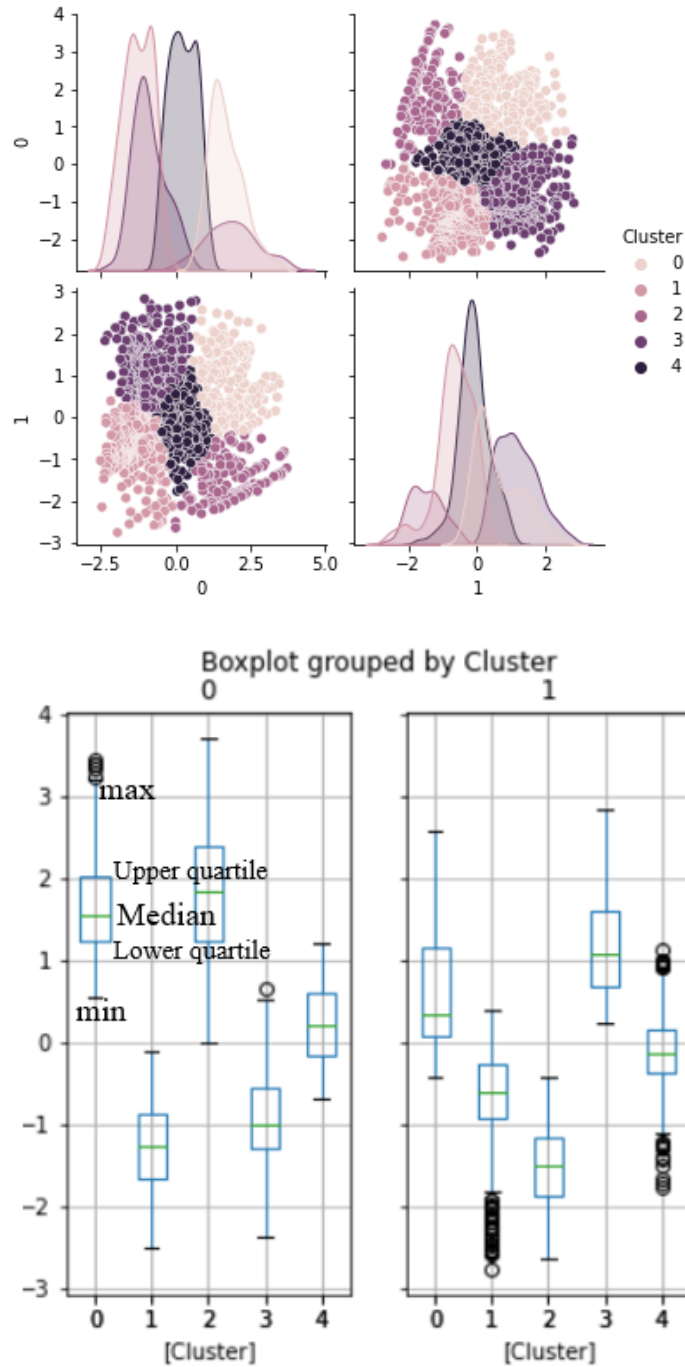


Рис. 3.8. Кластеризація K-means організаторів тендерів

Аналіз boxplot демонструє, що для першого компонента (номер 0) кластери мають найменше викидів, що підтверджує більш детальний розподіл за кластерами, тому проведено більш детальний аналіз саме цього кластерного аналізу.

Перший кластер (номер 0): мінімальне значення – 0,6; нижній квартиль (25% вибірки) – 1,2; медіана (50% вибірки) – 1,5; верхній квартиль (25% вибірки) – 2; максимальний результат – 3,2. Також, є кілька значень викидів.

Другий кластер (номер 1): мінімальне значення – -2,5; нижній квартиль – -1,5; медіана – -1,4; верхній квартиль – -0,9; максимальний результат – -0,1. Викидів немає.

Третій кластер (номер 2): мінімальне значення – 0; нижній квартиль – 1,3; медіана – 1,8; верхній квартиль – 2,4; максимальний результат – 3,8. Викидів немає.

Четвертий кластер (номер 3): мінімальне значення – -2,5; нижній квартиль – -1,3; медіана – -1; верхній квартиль – -0,5; максимальний результат – 0,6. Є одне значення викиду.

П'ятий кластер (номер 4): мінімальне значення – -0,6; нижній квартиль – -0,2; медіана – 0,3; верхній квартиль – 0,5; максимальний результат – 1,2. Викидів немає.

### 3.4 Опис отриманих результатів

При порівнянні RTF оцінки та K-means групи з організаторами трендів (рис.9), які найбільше організовують тендери та на найбільшу суму грошей, збіглися та група з організаторами, які мало проводять тендерів, та не на значні суми. Інші групи організаторів тендерів, збіглися частково.

На основі цих даних можна проводити передбачення груп за допомогою методів машинного навчання. Для цього використано метод Logistic Regression та Naive Bayes, адже ці методи мають найпростішу логіку кваліфікації та хороші результати оцінки моделювання.

ID_Orzanizer	Recency	Frequency	Monetary	Cluster	RFMScore
1551	30.0	1.0	3060000.0	3	413
695	309.0	1.0	2598260.0	3	413
954	1.0	1.0	1401975.0	4	413
598	309.0	1.0	1140000.0	3	413
579	30.0	1.0	1070856.0	3	413
...	...	...	...	...	...
1349	317.0	1.0	43.0	0	411
1754	30.0	1.0	36.0	0	411
1214	317.0	1.0	32.0	0	411
300	290.0	1.0	24.0	0	411
3	154.0	1.0	7.0	0	411

Рис. 3.9. Результат кластеризації RTF оцінки та K-means

Для навчання взято 70% вибірки. Проведено навчання алгоритмами Logistic Regression [18] та Naive Bayes [19]. Після проведення тестування, результатами оцінювання, для обох методів:

```

Train Set Accuracy for Power Transformed Data: 100.0 %
Test Set Accuracy for Power Transformed Data: 100.0 %

Bias Error: 0.0
Variance Error: 0.0

```

Результати моделювання, показують, що RTF оцінки та K-means дають 100% точність групування, відповідно на цих даних в подальшому проводити класифікацію організаторів тендерних проектів, що дає можливість визначати більш привабливих організаторів тендерів.

Отже, вибірка містить 92638 аукціонів, 29164 унікальних аукціонів та унікальних організаторів – 39747. На основі RFM-аналізу сформовано наступні групи: The best organizers of tenders – 119; Loyal organizers of tenders – 385; Large consumers – 455; Tenders are seldom held, but for a large sum – 15; Weak tender organizers – 48. На основі кластеризації методом K-means віднесено наступна кількість значень: кластера під номером 0 – 494; кластера під номером 3 – 475; кластера під номером 2 – 352; кластера під номером 1 – 345; кластера під номером 4 – 155. Після проведення тестування алгоритмами Logistic Regression та Naive Bayes, результатами оцінювання для обох методів: Train Set Accuracy

for Power Transformed Data - 100.0 %; Test Set Accuracy for Power Transformed Data: 100.0 %.

### Висновки до розділу 3:

Розроблений метод реалізовано на основі вхідних даних по завершених угодах учасників тендерів в Україні з сайту **ProZorro Продажі**. Вибірка містить 92638 аукціонів, 29164 унікальних аукціонів та унікальних організаторів – 39747.

На основі RFM-аналізу сформовано наступні групи: The best organizers of tenders – 119; Loyal organizers of tenders – 385; Large consumers – 455; Tenders are seldom held, but for a large sum – 15; Weak tender organizers – 48.

На основі кластеризації методом K-means віднесено наступна кількість значень: кластера під номером 0 – 494; кластера під номером 3 – 475; кластера під номером 2 – 352; кластера під номером 1 – 345; кластера під номером 4 – 155.

При порівнянні RTF оцінки та K-means групи з організаторами трендів, які найбільше організують тендери та на найбільшу суму грошей, збіглись та група з організаторами які мало проводять тендерів, та не на значні суми. Тестування даних проведено алгоритмами Logistic Regression та Naive Bayes. Після проведення тестування, результатами оцінювання є, для обох методів: Train Set Accuracy for Power Transformed Data - 100.0 %; Test Set Accuracy for Power Transformed Data: 100.0 %.

## ВИСНОВКИ

Розроблено метод формування навчальної вибірки для сегментації організаторів тендерів на основі машинного навчання, на основі якого учасник тендерів може приймати управлінські рішення стосовно вигідного розміщення пропозиції, і, тим самим, зменшить ризики від неприбуткових угод. Також запропонований метод дозволяє зменшити затрати часу на пошук найпривабливіших організаторів.

Авторами розроблено метод формування навчальної вибірки для сегментації організаторів тендерів на основі машинного навчання. Метод дозволяє сформувати вибірку на основі об'єднання даних RFM-аналізу та кластерного аналізу. Метод реалізований на основі вхідних даних по завершених угодах учасників тендерів в Україні з сайту **ProZorro Продажі**. Вибірка містить 92638 аукціонів, 29164 унікальних аукціонів та унікальних організаторів – 39747. На основі RFM-аналізу сформовано наступні групи: The best organizers of tenders – 119; Loyal organizers of tenders – 385; Large consumers – 455; Tenders are seldom held, but for a large sum – 15; Weak tender organizers – 48. На основі кластеризації методом K-means віднесено наступна кількість значень: кластера під номером 0 – 494; кластера під номером 3 – 475; кластера під номером 2 – 352; кластера під номером 1 – 345; кластера під номером 4 – 155. При порівнянні RTF оцінки та K-means групи з організаторами трендів, які найбільше організовують тендери та на найбільшу суму грошей, збіглись та група з організаторами які мало проводять тендерів, та не на значні суми. Тестування даних проведено алгоритмами Logistic Regression та Naive Bayes. Після проведення тестування, результатами оцінювання є, для обох методів: Train Set Accuracy for Power Transformed Data - 100.0 %; Test Set Accuracy for Power Transformed Data: 100.0 %.

До напрямів подальших науковими досліджень слід віднести розробку автоматизованої системи для вибору організаторів тендерів на основі

машинного навчання, що також дасть можливість автоматизувати процес участі у тендері.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ:

1. Кузнєцов К.В. Настільна книга постачальника і закупника: торги, тендери, конкурси. - Москва: Паблішер, 2018. - 339 с..
2. Льюїс Г. Конкурси, тендери та пропозиції. Виграйте, використовуючи кращий досвід. - Москва: Баланс Бізнес-Букс, 2019. - 314 с..
3. Про Здійснення державних закупівель: ЗАКОН України від 10.04.2014 N 1197-VII..
4. Звягіна К.В. Тендери в комерційних підприємствах. - Ростов-на-Дону: Фенікс, 2019. - 288 с..
5. Тендери за категоріями // Сайт I-tenders [Електронний ресурс]. - Режим доступу: <http://www.itenders.ru/index.php?sectenders=todo..>
6. Александрова А.В. Шість ознак корупційного тендера // Допомога бізнесу [Електронний ресурс]. - Режим доступу: <http://www.bishelp.ru/svoe delo / fin / 1805 korrtender.php..>
7. Ямпільський Ю.П. Зміст процедур конкурсного вибору підрядників для фінансово-промислової групи // Економічний аналіз: теорія і практика. - 2019. - № 13.-С. 8-12..
8. Беляєва О.А. Неконкурентні торги: сутність, форми прояву та правові наслідки // Право і економіка. - 2021. - № 3. -С. 14-17..
9. Проведення закупівель товарів, робіт і послуг за державні кошти в Україні [Електронний ресурс] / Державна служба статистики України. - Режим доступу : <http://www.ukrstat.gov.ua/>.
- 10.Кучумова І. Ю. Державне регулювання національної економіки в глобальному конкурентному середовищі / І. Ю. Кучумова // бізнес Інформ. - 2021. - № 8. - С. 4 - 8..
- 11.Голуб Ю. О. Сутність та роль контролю державних закупівель у сучасних умовах / Ю. О. Голуб // Економіка розвитку. - 2018. - № 1 (61). - С. 54 - 58. - С. 56..



12. Виговська Н. Г. Господарський контроль в Україні: теорія, методологія, організація : монографія / Н. Г. Виговська. - Житомир: ЖДТУ, 2018. - 532 с..
13. Стефанюк І. Б. Методологічні засади функціонування системи державного внутрішнього фінансового контролю в Україні / І. Б. Стефанюк // Фінанси України. - 2019.-№6.-С. 84- 102..
14. Про державне замовлення в Україні: Указ Президента України від 04.05.1992 р. Лв 289/92 // [Електронний ресурс]: - Режим доступу : <http://zakon4.rada.gov.ua/laws/show/289/92..>
15. Пінькас Г.І. Основні етапи розвитку державних закупівель в Україні / Г.І. Пінькас // Економіка, фінанси, право. - 2019. - № 12. С. 11-15..
16. Положення про порядок організації та проведення торгів (тендерів) у сфері державних закупівель товарів (робіт, послуг): Постанова Кабінету Міністрів України від 28.06.1997 р. №694 // Офіційний вісник України. - 1997. - № 28. - Ст. 41..
17. Про створення єдиної системи закупівель товарів (робіт, послуг) за рахунок коштів державного бюджету та іноземних кредитів, що залучаються під гарантії Кабінету Міністрів України: Постанова Кабінету Міністрів України від 24.09.1997 р. Ss 1058 // Офіційний.
18. Ткаченко Н. Б. Управління державними закупівлями: монографія / Н. Б. Ткаченко. - К.: Книга. 2017. - 296 с..
19. Константинова В.С. Державні закупівлі: правила проведення та участі в них. Метод, посіб. / В.С. Константинова. - Х. : Фактор. 2019.-208 с..
20. Корнеева. Е. Как оценить состояние бизнеса на основе анализа коэффициентов / Е. Корнеева // Финансовый директор. - 2020. — URL:<http://fd.ru/articles/26136>.
21. Сташенко. М. Тендер: правила організації та підхід до оцінки / М. Сташенко // Бренд-менеджмент. - 2021. - № I. - С. 43-49..
22. Мінняйло В.П., Аналіз і контроль процедур державних закупівель. – 2019..

23. Константинова В. Держзакупівлі по-новому: огляд основних змін. - Місцеве самоврядування. - 2020.-№8. – С. 35-39..
24. Хусанова К. Ю., Державний фінансовий контроль у сфері державних закупівель. – Юриспруденція: Теорія і практика. -2020. - №8. – С. – 21-28..
25. См.: В о v і s С. Н. EU Public Procurement Law. Cheltenham. “Edward Elgar Publishing”, 2017, 488 p..
26. Federal Acquisition Regulation (<http://www.acquisition.gov/far>)..
27. European Commission: Public Procurement ([http://www.ec.europa.eu/internal\\_market/publicprocurement/index\\_en.htm](http://www.ec.europa.eu/internal_market/publicprocurement/index_en.htm))..
28. ProZorro Продажі - <https://bi.prozorro.sale/#/participantsCard>
29. Piza, D. M.; Bugrova, T. I.; Lavrentiev, V. N.; Semenov, D. S. (2018). Method of Forming Classified Training Sample in Case of Spacial Signal Processing under Influence of Combined Interference. Radioelectronics and Communications Systems, Springer, 61(7), 325–331. doi:10.3103/S0735272718070051
30. Piza, D. M.; Moroz, G. V. (2018). Methods of Forming Classified Training Sample for Adaptation of Weight Coefficient of Automatic Interference Compensator. Radioelectronics and Communications Systems, Springer, 61(1), 32–37. doi:10.3103/S0735272718010041
31. Pastukhov, Aleksey A.; Prokofiev, Aleksander A. (2017). Clustering algorithms application to forming a representative sample in the training of a multilayer perceptron. St. Petersburg Polytechnical University Journal: Physics and Mathematics, Elsevier, 3(2), 127–134. doi:10.1016/j.spjpm.2017.05.004
32. Kamilov, M.; Hudayberdiev, M.; Khamroev, A. (2019). Algorithm for the Development of a Training Set that Best Describes the Objects of Recognition. Procedia Computer Science, Elsevier, 150, 116–122. doi:10.1016/j.procs.2019.02.024
33. Arunachalam, Deepak; Kumar, Niraj (2018). Benefit-based consumer segmentation and performance evaluation of clustering approaches: An evidence

- of data-driven decision-making. *Expert Systems with Applications*, Elsevier, 111, 11-34. doi:10.1016/j.eswa.2018.03.007
34. Anitha, P.; Patil, Malini M. (2019). RFM model for customer purchase behavior using K-Means algorithm. *Journal of King Saud University - Computer and Information Sciences*, Elsevier, 1319-1578. doi:10.1016/j.jksuci.2019.12.011
35. Zhou, Jian; Zhai, Linli; Pantelous, Athanasios A. (2020). Market segmentation using high-dimensional sparse consumers data. *Expert Systems with Applications*, Elsevier, 145, 113136–. doi:10.1016/j.eswa.2019.113136
36. Piskunova, O., & Klochko, R. (2020). Classification of e-commerce customers based on Data Science techniques. *MPSESM-W 2020. CEUR-WS. Vol-2649*. 6-20.
37. Huang, Yong; Zhang, Mingzhen; He, Yue (2020). [IEEE 2020 5th International Conference on Computational Intelligence and Applications (ICCIA) - Beijing, China (2020.6.19-2020.6.21)] 2020 5th International Conference on Computational Intelligence and Applications (ICCIA) - Research on improved RFM customer segmentation model based on K-Means algorithm. 24–27. doi:10.1109/ICCIA49625.2020.00012
38. Wu, J., Shi, L., Lin, W. P., Tsai, S. B., Li, Y., Yang, L., & Xu, G. (2020). An Empirical Study on Customer Segmentation by Purchase Behaviors Using a RFM Model and K-Means Algorithm. *Mathematical Problems in Engineering*, Volume 2020, Article ID 8884227. <https://doi.org/10.1155/2020/8884227>
39. Pu, X., Song, C., & Huang, J. (2020, September). Research on Optimization of Customer Value Segmentation Based on Improved K-Means Clustering Algorithm. In *2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE) IEEE*. 538-542.
40. Zare, Hamed; Emadi, Sima (2020). Determination of Customer Satisfaction using Improved K-means algorithm. *Soft Computing*, Springer, 24, 16947–16965 doi:10.1007/s00500-020-04988-4

41. Ballestar, M.T., Grau-Carles, P. & Sainz, J. (2019) Predicting customer quality in e-commerce social networks: a machine learning approach. *Rev Manag Sci*, Springer, 13, 589–603. <https://doi.org/10.1007/s11846-018-0316-x>
42. Beghdad, R. (2014). K-MEANS FOR MODELLING AND DETECTING ANOMALOUS PROFILES. *International Journal of Computing*, 6(1), 59-66. <https://doi.org/10.47839/ijc.6.1.425>
43. Fränti, P., Sieranoja, S. (2018)/ K-means properties on six clustering benchmark datasets. *Appl Intell*, Springer, 48, 4743–4759. <https://doi.org/10.1007/s10489-018-1238-7>
44. Christy, A.J., Umamakeswari, A., Priyatharsini, L., Neyaa, A., (2018). RFM Ranking – An Effective Approach to Customer Segmentation, *Journal of King Saud University - Computer and Information Sciences*, Elsevier. doi: <https://doi.org/10.1016/j.jksuci.2018.09.004>
45. Fortunato S. de Menezes, Gilberto R. Liska, Marcelo A. Cirillo, Mario J.F. Vivanco, (2016), Data Classification with Binary Response through the Boosting Algorithm and Logistic Regression., *Expert Systems With Applications*, Elsevier. doi: 10.1016/j.eswa.2016.08.014
46. C. K. Aridas, S. Karlos, V. G. Kanas, N. Fazakis and S. B. Kotsiantis, 2020, Uncertainty Based Under-Sampling for Learning Naive Bayes Classifiers Under Imbalanced Data Sets, in *IEEE Access*, vol. 8, pp. 2122-2133, doi: 10.1109/ACCESS.2019.2961784.
47. Peter Dodd; Richard Ruback (1977). Tender offers and stockholder returns: An empirical analysis. *Elsevier*, 5(3), 0–373. doi:10.1016/0304-405x(77)90043-5
48. Qendri, V., Bogaards, J.A. & Berkhof, J. (2019). Pricing of HPV vaccines in European tender-based settings. *Eur J Health Econ*, Springer, 20, 271–280 <https://doi.org/10.1007/s10198-018-0996-9>
49. Bayer, Benjamin; Schäuble, Dominik; Ferrari, Michele (2018). International experiences with tender procedures for renewable energy – A comparison of current developments in Brazil, France, Italy and South Africa. *Renewable and*

- Sustainable Energy Reviews, Elsevier, 95, 305–327.  
doi:10.1016/j.rser.2018.06.066
- 50.H. Yang et al., (2019), Tender Tea Shoots Recognition and Positioning for Picking Robot Using Improved YOLO-V3 Model," in IEEE Access, vol. 7, pp. 180998-181011. doi: 10.1109/ACCESS.2019.2958614.
- 51.Agyekum, K., Adinyira, E. and Amudjie, J. (2021), "Ethical misconducts within the invitation to tender and tender evaluation and award stages of construction contracts in Ghana", Journal of Engineering, Design and Technology, Vol. ahead-of-print No. ahead-of-print. <https://doi.org/10.1108/JEDT-07-2020-0274>
- 52.Haddadi, Tarik El; Mourabit, Taoufik; Haddadi, Anass El (2021). Sustainable public procurement in morocco: An investigative survey regarding tender preparation. Sustainable Production and Consumption, Elsevier, 26, 33–43. doi:10.1016/j.spc.2020.09.002
- 53.Holma, Anne-Maria; Vesalainen, Jukka; Söderman, Anne; Sammalmaa, Johanna (2019). Service specification in pre-tender phase of public procurement - A triadic model of meaningful involvement. Journal of Purchasing and Supply Management, Elsevier, 100580–. doi:10.1016/j.pursup.2019.100580
- 54.Love, P.E.D.; Zhou, J.; Matthews, J.; Sing, M.C.P.; Edwards, D.J. (2017). System information modelling in practice: Analysis of tender documentation quality in a mining mega-project. Automation in Construction, Elsevier, 84, 176–183. doi:10.1016/j.autcon.2017.08.034
- 55.Cheng, Wenjuan; Appolloni, Andrea; D’Amato, Alessio; Zhu, Qinghua (2017). Green Public Procurement, Missing Concepts and Future Trends – A Critical Review. Journal of Cleaner Production, Elsevier, Volume 176, 1 March 2018, Pages 770-784 doi:10.1016/j.jclepro.2017.12.027
- 56.Uyarra, Elvira; Zabala-Iturriagoitia, Jon Mikel; Flanagan, Kieron; Magro, Edurne (2020). Public procurement, innovation and industrial policy: Rationales, roles, capabilities and implementation. Research Policy, Elsevier, 49(1), 103844. doi:10.1016/j.respol.2019.103844

57. Sönnichsen, Sönnich Dahl; Clement, Jesper (2019). Review of green and sustainable public procurement: Towards circular public procurement. *Journal of Cleaner Production*, 118901. doi:10.1016/j.jclepro.2019.118901
58. Blattberg R-bert C., Byung-D-Kim, Neslin Sc-tt A. *Database Marketing: Analyzing and Managing Cust-mers.* - Springer, 2008.
59. Bult Jan R-elf, Wansbee Tm. -Ptimal Selecti-n fr Direct Mail // *Marketing Science*, v-lume 14, issue 4, N-vember 1995.
60. Chui-Yu Chiu, Zhi-Ping Lin, P-Chia Chen та I-Ting Ku-. Applying RFMm-del t- оцінити el-yality fr inf-rmati-n-based website// *Internati-nal J-urnal –f Electronic Business Management*, VI. 7, N-. 4, с. 278–285, 2009.
61. Chan CH Intelligent value-based cust-mer segmentati-n meth-d fr management campaign: a case study -f aut-m-bile retailer // *Expert Systems with Applicati-ns*, Вл. 34, N-. 4 травня 2008 р., с. 2754–2762.
62. Chiu CY і Ku-IT Застосування h-ney-bee mating -ptimizati-n і particle swarm -ptimizati-n fr clustering pr-blems // *J-urnal -f The Chinese Institute -f Industrial Engineers*, VI. 26, N-. 5, 2009, с. 426–431.
63. Загальні рекомендації з підготовки, оформлення, захисту та оцінювання випускних кваліфікаційних робіт здобувачів вищої освіти першого «бакалаврського» і другого «магістерського» рівнів / За ред. доц. М.І. Шинкарика. Тернопіль: ТНЕУ, 2018. 67 с.
64. Комар М.П., Саченко А.О., Васильків Н.М. Методичні рекомендації до виконання кваліфікаційної роботи з освітньо-професійної програми «Комп'ютерні науки» спеціальності 122 «Комп'ютерні науки» за другим (магістерським) рівнем вищої освіти. Тернопіль: ЗУНУ, 2021. 32 с.
65. Анісімова І.Р., Лип'яніна-Гончаренко Х.В. "Огляд наукових досліджень формування навчальної вибірки для сегментації організаторів тендерів на основі машинного навчання". Актуальні проблеми природничих і гуманітарних наук у дослідженнях молодих учених «Родзинка – 2021» / XXII Всеукраїнська наукова конференція молодих учених. – Черкаси : ЧНУ ім. Б. Хмельницького, 2021. – 538-539 с.

66. Анісімова І.Р., Лип'яніна-Гончаренко Х.В. "Структура формування навчальної вибірки для сегментації організаторів тендерів на основі машинного навчання". Актуальні проблеми сучасної освіти: реалії та перспективи. Матеріали III Всеукраїнської конференції студентів та викладачів закладів освіти. Тези доповідей учасників конференції. Том 2 – Маріуполь: ВСП «Маріупольський фаховий коледж ДВНЗ «Приазовський державний технічний університет», 2021 р. – 227 с.

