

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ ТЕЛЕКОМУНІКАЦІЙ**  
**НАВЧАЛЬНО–НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**  
Кафедра інженерії програмного забезпечення

**Пояснювальна записка**

до магістерської роботи  
на ступінь вищої освіти магістр

на тему: «**Удосконалення інформаційної системи для автоматизації  
аналітичної діяльності компанії Velo Kyiv із застосуванням машинного  
навчання**»

Виконав: студент 6 курсу, групи ПДМ–61  
спеціальності

121 Інженерія програмного забезпечення  
(шифр і назва спеціальності/спеціалізації)

Максименко Р.С.

(прізвище та ініціали)

Керівник Золотухіна О.А.

(прізвище та ініціали)

Рецензент \_\_\_\_\_

(прізвище та ініціали)

Київ –2022

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ ТЕЛЕКОМУНІКАЦІЙ**  
**НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ**  
**ТЕХНОЛОГІЙ**

Кафедра Інженерії програмного забезпечення  
Ступінь вищої освіти -«Магістр»  
Спеціальність підготовки – 121 «Інженерія програмного забезпечення»

**ЗАТВЕРДЖУЮ**

Завідувач кафедри  
Інженерії програмного  
забезпечення

Негоденко О.В.

“ \_\_\_\_\_ ” \_\_\_\_\_ 2022 року

**З А В Д А Н Н Я**  
**НА МАГІСТЕРСЬКУ РОБОТУ СТУДЕНТА**

**Максименко Руслан Сергійович**

(прізвище, ім'я, по батькові)

1. Тема роботи: «Удосконалення інформаційної системи для автоматизації аналітичної діяльності компанії VeloKyiv із застосуванням машинного навчання»

Керівник роботи: Золотухіна О.А., кандидат технічних наук, доцент

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом закладу вищої освіти від «11» жовтня 2021 року  
№170.

2. Строк подання студентом роботи “28” грудня 2021 року

3. Вхідні дані до роботи

Методи машинного навчання;

Науково-технічна література з удосконалення інформаційної системи для автоматизації аналітичної діяльності із застосуванням машинного навчання;

4. Зміст розрахунково-пояснювальної записки(перелік питань, які потрібно розробити).

4.1 Аналіз можливостей машинного навчання під час автоматизації

4.2 Автоматизація аналітичної діяльності компанії VeloKyiv.

4.3 Методологія впровадження методів машинного навчання для автоматизації аналітичної діяльності на ринку велосипедів та конверсія.

5. Перелік демонстраційного матеріалу (назва основних слайдів)

1. Актуальність проблеми
2. Кластеризація
3. Пошук асоціативних правил
4. Критеріальна база клієнта
5. Процес формування індивідуальних пропозицій для клієнтів
6. Приклад згенерованого повідомлення
7. Аналіз ефективності комерційної пропозиції
8. Порівняльний аналіз конверсії продажів

6. Дата видачі завдання «    » 2021

### КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів Магістерської роботи	Строк виконання етапів роботи	Примітка
1	Підбір науково-технічної літератури		Виконано
2	Вимоги до удосконалення інформаційної системи	08.11-10.11	Виконано
3	Побудування повної моделі компанії.	11.11-18.11	Виконано
4	Удосконалення обробки даних при формуванні цільової аудиторії компанії	19.11-28.11	Виконано
5	Удосконалення аналізу ефективності комерційної пропозицій	29.11-04.12	Виконано
6	Вступ, висновки, реферат	05.12-12.12	Виконано
7	Розробка обов'язкових демонстраційних матеріалів	20.12	Виконано
8	Попередній захист роботи	22.12	Виконано
9	Здача роботи	28.12	Виконано

Студент \_\_\_\_\_  
( підпис ) (прізвище та ініціали)

Керівник роботи \_\_\_\_\_  
( підпис ) (прізвище та ініціали)





## РЕФЕРАТ

### АВТОМАТИЗАЦІЯ АНАЛІТИЧНОЇ ДІЯЛЬНОСТІ, МАШИННЕ НАВЧАННЯ, КЛАСТЕРИЗАЦІЯ, ПОШУК АСОЦІАТИВНИХ ПРАВИЛ

Магістерська робота виконана на 70 сторінках, містить: 36 рисунків, 5 таблиць, 3 формули та 9 використаних джерел.

*Об'єкт дослідження* – Процес автоматизації аналітичної діяльності велосипедного інтернет-магазину.

*Предмет дослідження* – Методи машинного навчання для автоматизації аналітичної діяльності інтернет-магазину.

*Мета роботи* – є підвищення якості автоматизованої аналітичної діяльності інтернет-магазину VELOKYIV за рахунок удосконалення інформаційної системи на основі методів машинного навчання.

*Методи дослідження* – методи машинного навчання.

У роботі проведено аналіз можливостей машинного навчання при автоматизації аналітичних даних.

Для автоматизації аналітичної діяльності інтернет-магазину VeloKyiv потрібно вирішити наступні завдання:

- вивчити діяльність організації та побудувати її повну модель
- удосконалити обробку даних при формуванні цільової аудиторії компанії;
- удосконалити аналіз ефективності комерційної пропозицій.
- Оцінити конкурентоспроможність підприємства
- Підвищити конверсію продажів

Для більшої конверсії необхідно впровадити систему персонального підбору пропозицій клієнтам за допомогою методів машинного навчання.

У якості вихідних даних є індивідуально пропозиція для клієнта, та аналіз ефективності цієї пропозиції.

*Галузь використання* – автоматизація аналітичної діяльності

## ЗМІСТ

<b>ВСТУП .....</b>	<b>3</b>
<b>1 АНАЛІЗ ІСНУЮЧИХ РІШЕНЬ ТА АНАЛІЗ МОЖЛИВОСТЕЙ МАШИННОГО НАВЧАННЯ ПРИ АВТОМАТИЗАЦІ.....</b>	<b>5</b>
1.1 Аналіз існуючих рішень.....	5
1.2 Джерела інформації для аналітичної діяльності .....	9
1.3 Методи машинного навчання .....	16
1.3.1 Нейронні мережі .....	16
1.3.2 Дерево рішень .....	19
1.3.3 «Випадковий ліс» .....	20
1.3.4 Кластеризація.....	20
1.3.5 Пошук асоціативних правил.....	21
1.3.6 Попередня обробка даних .....	21
1.4 Аналіз джерел інформації і можливостей їх використання для автоматизованої обробки даних .....	23
<b>2 АВТОМАТИЗАЦІЯ АНАЛІТИЧНОЇ ДІЯЛЬНОСТІ КОМПАНІЇ VELOKYIV .....</b>	<b>33</b>
2.1 Загальна характеристика підприємства .....	33
2.2 Автоматизація бізнес-процесів обробки даних при формуванні цільової аудиторії компанії .....	42
2.3 Застосування методів машинного навчання для обробки інформації у компанії VELOKYIV.....	44
2.4 Оцінка конкурентоспроможності підприємства за допомогою методів машинного навчання.....	54
<b>3 МЕТОДОЛОГІЯ ВПРОВАДЖЕННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ АВТОМАТИЗАЦІЇ АНАЛІТИЧНОЇ ДІЯЛЬНОСТІ НА РИНКУ ВЕЛОСИПЕДНОГО РІТЕЙЛЮ ТА КОНВЕРСІЯ .....</b>	<b>65</b>
3.1 Розробка методики впровадження методів машинного навчання для автоматизації аналітичної діяльності.....	65
3.2 Конверсія продажів.....	68

<b>ВИСНОВОК .....</b>	<b>70</b>
<b>СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ .....</b>	<b>71</b>



## ВСТУП

*Обґрунтування вибору теми та її актуальність:* У зв'язку з цифровізацією суспільства, для збереження своєї конкурентоспроможності на ринку велосипедного ринку, необхідно аналізувати бізнес-процеси, усувати їхні недоліки, шукати шляхи оптимізації, зараз це необхідно у підприємств будь-якої галузі. На даний час, для вилучення даних часто використовується робота звичайних операторів ПК, що займає багато часу. Через людський фактор, не виключається і наявність помилок у вилучених даних.

Існуючі системи автоматизації аналітичних даних дозволяють спростити роботу операторів, прискорюють час обробки документів, зменшують кількість помилок за рахунок відсутності людського фактору, проте не виключають їх повністю, через що присутність оператора все ще важлива.

Для пришвидшення введення даних та їх обробки необхідно удосконалити інформаційну систему, яка автоматично буде вилучати необхідні дані, які були отримані шляхом відстежуванням активності користувача.

*Ступінь вивчення проблеми:* На даний момент існує досить багато продуктів, які займаються автоматизацією аналітичної діяльності. Але, майже усі вони не використовують машинне навчання, якщо його залучити, то завдяки отриманим даним можливо отримати набагато більше інформації про клієнта.

*Об'єктом дослідження* є процес автоматизації аналітичної діяльності велосипедного інтернет-магазину.

*Предметом дослідження* є методи машинного навчання для автоматизації аналітичної діяльності інтернет-магазину

*Метою роботи* є підвищення якості автоматизованої аналітичної діяльності інтернет-магазину VELO KYIV за рахунок удосконалення інформаційної системи на основі методів машинного навчання .

*Завданням роботи* є удосконалення інформаційної системи для автоматизації аналітичної діяльності компанії із застосуванням машинного навчання.

*Методика дослідження:* Перш за все, потрібно було визначити правильну методику розпізнавання та класифікації інформації, вибрати правильну архітектуру та стек технологій із забезпеченням коректної працездатності продукту, за можливості підтримки його розробником з метою ефективного додавання нового функціоналу або виправлення виявлених недоліків.

Враховуючи вимоги до специфіки програмного продукту, найкращим рішенням є додаток, що має клієнт-серверну архітектуру, де клієнтська сторона відповідає за роботу та логіку програмного продукту, а серверна – за обчислення. Такий тип технології дає можливість використання складних обчислень, не призводячи до необхідності введення великої кількості даних.

Для клієнтської сторони, інформаційна система має включати в себе бібліотеку вже вилучених даних, та сторінку для обробки документів. Сторінка для обробки інформації представлятиме з себе форму для введення/редагування даних та аналітику цих даних.

Результати обробки інформації можна переглядати всередині інформаційної системи.

*Наукова новизна роботи:* Таким чином, наукова новизна полягає в створенні персональних пропозиції з інформації отриманої з активності користувача, що дозволить ретельніше підбирати товар.

*Практична значущість результатів:* Даний продукт може бути використаний у всіх сферах діяльності, де потрібна створити персональну пропозицію та проаналізувати ефективність пропозиції

# 1 АНАЛІЗ ІСНУЮЧИХ РІШЕНЬ ТА АНАЛІЗ МОЖЛИВОСТЕЙ МАШИННОГО НАВЧАННЯ ПРИ АВТОМАТИЗАЦІЇ

## 1.1 Аналіз існуючих рішень

Створення оптимальної маркетингової стратегії починається з аналізу даних.

У маркетингу аналіз даних - це процес організації, пояснення та інтерпретації даних, щоб відповісти на питання про маркетингову стратегію, помножити її переваги та зміцнити її слабкі місця. Наприклад, за допомогою аналізу, ви зможете знайти відповіді на такі запитання:

- Скільки унікальних відвідувачів у вашого сайту в середньому за день, тиждень і місяць?
- Скільки із них нових?
- Які сторінки вони приходять?
- Який із ваших рекламних каналів найефективніший?
- На яких каналах вам слід зосередитись, щоб збільшити дохід?

Маркетологам потрібен аналіз даних. З його допомогою вони можуть уникнути припущень і побачити, які області їхньої маркетингової стратегії вимагають особливої уваги. Аналіз даних можна використовуватиме виявлення тенденцій, закономірностей та цінної інформації, щоб спростити робочий процес, коригувати рекламні кампанії та підвищувати дохід.

Правильно аналізуючи дані, ви зможете досягти таких маркетингових цілей:

*Зменшити витрати на рекламу.* Знайдіть неефективні кампанії, ключові слова, регіони, посадкові або товари, які погано конвертують, припиніть вкладати в них кошти та зосередьтеся на каналах, які приносять більше доходу.

*Збільшити прибуток від відвідувачів.* Прогнозуйте життєву цінність клієнта та ймовірність конверсії; пропонуйте користувачам товари, які дійсно їм потрібні завдяки підвищенню рівня персоналізації.

*Вчасно приймати рішення на основі даних.* Отримуючи якісні дані, ви матимете міцну основу для своєчасного прийняття рішень, швидкого тестування гіпотез і впевнені у своїх наступних кроках.

Інструменти для роботи з даними в маркетингу відрізняються за рівнем деталізації та методами інтерпретації. Як правило, вони зосереджені на збиранні, аналізі чи візуалізації інформації. Тим не менш, кількість варіантів на ринку досить велика, і знайти ідеальний інструмент не так просто. Давайте розберемося, як відбувається аналіз і що слід враховувати при виборі інструмента.

Ручний збір, очищення та систематизація даних займають багато часу, до того ж, припускають високу ймовірність людської помилки. Низька якість даних може призвести до втрати вашого маркетингового бюджету. Ви можете годинами шукати помилки у купі таблиць без гарантії якості результатів. Щоб заощадити час та переконатися в точності отриманої інформації, автоматизуйте аналіз за допомогою спеціалізованих інструментів.

Перед вибором інструмента звуйте список можливих варіантів, визначивши:

- типи даних, які ви хочете аналізувати
- мета, яку ви хочете досягти за допомогою аналізу
- рівень деталізації, якого ви хочете досягти у своєму аналізі

Тепер погляньмо на три найпопулярніших інструмента аналізу даних.

Першим розглянемо платформу бізнес-аналітики Datorama (рис 1.1) на базі штучного інтелекту з вбудованими функціями звітності та автоматичними прогнозами щодо досягнення цілей. Він збирає та поєднує маркетингові дані на одній централізованій платформі, аналізує їх по різних каналах та кампаніях, а також будує звіти на основі результатів.

Ключові характеристики:

- Підходить для B2B-маркетологів та підприємств

- Дозволяє кросплатформне управління: налаштовувати кампанії в Google, Facebook та інших інструментах, підключених до Datorama
- Висока ціна планів без можливості платити щомісяця — доступна лише щорічна оплата



Рисунок 1.1 Datorama

Другим розглянемо хмарну ETL платформу Supermetrics (рис 1.2) , яка збирає та аналізує ваші маркетингові дані за допомогою API-інтерфейсів та відправляє їх у сховище або інструмент візуалізації, звітності чи бізнес-аналітики на ваш вибір. Supermetrics дозволяє автоматизувати звітність та передавати результати туди, куди ви хочете.

Ключові характеристики:

- Не пропонує моніторингу якості та звітів про атрибуцію
- Можна інтегрувати з Google Data Studio
- Пропонує кілька інтеграцій, у тому числі з такими джерелами, як Google Analytics та Google Реклама



Рисунок 1.2 Supermetrics

Третім розглянемо автоматизовану ETL платформу Fivetran (рис 1.3) для інтеграції та аналітики, що дозволяє об'єднувати дані з кількох джерел в одному сервісі.

Ключові характеристики:

- Завдяки функції реплікації може аналізувати інформацію, якої вже не існує в системі джерела
- Не зберігає дані, а завантажує їх у вибране вами сховище
- Пропонує історичний режим для аналізу даних за певний момент часу, але цей параметр доступний не для всіх конекторів



Рисунок 1.3 FiveTran

Завдяки усім цим 3 платформам , авжеж можна автоматизувати аналітичну діяльність , але якщо додати машинне навчання , то можливо отримати значно точнішу інформацію, та й ще удосконалити обробку даних при формуванні цільової аудиторії компанії та удосконалити аналіз ефективності комерційної пропозицій. Докладніше порівняння в таблиці 1

Таблиця 1 Порівняння способів автоматизації аналітичної діяльності.

НАЗВА	ТОЧНІСТЬ	ФОРМУВАННЯ РЕКОМЕНДАЦІЙ	АНАЛІЗ ЕФЕКТИВНОСТІ	ШВИДКІСТЬ
FiveTran	83%	—	—	1.9 с
Supermetrics	87%	—	—	2.2 с
Datorama	89%	—	—	1.7 с
Froid	92%	+	+	2.1 с

## 1.2 Джерела інформації для аналітичної діяльності

Інформація, що є сукупністю відомостей і даних про діяльність підприємства, його організаційну структуру, зовнішні і внутрішні зв'язки, технології виробництва і асортимент продукції, що випускається, а також інші моменти, що стосуються його діяльності, називається комерційною.

До комерційної інформації відносяться наступні відомості:

- дані про зовнішнє середовище підприємства (ринок, законодавчі норми, а також діяльність конкурентів);
- внутрішня інформація (про виробничу діяльність, резерви компанії);
- відомості про споживачів продукції та послуг організації (їх фінансові можливості, а також віковий, національний, географічний склад);
- дані про конкурентів, (асортимент продукції, прийоми збуту, цінова політика);
- інформація про товари та послуги, що реалізуються на ринку (якісний, кількісний склад, вартість);
- дані про ринковий рівень цін;
- інформація про канали збуту та розподілу продукції, механізми доведення їх до кінцевого споживача;
- рекламні матеріали, а також відомості про їх ефективність.

Комерційну інформацію прийнято поділяти на 6 основних груп:

- ділова (ставиться до роботи з контрагентами);
- науково-технічна (описує нововведення та технологію виробництва);
- виробнича (планові та фактичні показники з випуску продукції);
- управлінська (організація керівництва підприємством);

- маркетингова (відомості про ринок);
- фінансова (грошові потоки).

Інформація в комерційній діяльності може поширюватися двома основними шляхами:

- за допомогою періодичного друку - у вигляді спеціалізованих журналів і каталогів (варто відзначити, що вони потребують постійного оновлення у міру зміни ситуації на підприємстві);
- шляхом створення електронних баз даних, які можуть містити як комплексні відомості щодо кожного підприємства, так і деталізацію за окремими напрямками їхньої комерційної діяльності.

Важливо розуміти, що у складі комерційної інформації входять як дані, що у загальному доступі, і закриті. До перших можна віднести контактні дані організації, імена її керівників, напрями виробничої та інших видів діяльності, а також відомості про деякі активи. Що стосується закритих даних, то вони визначаються організаційною формою підприємства - це може бути інформація про кредити, партнерські угоди, технології виробництва, і не підлягають поширенню.

Інформація комерційної організації поділяється на кілька різновидів:

- відкриті дані загального користування: реєстраційні та контактні дані організації, види її діяльності, асортимент продукції, що випускається, або перелік послуг, що надаються, та інші дані, що знаходяться у відкритому доступі;
- інформація, призначена для контролюючих органів (різноманітні форми звітності);
- закриті відомості, доступ до яких має лише безпосереднє керівництво фірми, а також обмежене коло співробітників: технологія виробництва та матеріали, договірні відносини, інша інформація, яка є комерційною таємницею.



Комерційна інформація має бути достовірною та актуальною для забезпечення безперебійної роботи підприємства, а особливо його управлінської ланки. Вона дозволяє правильно оцінити ситуацію, що склалася, і вжити відповідних заходів. У сучасних умовах комерційна інформація набуває особливої цінності, адже вона характеризується цілою низкою переваг:

- дає можливість постійно аналізувати економічну та виробничу ситуацію;
- використовується при складанні планів, прогнозів, а також на стадії контролю;
- своєчасне надходження достовірної інформації дає організації низку конкурентних переваг;
- відображає поточне ставлення до підприємства як з боку клієнтів, так і з боку партнерів та конкурентів;
- постійний контроль ситуації у внутрішньому та зовнішньому середовищі;
- достатній обсяг відомостей про підприємство у відкритому доступі покращує його імідж та підвищує рівень довіри;
- дозволяє вчасно вносити корективи у стратегічні цілі організації;

Комерційна інформація може бути отримана з таких джерел:

- маркетингові дослідження (проводиться самостійно, а також може бути надана партнерами та контрагентами);
- внутрішня документація фірми (різного роду звітні, а також поточна документація, що оформляється для підтвердження проведення виробничих операцій);
- зовнішні джерела (статистична інформація спеціальних служб, організацій та агентств, що публікується в періодичних виданнях та знаходиться у відкритому доступі).

Варто зазначити, що комерційна інформація досить неоднорідна. У зв'язку з цим було розроблено таку класифікацію:

1) по признаку часу

- ретроспективна (дозволяє оцінити ситуацію, що склалася
- у попередніх періодах для коригування майбутніх планів);
- поточна (дає уявлення про нинішні результати діяльності);
- прогнозна (визначає передбачуваний стан організації у майбутніх періодах);

2) відносно етапу прийняття управлінських рішень:

- констатуюча (описує поточну ситуацію на підприємстві в цілому, а також в окремих його підрозділах);
- пояснювальна (необхідна розуміння чинників, які призвели до тієї чи іншої ситуації);
- планова (дає уявлення про майбутній стан підприємства);
- контрольна (дозволяє оцінити відповідність поточних показників заявленим);

3) за охопленням:

- загальна (вміщає дані по підприємству загалом);
- локальна (описує стан окремих структур та підрозділів
- організації);

4) відповідно до призначення:

- планова (використовується визначення майбутніх показників);
- облікова (заноситься до документів для фіксації нинішнього стану справ);
- нормативно-довідкова (використовується для приведення виробничих процесів у відповідність до правил);
- статистична (дозволяє відстежити динаміку показників за останні кілька періодів);

- маркетингова (описує ситуацію над ринком).

5) за частотою надходження:

- дискретна (надходить з певною періодичністю);
- епізодична (надходить одноразово, у разі виникнення необхідності);
- безперервна (застосовується для моніторингу поточної ситуації).

Використання комерційної інформації багато в чому зумовлює ефективне функціонування фірми. Це повна сукупність відомостей про діяльність підприємства, яку використовують керівники для ухвалення управлінських рішень, а також зовнішні користувачі з метою ознайомлення з організацією.

При проведенні економічного аналізу та розроблення планів на перспективу використовуються різні джерела інформації, серед яких основними є централізовані (нормативні), планові, обліково-звітні та поза облікові.

До централізованих джерел відносяться всі нормативи, що діють на дату проведення аналізу (ставки податків, тарифи на оплату послуг, норми амортизації тощо).

Усі показники поточних і стратегічних планів, які розробляють безпосередньо на підприємстві, відносяться до планових джерел.

Обліково-звітними джерелами є дані бухгалтерського, статистичного, оперативного обліку та звітності.

Бухгалтерська інформація являє собою сукупність показників, які відображаються у фінансовій звітності та дають відомості про обіг усіх коштів

та ресурсів підприємства на підставі обліку та суворого документування на рахунках, балансі та інших облікових формах.

Основними документами бухгалтерської звітності, що використовуються в аналізі та плануванні, є форма № 1 «Бухгалтерський

баланс», форма № 2 «Звіт про прибутки та збитки». Крім цього, можуть використовуватися й інші документи: форма № 3 «Звіт про зміни капіталу», форма № 4 «Звіт про рух грошових коштів», форма № 5 «Додаток до бухгалтерського балансу».

Статистична інформація - це відомості, що відображаються в офіційних формах статистичної звітності. Для аналізу може використовуватися форма №5-з «Відомості про витрати на виробництво та реалізацію продукції (робіт, послуг)».

Оперативна інформація обслуговує поточну, щоденну діяльність підприємства та виконує в основному функцію регулювання та контролю за виробничо-торговельною діяльністю. Це можуть бути товарні звіти, дані про штатну чисельність персоналу, про кредиторську заборгованість тощо.

Решта джерел аналізу вважаються поза обліковими. До них зазвичай належать офіційні документи, якими користується організація у своїй діяльності (закони держави, укази президента, постанови уряду, акти перевірок та ревізій та ін.), господарсько-правові документи (статут, установчий договір, договори на постачання товарів, оренду приміщень або обладнання та ін.), пояснювальні записки, протоколи нарад, дані маркетингових досліджень.

Аналітична інформація є сукупність даних про стан досліджуваного об'єкта та ступеня впливу факторів на його зміну.

До аналітичної інформації мають пред'являтися найвищі вимоги. До основних вимог належать такі:

- інформація повинна відображати мету економічного аналізу,
- аналізовані процеси та явища;
- щодо аналізу необхідний оптимальний обсяг інформації;
- інформація має бути оперативною, тобто безпосередньо відображає зміну параметрів економічних явищ;
- інформація має бути якісною (достовірною, забезпечуватися

- системою обліку, легко оброблятися тощо).

Економічний аналіз діяльності суб'єктів господарювання виробив свій метод, заснований на діалектиці і є способом її застосування до вивчення його предмета. Характерними рисами методу економічного аналізу є:

- використання системи показників, що всебічно характеризують господарську діяльність,
- вивчення причин зміни цих показників,
- виявлення та вимірювання взаємозв'язку між ними з метою підвищення соціально-економічної ефективності.

Діалектичність підходу до вивчення господарських процесів передбачає їх розгляд у становленні та розвитку. Їм властиві закономірність переходу кількості в якість, поява нової якості, заперечення заперечень, боротьба протилежностей, відмирання старого і поява нового, прогресивнішого, тобто. всі методи та прийоми діалектики проявляються у дослідженні господарських процесів, господарської діяльності на всіх рівнях.

Перевагою такого визначення методу аналізу є те, що в ньому наголошується на значенні розробки та використання системи показників як найбільш характерна особливість методу аналізу. Це привернуло увагу широкого кола фахівців до проблем удосконалення системи показників з метою оцінки ефективності діяльності об'єкта, що аналізується, для вимірювання ступеня використання економічного потенціалу та виявлення резервів підвищення ефективності господарювання.

Водночас у сучасних умовах необхідно чіткіше підкреслити системність та комплексність аналізу. Системний комплексний підхід враховує цілісність діяльності суб'єкта господарювання, взаємозв'язок і взаємодію окремих її сторін та їх суперечливий вплив на кінцеві результати цієї діяльності.

Таким чином, метод аналізу це системний комплексний підхід до вивчення результатів діяльності суб'єкта господарювання, виявлення та

вимірювання суперечливого впливу на них окремих факторів, узагальнення матеріалів аналізу у вигляді висновків та рекомендацій на основі обробки спеціальними прийомами всієї наявної інформації про результати цієї діяльності.

### **1.3 Методи машинного навчання**

Машинне навчання вважається гілкою штучного інтелекту, основна ідея якого полягає в тому, щоб комп'ютер не просто використовував наперед написаний алгоритм, а сам навчився вирішення поставленого завдання. У традиційних обчисленнях алгоритми є наборами явно запрограмованих інструкцій, використовуваних комп'ютерами до виконання розрахунків чи розв'язання проблеми. А алгоритми машинного навчання дозволяють комп'ютерам навчатися на вхідних даних та використовувати статистичний аналіз для виведення значень, які потрапляють у певний діапазон. Завдяки цьому комп'ютери стають більш продуктивними у побудові моделей із вибірових даних, що дозволяє автоматизувати процеси прийняття рішень на основі заданої інформації.

Машинне навчання широко застосовується в багатьох областях, які, так чи інакше, займаються збиранням та аналізом даних. Машинне навчання є, по суті, розвитком методів апроксимації функції, але як точки виступають складніші об'єкти, елементи складно описаних просторів, а відповідями може бути лише числа, а й безлічі. Завдання машинного навчання зводяться до завдання знаходження невідомої залежності між відомим безліччю об'єктів та безліччю відповідей. Тобто. Необхідно побудувати таку функцію, яка досить точно наближала значення безлічі відповідей у точках безлічі об'єктів і всьому іншому просторі.

Машинне навчання потребує надання досвіду - інакше кажучи, йому

необхідні дані. Чим більше в систему надходить даних, тим точніше комп'ютер взаємодіє з ними, а також з даними, що отримує надалі. Чим вища точність взаємодії, тим успішнішим буде виконання поставленого завдання, і вищий ступінь прогностичної точності.

У машинному навчанні завдання зазвичай поділяються на широкі категорії. Ці категорії засновані на тому, як проходить навчання або як розроблена система отримує зворотний зв'язок.

Два методи машинного навчання що найбільш широко використовуються:

- навчання з вчителем, де задіяні алгоритми, що ґрунтуються на прикладах вхідних та вихідних даних, які марковані людьми;
- навчання без вчителя, де використовуються алгоритми без маркованих даних, що дозволяє машинам самостійно знаходити структуру всередині своїх вхідних даних.

Машинне навчання з вчителем - під час навчання з учителем комп'ютер отримує зразки вхідних даних із зазначеними бажаними результатами. Мета цього методу полягає в тому, щоб алгоритм міг «вчитися» шляхом порівняння свого висновку із заздалегідь зазначеними результатами та відповідним чином коригувати модель. Таким чином, метод навчання використовує шаблони для прогнозування значень додаткових немаркованих даних.

Як правило, машинне навчання з учителем використовується для прогнозування статистично ймовірних майбутніх подій на основі історичних даних. Такі алгоритми можуть використовувати історичну інформацію про фондовий ринок для прогнозування майбутніх коливань або фільтрувати спам. Під час навчання з учителем марковані фотографії собак можуть використовуватися як вхідні дані для класифікації немаркованих зображень собак.

Машинне навчання без вчителя - при навчанні без вчителя алгоритм отримує немарковані дані, тому потрібно самостійно знайти схожість у його вхідних даних. Немаркованих даних зазвичай набагато більше, ніж маркованих, тому особливо корисними сьогодні є методи машинного навчання, які полегшують навчання без вчителя.

Мета навчання без вчителя – виявлення у наборі даних явних і прихованих шаблонів і характеристик, що дозволяє обчислювальній машині автоматично виявляти схожість, необхідну класифікації необроблених даних.

Навчання без вчителя зазвичай використовується для транзакційних даних. На основі отриманих даних алгоритм навчання без вчителя може з'ясувати, що дали певного вікового діапазону, які купують велосипедні костюми, ймовірно мають велосипед, або хочуть його купити, і тому ця аудиторія може стати цільовою для маркетингової кампанії, пов'язаної з велосипедами та аксесуарами для них, що збільшить кількість замовлень.

Алгоритм, який не знає «правильних» відповідей, може аналізувати та організовувати ширші та непов'язані набори даних. Навчання без вчителя часто використовується для виявлення аномалій (у тому числі виявлення для шахрайства з кредитними картками) та створення систем рекомендацій, які радять користувачеві, які продукти купити/який фільм подивитися на основі його переваг. При навчанні без вчителя немарковані зображення собак можуть використовуватися як вхідні дані для алгоритму пошуку подібностей і класифікації всіх фотографій собак.

Розглянемо основні п'ять методів машинного навчання.

### 1.3.1 Нейронні мережі

Нейронні мережі імітують структуру мозку: кожен штучний нейрон з'єднується з кількома іншими нейронами. Нейронні мережі мають багатосарову структуру: нейрони одному шарі передають дані кільком



нейронам наступного тощо. буд. Зрештою, дані досягають вихідного шару, де мережа

видає припущення у тому, як розв'язати завдання чи класифікувати об'єкт.

Нейронні мережі застосовують у цілому ряді галузей. У охороні здоров'я їх використовують під час аналізу медичних знімків з метою прискорення діагностичних процедур та пошуку ліків. У телекомунікаційній галузі та медіа індустрії нейронні мережі можна застосовувати для машинного перекладу, розпізнавання шахрайства та надання послуг віртуальних помічників. У фінансовій галузі їх використовують для розпізнавання шахрайства, управління портфелями та аналізу ризику. У роздрібній торгівлі — для позбавлення черг до каси та для персоналізації обслуговування покупців.

### 1.3.2 Дерево рішень

Алгоритм дерева рішень класифікує об'єкти, відповідаючи на «питання» про їх атрибути, розташовані у вузлових точках. Залежно від відповіді вибирається одна з гілок, і так доти, доки не буде досягнуто «аркуш» — остаточна відповідь.

Серед застосувань дерева рішень – платформи управління знаннями для клієнтського обслуговування, прогностного призначення цін та планування випуску продукції.

У страховій компанії дерево рішень допоможе з'ясувати, які види страхових продуктів та премій краще використовувати з урахуванням можливого ризику. Використовуючи дані про місцезнаходження та відомості про страхові випадки з урахуванням погодних умов, система може визначати категорії ризику на підставі поданих вимог та витрачених сум. Потім, використовуючи моделі, система оцінюватиме нові заяви про страховий захист, класифікуючи їх за категорією ризику та можливими фінансовими збитками.

### 1.3.3 «Випадковий ліс»

Щоб одиночне дерево рішень давало точні результати, його треба навчати, алгоритм випадкового лісу (random forest) використовує «комітет» випадковим чином створених вирішальних дерев з різними наборами атрибутів і дає можливість їм проголосувати, щоб вибрати найпопулярніший клас.

Випадковий ліс — універсальний механізм, що швидко навчається для виявлення зв'язків усередині набору даних. У приклад можна навести небажані масові розсилки, що створюють проблеми не тільки користувачам, а й провайдером Інтернету, яким через спам доводиться мати справу з підвищеним навантаженням на сервери. Для боротьби з проблемою було розроблено автоматизовані методи фільтрації спаму, які за допомогою ансамблю вирішальних дерев швидко та ефективно визначають небажані листи.

Серед інших застосувань — діагностика захворювань шляхом аналізу медичної картки пацієнта, розпізнавання банківського шахрайства прогнозування кількості дзвінків у колл-центрах та прогнозування ймовірності прибутку та збитку при купівлі певних акцій.

### 1.3.4 Кластеризація

Кластеризація - це групування елементів даних, що мають подібні характеристики, за допомогою статистичних алгоритмів. Це спосіб навчання без вчителя, який можна використовувати для розв'язання завдань класифікації. Наприклад, сегментування купівельної аудиторії залежно від показників для уточнення адресації рекламних кампаній рекомендації новини конкретним читачам ; допомога у роботі правоохоронних органів.

Кластеризація також дієва, коли у складних наборах даних потрібно виявити групи, які важко помітити без спеціальних засобів. Область

застосування дуже широка: від групування схожих документів у базі даних до виявлення за кримінальними новинами території із підвищеним рівнем злочинності.

### 1.3.5 Пошук асоціативних правил

Пошук асоціативних правил - це метод навчання без вчителя, що дозволяє знаходити відносини між змінними. Використовується в двигунах видачі рекомендацій - саме цей метод застосовується в багатьох інтернет-магазинах для складання фрази "Разом з цим товаром зазвичай купують ...".

Конкретний приклад - підвищення продажів у магазин. Вивчивши купівельну поведінку шляхом пошуку асоціативних правил, можна пропонувати спеціальну упаковку та набори для свят та інших особливих випадків. Асоціативні правила дозволяють з'ясувати, коли і за яких обставин покупці набувають тих чи інших поєднань товарів. Використовуючи відомості про минулі покупки та час їх здійснення, можна скласти програму знижок та сформулювати індивідуальні пропозиції з розрахунком на підвищення продажів.

### 1.3.6 Попередня обробка даних

Після визначення найкращого алгоритму для вирішення задачі важливими етапами, які забезпечують ефективне використання набору даних для машинного навчання, є попередня обробка та очищення даних. Необроблені дані часто спотворені та ненадійні, і в них можуть бути пропущені значення. Використання таких даних при моделюванні може призвести до неправильних результатів. Ці завдання є частиною процесу обробки та аналізу даних групи і зазвичай мають на увазі початкове вивчення набору даних, що використовується для визначення та планування необхідної попередньої обробки.

Для того, щоб перевірити якість даних, необхідно оцінити таке:

- кількість записів;
- кількість атрибутів (або компонентів);
- типи даних атрибуту (номінальні, порядкові чи безперервні);
- кількість пропущених значень;
- правильність формату даних;
- неузгоджені записи даних.

При виявленні проблем з даними необхідно виконати обробку, яка часто включає очищення пропущених значень, нормалізацію даних, дискретизацію, обробку тексту для видалення та/або заміни впроваджених символів, які можуть впливати на вирівнювання даних, змішані типи даних у загальних полях та ін.

Головні завдання попередньої обробки даних:

- Очищення даних. Заповнення пропущених значень, виявлення та видалення спотворених даних та викидів.
- Перетворення даних. Нормалізація даних зниження вимірювань і спотворень.
- Ущільнення даних. Створення вибірки даних або атрибутів для спрощення обробки даних.
- Дискретизація даних. Перетворення безперервних атрибутів на категоріальні, щоб простіше було використовувати деякі методи машинного навчання.
- Очищення тексту. Видалення впроваджених символів, які можуть порушувати вирівнювання даних.

#### **1.4 Аналіз джерел інформації і можливостей їх використання для автоматизованої обробки даних**

Фінансовий успіх підприємства залежить від оперативності прийняття виробничих рішень та здатності перебудувати бізнес-процеси в максимально стислий термін, якщо цього вимагає ситуація на ринку. Автоматизована обробка інформації забезпечить продуктивну діяльність компанії з усіх обраних напрямів.

Під автоматизованою обробкою даних розуміється їхня обробка за допомогою засобів обчислювальної техніки. Засобами обчислювальної техніки можуть бути електронні обчислювальні машини, комплекси та мережі, допоміжні та периферійні пристрої, у тому числі встановлене програмне забезпечення.

Стратегія розвитку фірми безпосередньо пов'язана зі збором та аналізом даних, що надходять від зовнішніх та внутрішніх джерел.

Впровадження автоматизованих систем управління та обробки даних у компанію для роботи з необхідною інформацією має ряд переваг:

- зменшується кількість працівників, що призводить до зниження витрат на заробітну плату (ключова ознака);
- при зміні будь-яких окремих параметрів уже готової звітності нові значення перераховуються в найкоротші терміни;
- при дослідженні конкурентів, зрізів ринку та внутрішніх бізнес-процесів комп'ютеризовані системи формують єдину базу даних з можливістю сортувати інформацію по-різному для подальшого порівняльного аналізу за вибраними параметрами;
- максимальна оперативність аналізу.

Формалізовані бази даних, що утворюються в результаті автоматичного збору інформації, включають:

- класифікацію об'єктів, що містяться, відповідно до офіційно
- затверджених класифікаторів;
- шаблонний опис параметрів;
- кодування та інші засоби безпеки для захисту інформації
- ідентифікацію кожного об'єкта виходячи з його унікальних характеристик;

Для автоматизованої обробки даних найкраще підходять обліково-звітні джерела, наприклад, бухгалтерська звітність, так як у повному та достовірному обсязі подає інформацію про фінансове становище організації, фінансові результати її діяльності та зміни у її фінансовому становищі, і має чітко визначену форму надання.

Автоматизація облікового процесу може частково змінити методіку бухгалтерського обліку. Це проявляється в наступному:

- автоматизоване документування (у бухгалтерському обліку документування – це найбільш трудомісткий процес) може одночасно з розробкою документа формувати типові бухгалтерські проводки, що значно скорочує обліковий процес та дозволяє своєчасно створювати необхідні звіти;
- підвищується можливість аналізу розрахунків завдяки можливості додавання до окремих бухгалтерських рахунків додаткових ознак аналітики, внаслідок чого цих рахунках крім облікових даних можуть відбиватися планові чи нормативні показники;
- прискорюється процес калькулювання з допомогою збільшення швидкості виконання арифметичних розрахунків;

- швидше відбувається процес формування бухгалтерської звітності, отже, скорочується звітний період (баланс за потреби можна за запитом скласти щотижня і щодня), і навіть підвищується достовірність і зручність аналізу звітності.

Таким чином, автоматизація процесів управління первинною документацією допомагає домогтися прозорості в роботі, позбавляє рутинної роботи, зменшує кількість помилок, практично повністю ліквідує ризики втрат і перепусток документів, значно збільшує швидкість введення даних в облікову систему та отримання звітності, а зрештою підвищує ефективність роботи всієї компанії.

Розглянемо вплив обробки інформації за умов розвитку цифрової економіки з прикладу ринку велосипедної роздрібною торгівлі.

Минула криза показала, що ринок велосипедів — це один з стійкіших ринків . У 2008-2009 роках він виріс на 5% (швидше зростав тільки продуктивний роздріб — на 11%). У 2016-2017 роках ринок велосипедів — один із небагатьох у непродовольчому роздробі, який показав зростання не лише в обсязі, а й у грошовому еквіваленті. Це актуалізувало вибір об'єкта та проведення компаративного аналізу.

Для вирішення поставленого завдання було використано метод побудови трикомпонентної фазової діаграми. При компаративному аналізі, зазвичай, проводять класифікацію досліджуваної сукупності об'єктів з визначення груп мереж на ринку велосипедної роздрібною торгівлі, з однаковим чи близьким типом поведінки. Класифікація може проводитись у різний спосіб. Найбільш відомим є метод кластерного аналізу. Кінцевою метою дослідження фінансової та бухгалтерської звітності мереж є покращення фінансових показників, підвищення ліквідності компаній на ринку.

Інформаційною основою стали бухгалтерська звітність та фінансові діяльності компаній за 2011-2017 роки. Вибрані показники відповідали вимогам: по-перше, аналізовані показники мереж повинні мати однорідну структуру; по-друге, мають бути представлені по всіх аналізованих мережах; по-третє, кількість спостережень має бути достатньо для аналізу та обґрунтованих висновків. Безумовно, дані показники відбивають лише окремі боку діяльності компаній, але найістотніші, які мають позиції мереж над ринком та його динаміку розвитку.

Однак за існуючих відмінностей у підходах ведення економічної політики компаніями загальними рисами залишаються: активний розвиток маркетингових кампаній; збереження та збільшення обсягів продажу в умовах стагнації; зростання низького цінового сегмента; розвиток сегмента аксесуарів, при загальному скороченні життєвого циклу комплектуючих; розвиток унікальної товарної пропозиції (УТП) та власних торгових марок (ВТМ).

Ці пріоритети дозволили виділити три типи основних стратегій фінансового розвитку магазинів велосипедів, що відповідають конкретним умовам та набору певних факторів:

- стрес-толеранти (S) («патієнти»), що мають низькі темпи розвитку, зумовлені значним дефіцитом ресурсів;
- конкуренти (C) («віоленти»), моделі поведінки яких характеризуються володінням ефективними механізмами забезпечення доступу до ресурсів та інтенсифікацією виробництва;
- рудерали (R) («експлеренти»), які прагнуть максимізації надходжень (доходів) в умовах слабкої конкуренції та значного



обсягу ресурсів, ефективність використання останніх має другорядне значення.

Учасники ринку із чітко вираженими первинними стратегіями зустрічаються вкрай рідко. Найчастіше мають місце звані вторинні стратегії, які є змішанням первинних стратегій. При цьому пояснення моделей економічної поведінки учасників в умовах конкурентної боротьби вимагає ідентифікації їх стратегій.

Методика ідентифікації стратегій конкурентної поведінки економічних агентів використовує трикомпонентну фазову діаграму, що зображується на поверхні. Класифікаційна діаграма є області значень відносних індексів ( $I_C$ -конкуренції;  $I_S$  - стресу;  $I_D$  - порушення), кожен з яких відповідає за належність до відповідної первинної стратегії. Таким чином, агенти з первинними стратегіями розташовуються в кутах (R, S або C), а поле трикутника розбивається на чотири вторинних типи, між якими у свою чергу також можливі переходи.

При побудові діаграми з її осей відкладаються з трьох основних індексів (рис. 1). Третій індекс завжди є залежним, оскільки всі свої можливості та зусилля агент використовує для протистояння трьом основним типам зовнішнього впливу – стресу, порушенням та конкурентам (відповідно S-, R-, C – стратегії). Взнявши значення загальної величини цих зусиль за 100%, можна записати таку формулу (1)

$$I_S + I_R + I_C = 100\% \quad (1)$$

де  $I$  – інтенсивність зусиль щодо реалізації відповідних (S, R, C) стратегій поведінки агента.

Для переходу від моделі до практичних методів виявлення стратегій необхідно знайти відповідні вимірювані характеристики агентів. Таким

чином, стає можливим виявляти конкурентні стратегії поведінки агентів за емпіричними даними на підставі двох незалежних параметрів. При побудові класифікаційної діаграми як індекс конкуренції (ліва похила вісь) використовується абсолютне значення деякого економічного параметра  $M$ , а як індекс стресу горизонтальна вісь) – її відносний приріст  $M'$ , що розраховується за формулою (2):

$$M'_i = \frac{100 \cdot (M_i - M_{i-1})}{M_{i-1}}, \quad (2)$$

Де  $M$  - деякий економічний параметр

Масштаб діаграми на рисунку 1.4 задається виходячи з діапазону значень  $M$  і  $M'$  вихідних даних. Лінії розмітки проводяться паралельно осям: на лівій похилій проводяться паралельно горизонтальній, а на горизонтальній - паралельно лівій похилій.

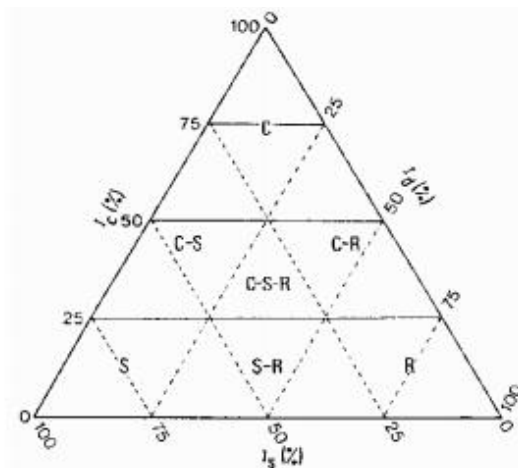


Рис 1.4 Класифікаційна діаграма конкурентних стратегій поведінки агентів, що кількісно характеризує співвідношення конкуренції, стресу та порушення, що відповідають стратегіям різних типів

Для побудови діаграми необхідно розрахувати індекси  $M$  і  $M'$ , ґрунтуючись на економічних даних, що використовуються. Відклавши отримані значення відповідних осей діаграми, на перетині цих значень отримаємо місце розташування об'єкта на діаграмі. Так само можна побудувати динамічну діаграму, використовуючи дані для одного об'єкта в різні моменти часу.

На рисунку 1.5 зображено комбінований графік, що характеризує кожну з компаній за його основними економічними показниками за 2017 рік. З представленого графіка видно, що у всіх учасників, крім 'Спортмастер', частка С стратегії у загальному обсязі протягом усього часу не перевищує 20%. Це вказує на те, що дана компанія дотримується R-S стратегії та адаптується до непродуктивних, нерівноважних умов на ринку. Але тільки 'Спортмастер' протягом практично всього часу дотримується C-S-R стратегії, тобто адаптується до умов, що змінюються, де гострота конкуренції, нерівноважний стан ринку обумовлені середньо інтенсивним стресом і порушенням.

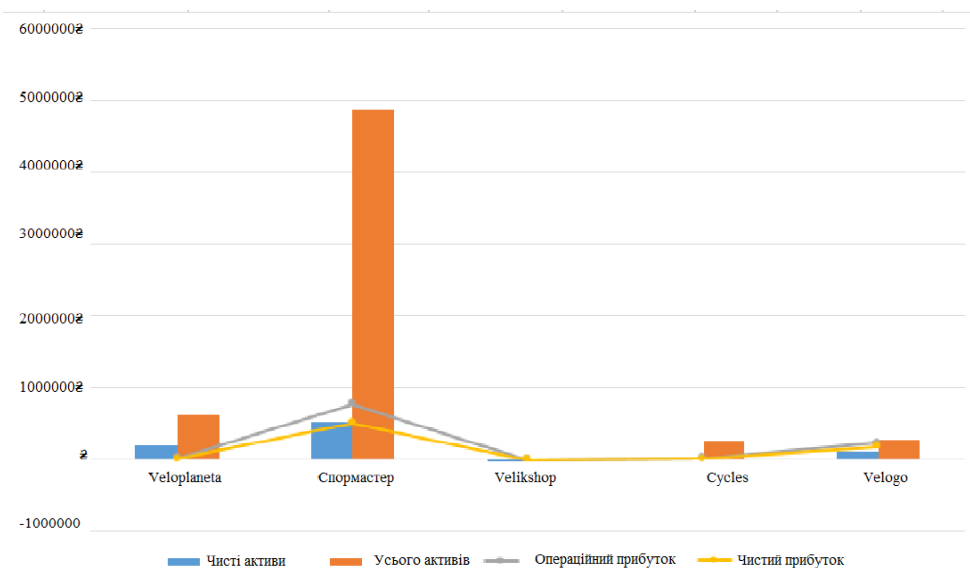


Рис 1.5 Основні економічно показники компаній за 2017 р.

На рисунку 1.6 зображено статичну діаграму чистих активів компаній за 2017 р., а на рис. 1.7 представлено динамічну діаграму чистих активів компаній за 2012–2017 роки. Статична діаграма дозволяє оцінити співвідношення стратегій у час, а загальна динамічна діаграма дозволяє спостерігати зміну стратегій компаній на ринку у проміжок часу з 2012 по 2017 роки.

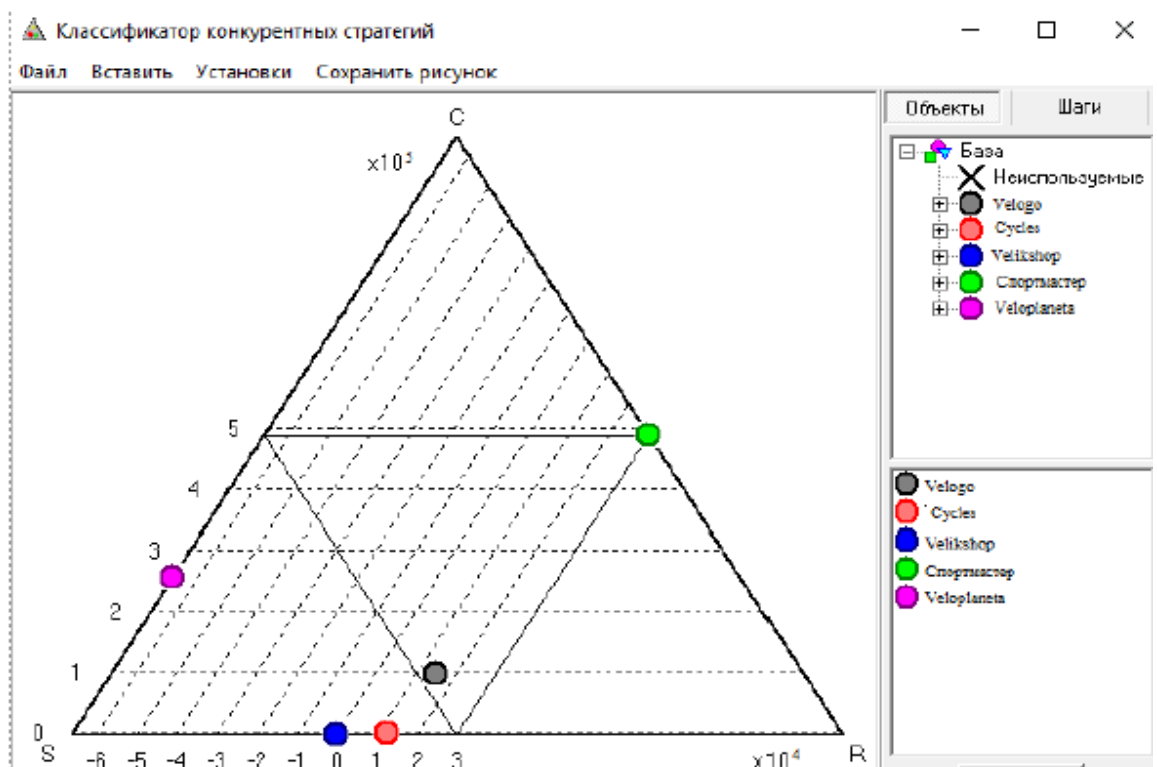


Рис 1.6 Основні економічно показники компаній за 2017 р.

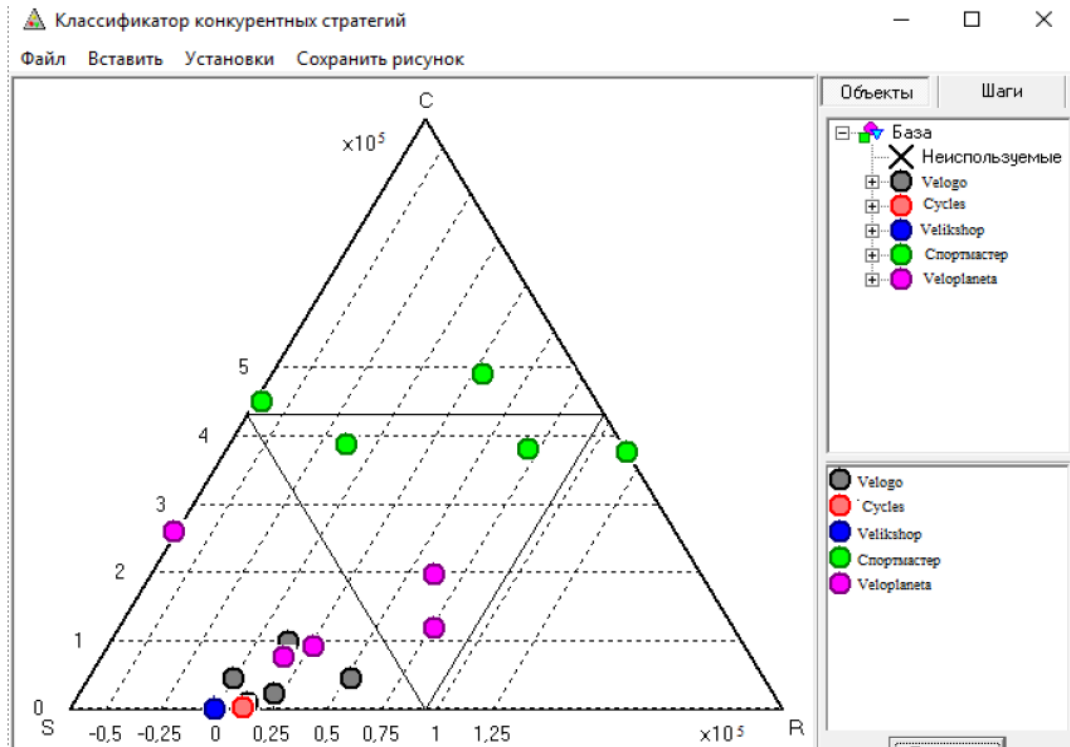


Рис 1.7 Класифікаційна динамічна діаграма чистих активів компанії за 2012-2017 рр.

На основі аналізу даних представленої діаграми можна відзначити, що три мережі, Velikshop, Velogo та Cycles дотримувалися однієї стратегії весь період часу. При цьому хаотично розташовані елементи, що символізують Спортмастер та Veloplaneta, вказують на зміну стратегій за цей проміжок часу.

Таким чином, по-перше, взаємозв'язок між виділеними факторами та економічним розвитком компаній істотно складніший, ніж традиційно представляється у науковій літературі. Це зумовлюється тим, що вплив зазначених факторів опосередковується одночасною присутністю широкого кола інших процесів та умов, тенденцій у суспільстві. По-друге, той самий фактор (наприклад, випуск продукції під власною торговою маркою (ВТМ)) по-різному проявляється у фінансових успіхах різних компаній. Універсальний підхід тут не працює. У цьому сенсі проблема полягає не лише

у додаткових фінансових вливаннях у розвиток бізнесу, а створенні умов, у яких вони давати найбільший ефект у розвиток компанії.

По-третє, як свідчить практика, роль аналізованих чинників протягом може істотно змінюватися. Стійкість фінансових показників компаній проявляється у стабільності економічної та соціальної ситуації в країні, поточних трендах та динаміці розвитку ринку велосипедної роздрібною торгівлі.

В даний час без використання систем, що дозволяють проводити автоматизований, а ще краще автоматичний аналіз потенційних споживачів і підбирати для них індивідуальні пропозиції з урахуванням їх інтересів і особливостей, уявити сучасний бізнес неможливо. Впровадження даного методу збору та аналізу інформації допоможе сформувати конкурентні переваги, а його використання можливе не тільки на ринку велосипедної роздрібною торгівлі, а й на інших ринках збуту

## 2 АВТОМАТИЗАЦІЯ АНАЛІТИЧНОЇ ДІЯЛЬНОСТІ КОМПАНІЇ VELOKYIV

### 2.1 Загальна характеристика підприємства

Товариство з обмеженою відповідальністю «VeloKyiv» надає свої послуги з 2012 р., за більш ніж 5-річний термін компанія нагромадила величезний досвід, знання та клієнтську базу .

Інтернет-магазин орендує приміщення під склад, що знаходиться в м. Київ загальною площею 228 м<sup>2</sup>.

Унікальність ТОВ «VeloKyiv» полягає у наданні повного сервісу:

- по прийманню ваших велосипедів (можна здати його іншого міста поштою, здати речі через кур'єра, викликати оцінювача додому, самостійно приїхати в день прийому з оцінювачем на склад), оцінка велосипедів відповідно до стану, бренду, поточних цін на даний асортимент;
- з продажу ваших велосипедів (занесення речей на сайт з фотографіями, описом, вимірами, продажів через інтернет або зі складу, доставка в межах Києва до дверей покупця, а також відправка замовлень поштою по Україні).

Інтернет-магазин зарекомендував рівень свого сервісу та якість результату серед клієнтів із різних міст України. Основні послуги компанії – це продаж та прийом нових та б/у велосипедів через інтернет-магазин та інших торгових майданчиків.

Місія велосипедного інтернет-магазину полягає у пошуку нових господарів для брендових та якісних велосипедів за невеликою вартістю.

Для досягнення цього результату інтернет-магазин розвиває клієнтську базу, скорочує час виконання кожного етапу внутрішніх бізнес-процесів,

удосконалює технологічні процеси та підбирає найкращі трудові кадри. У свою чергу, все це стає запорукою зростаючого прибутку, а також забезпечує стабільність і впевнене зростання в майбутньому.

Стратегія велосипедного онлайн-майданчика як клієнт орієнтованої компанії – вибудовування довгострокових відносин з клієнтами та покупцями, створення максимального психологічного комфорту у співпраці та зміцнення позицій на ринку велосипедів.

На рисунках 5 та 6 представлені основні стратегічні цілі інтернет-магазину.

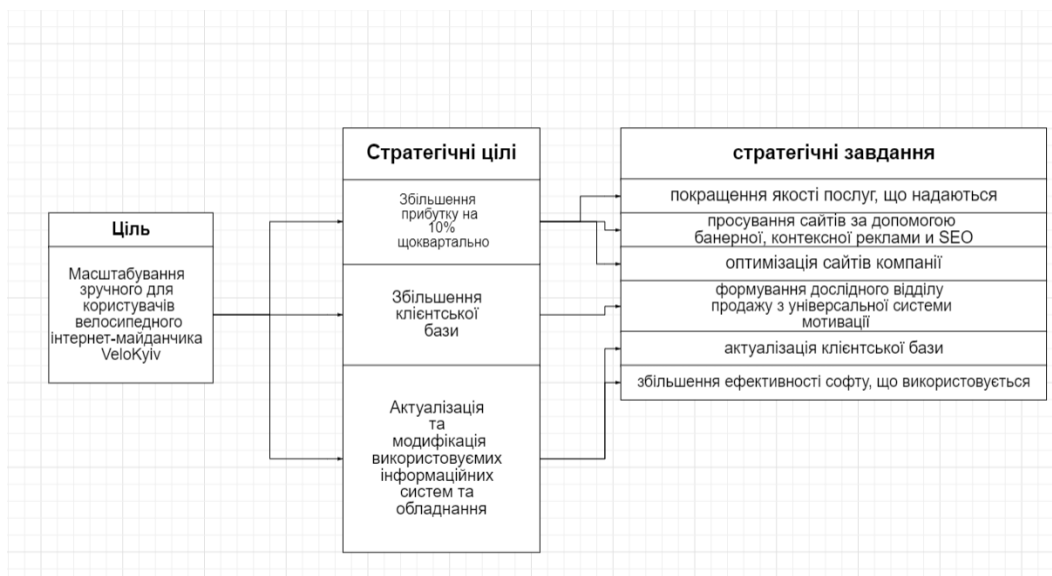


Рис 2.5 - Основні стратегічні цілі



Рис 2.6 Основні стратегічні завдання



Стратегічні цілі велосипедного інтернет-магазину на найближчий рік:

- розширення розмірного асортименту велосипедного одягу та взуття;
- пошук нового приміщення (офісу) у м. Харків у зв'язку зі збільшенням кількості речей, що приймаються на реалізацію;
- пошук та навчання нових контент-менеджерів зі зростанням кількості товарів, що приймаються на реалізацію (збільшення швидкості занесення та прийому на комісію нових товарів).

Бізнес-процеси ТОВ «VeloKyiv» можна згрупувати за трьома основними напрямками (рисунок 7):

- процеси менеджменту – сукупність управлінських видів діяльності (дій, рішень, впливів, аналізу) організації, що забезпечує досягнення поставленої мети організації;
- основні процеси – сукупність взаємозалежних та взаємодіючих видів діяльності, в керованих умовах та за допомогою певних ресурсів;
- процеси забезпечення (допоміжні процеси) – процеси, призначені задля забезпечення результативного функціонування процесів організації.

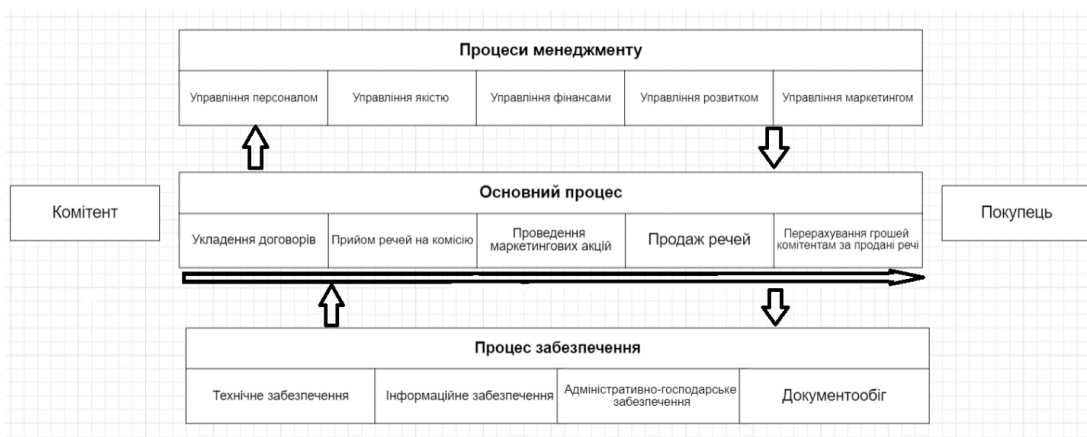


Рис 2.7 Бізнес процеси

Бізнес-процеси велосипедного інтернет-магазину:

- укладання договору з комітентом;
- прийом на комісію велосипедів та аксесуарів;
- обробка та завантаження нових товарів на сайт;
- отримання замовлення та його подальше узгодження;
- доставка товарів покупцю;
- перерахування грошей за продані речі на розрахунковий рахунок комітента;
- підписання розрахункової відомості.

Модель описаного бізнес-процесу зображено рисунку 2.8.



Рис 2.8 Модель бізнес процесів

Виходячи з описаних вище бізнес-процесів, можна виділити основні бізнес функції компанії та показати їх зв'язок із бізнес-процесами(рисунк 9).



Рис 2.9 Зв'язок бізнес-процесів та бізнес-функцій

Найвищу ланку в організаційній структурі ТОВ «VeloKyiv» займає директор. У його підпорядкуванні перебувають:

- начальник відділу продажів;
- начальник відділу контенту;
- оцінювач (1 чол.);
- менеджери (3 чол.);
- контент-менеджери (3 чол.);
- кур'єр (2 чол.);
- інтернет-маркетолог (1 чол.);
- бухгалтер (аутсорс);
- юрист (аутсорс).

Менеджерами безпосередньо керує начальник відділу продажу, оцінювач та контент-менеджери перебувають у підпорядкуванні начальника відділу контенту.

Організаційна структура управління є ієрархічною структурою з чітким поділом праці, а також використання на кожній посаді кваліфікованих фахівців.

На рисунку 10 зображено організаційну структуру інтернет-магазину.

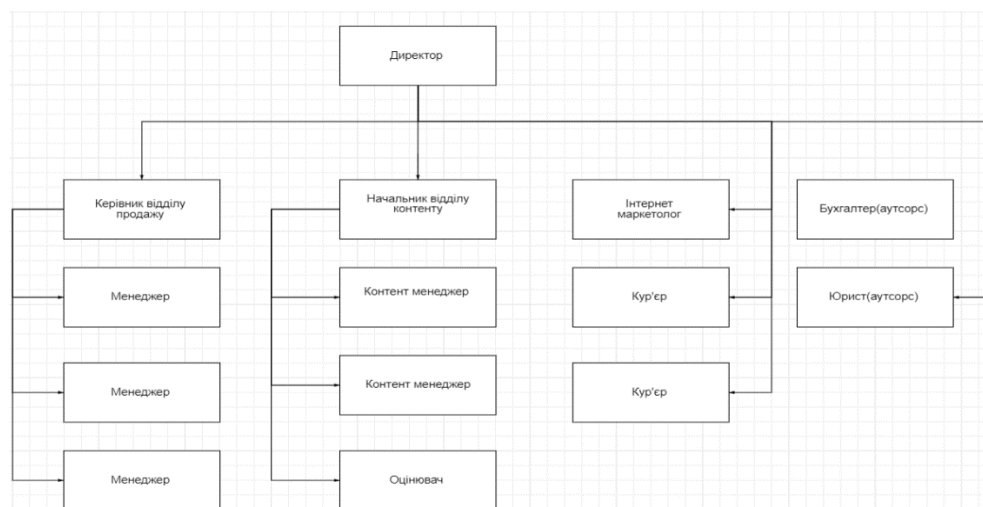


Рис 2.10 Організаційна структура

Велосипедний інтернет-магазин надає кілька видів послуг (рис. 11):

- – продаж велосипедів;
- – продаж велосипедного взуття;
- – продаж велосипедного одягу;
- – продаж аксесуарів для велосипедів;
- – оцінка та прийом на велосипедів;
- – організація кур'єрської доставки товарів по Києву;
- – доставка Поштою по всій території України;
- – здавання велосипедів на комісію через кур'єра;

Розіб'ємо та розглянемо весь спектр послуг компанії в наступних групах:

- – продаж велосипедів, велосипедного одягу та взуття, аксесуарів для велосипеда – щодня десятки нових та майже нових речей знаходять нових господарів по всій Україні;
- – оцінка та прийом велосипедів на комісію (в офісі та через кур'єра) – щомісяця понад тисячу речей дбайливо відбираються, інтернет-магазин приймає на комісію лише 40% речей, які привозять з усієї України;
- – кур'єрська доставка по Києву та поштою по всій території України;

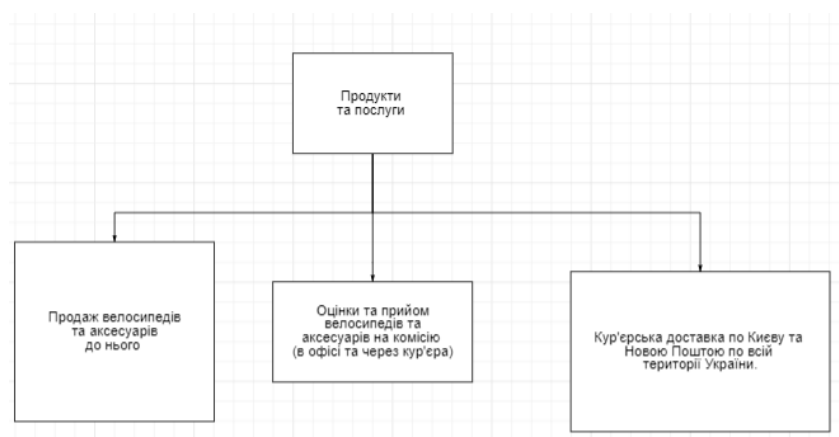


Рис 2.11 Товари та послуги

Корпоративна мережа компанії ТОВ «VeloKyiv» включає:

- файловий сервер;
- сервер 1С;
- локальні комп'ютери;
- принтери Kyocera;
- маршрутизатор D-link;
- комутатори D-link.

На рисунку 12 зображено мережеву структуру офісу.

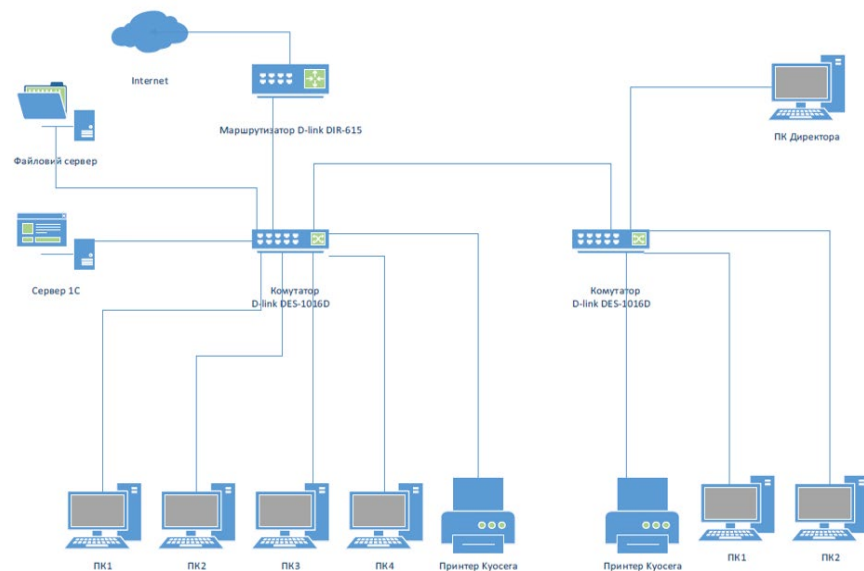


Рис 2.12 мережеву структуру офісу

У компанії використовують такі програмні продукти:

- Бек-офіс, розроблений ЗАТ «FROID»;
- Microsoft Office;
- 1С: Підприємство;
- Adobe Photoshop;
- Adobe Acrobat;
- IP-телефонія.

Модель інформаційних систем наочно представлена рисунку 13

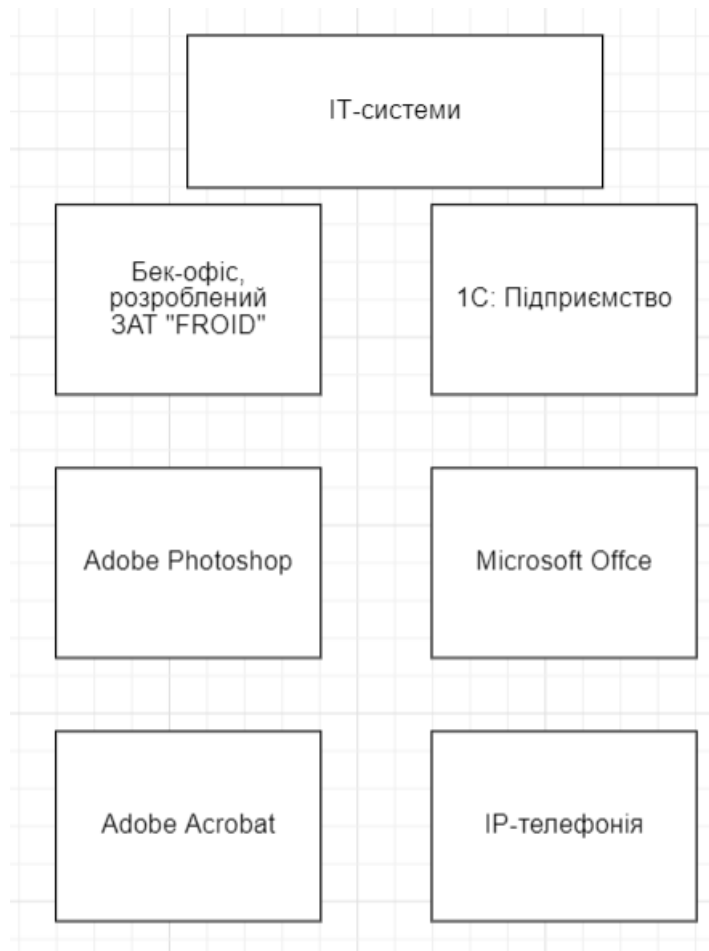


Рис 2.13 Модель інформаційних систем

Співробітники відділу продажів, відділу контенту, а також бухгалтер використовують під час роботи програми 1С: Підприємство, бек офіс та програми пакета Microsoft Office.

У своїй роботі контент-менеджери використовують такі програми для обробки фотографій Adobe Photoshop та Adobe Acrobat.

Усі співробітники компанії використовують ІР-телефонію.

Використання додатків співробітниками представлено у матриці залежностей (таблиця 2).

Таблиця 2 - Матриця залежностей

Додаток	Співробітники відділу продажу	Співробітник відділу контенту	Інтернет - маркетолог	Бухгалтер
Бек-офіс, розроблений ЗАТ FROID	+	+		+
Microsoft Office	+	+	+	+
1С: Підприємство	+	+		+
Adobe Photoshop		+		
Adobe Acrobat		+		
IP-телефонія	+	+	+	+

На рисунку 14 представлена повна модель ТОВ «VeloKyiv», що складається з 3-х рівнів:

- бізнес-стратегія;
- операційний рівень;
- IT-інфраструктура.

Перший рівень – бізнес-стратегія, він включає місії, цілі, завдання організації та зв'язок бізнес-процесів з бізнес-функціями.

другий рівень - операційний рівень, він складається з організаційної структури, наданих продуктів і послуг, а також моделі бізнес-процесів.

Третій рівень – IT-інфраструктура, він є сукупністю програмних продуктів, що використовуються в організації, мережева структура офісу та матриця залежностей.

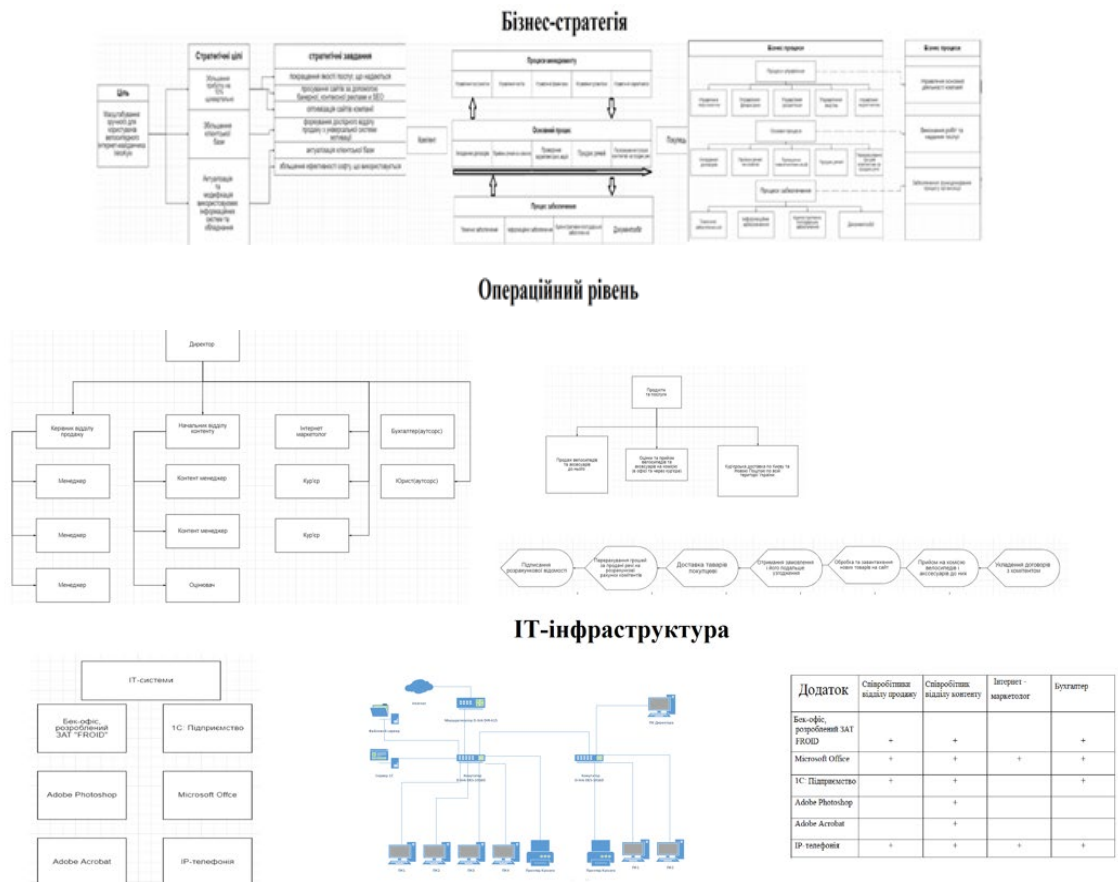


Рис 2.14 Повна модель організації

## 2.2 Автоматизація бізнес-процесів обробки даних при формуванні цільової аудиторії компанії

Моделювання бізнес-процесів – це ефективний засіб, що дозволяє дати оцінку діяльності організації та знайти можливості для її оптимізації, виявити поточні проблеми та недоліки у існуючих бізнес-процесах .

Модель поточного стану організації – це модель AS-IS, тобто модель «як є».

Модель систематизує бізнес-процеси, що протікають в даний момент, а також інформаційні об'єкти, що використовуються в організації. Грунтуючись на цьому можна виявити вузькі місця в організації та взаємодії бізнес-процесів, визначається необхідність тих чи інших змін існуючої структури.



Таку модель зазвичай називають функціональною та виконують з використанням різних графічних нотацій та case-засобів. Важливим моментом на етапі побудови моделі AS-IS вважається створення максимально наближеної до дійсності моделі, що базується на реальних потоках процесів, а не на їхньому ідеалізованому поданні.

Проектування інформаційних систем та управління процесами передбачає побудову моделі AS-IS та подальший перехід до моделі TO-BE.

На даному етапі розвитку компанії не використовує методи машинного навчання для формування індивідуальних пропозицій клієнтам, а лише збирає базу даних щодо клієнтів, їх замовлень та реалізованих речей, тобто необхідно тільки побудувати модель TO-BE.

Модель бізнес-процесів TO-BE – це модель як буде. Відсутність системи формування індивідуальних пропозицій клієнтам негативно б'є по розвитку клієнтської бази та підприємства у цілому. Для усунення цього недоліку існує шлях поліпшення компанії – використання методів машинного навчання на формування індивідуальних пропозицій на підставі обробки інформації по клієнтам за певними критеріями, які будуть відображені у категоріальній базі оцінки.

На рисунку 15 зображено модель TO-BE формування індивідуальних пропозицій клієнтам компанії VeloKyiv.

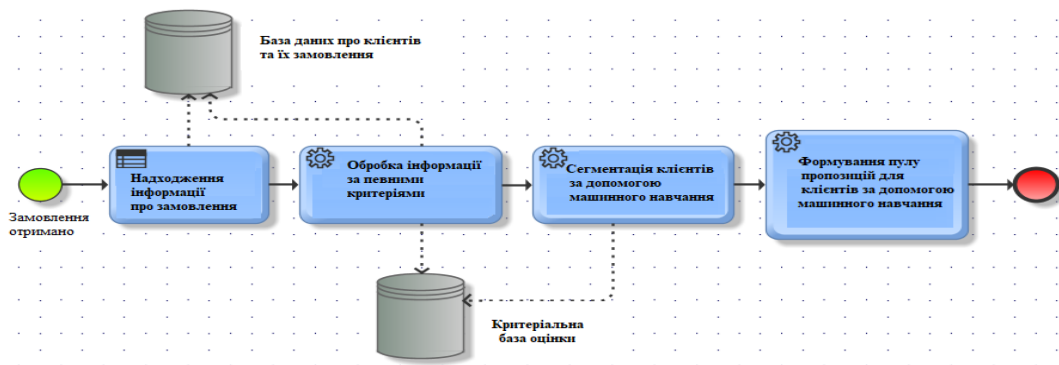


Рис 2.15 Процес формування індивідуальних пропозицій для клієнтів компанії VeloKyiv (to-be)

Модель бізнес-процесу TO-BE компанії VeloKyiv побудована у програмному продукті для моделювання бізнес-процесів організацій – ArisExpress. При побудові була використана нотація BPMN, що відображає хід виконання бізнес-процесів та дозволяє детально та точно описати їх виконання.

Таким чином, компанія зможе пропонувати унікальні персоналізовані пропозиції для своїх клієнтів майже щодня з урахуванням зміни потреб та інтересів у цей час.

### **2.3 Застосування методів машинного навчання для обробки інформації у компанії velokiuv**

Сьогодні машинне навчання, засноване на зборі та аналізі даних, і штучний інтелект, як його логічне продовження, – вже не атрибут фантастичних романів та фільмів, а працююча технологія, яка вирішує все більше завдань у найрізноманітніших галузях.

Машинне навчання вже довело свою життєздатність – воно вирішує все більше завдань у різних областях. У міру розвитку технологій сфера застосування інтелектуальних систем з машинним навчанням зростатимуть – і що далі, то складнішими будуть завдання, які вирішуються за допомогою інтелектуального аналізу даних.

Для даного бізнесу у сфері велосипедного ринку машинне навчання буде застосоване для збору, обробки та створення унікальних персоналізованих комерційних пропозицій як для постійних клієнтів, так і для нових тих, хто здійснив не більше однієї покупки.

При створенні персоналізованих пропозицій компанія має такі цілі:

- інформування клієнтів про компанію (нагадування про себе)
- експрес-інформування про товари (решту клієнт за бажання отримає, перейшовши за посиланням);

- продаж товарів (моментальний, тому що клієнту пропонується така пропозиція, яка враховує його потреби та інтереси);
- розвага споживачів, привернення уваги.

Далі наведемо переваги та недоліки даної моделі роботи компанії з клієнтами у таблиці 3.

Таблиця 3 - Переваги та недоліки

Переваги	Недоліки
Можливість персоналізації пропозицій	Висока вартість
Охоплення виключно цільової аудиторії	Необхідність постійного оновлення
Можливість контролю ефективності	Складна сегментація клієнтів, 1-2 ознака не показова
Миттєвий відгук на пропозицію	Негативний ефект "стеження"

Найважливіше у цій моделі – це отримання найповнішої інформації про кожного клієнта. На основі цієї інформації (попередніх замовлень) буде зроблена персональна пропозиція.

Для створення такої пропозиції було складено список критеріїв для підбору індивідуальної пропозиції для кожного конкретного клієнта («Критеріальна база»). «Критеріальна база» включає шість параметрів сегментування всіх клієнтів компанії. Дана база представлена в таблиці 4 із зазначенням типу та значень параметрів, що приймаються.

В даному випадку будуть використані два методи машинного навчання: кластеризація та пошук асоціативних правил. Так як обидва методи відносяться до методів навчання без вчителя, то для більш правильної роботи алгоритму, необхідно буде провести додаткове навчання, коригування помилок кластеризації клієнтів та їх потреб, а також при побудові асоціативних правил.

Таблиця 4 – Критеріальна база

Параметр	Тип	Значення
Стать	Текстовий	Чоловік/жінка
Зріст	Числової	Від 140 до 185 см (крок – 15 см)
Стан речі	Текстовий	Нове/бв
Сезонність	Текстовий	Літо, Весна, Осінь
Рівень бренду	Текстовий	Дешеві, Середні, Преміум
Лояльність клієнта	Логічний	Постійні/непостійні

На першому етапі використовується метод кластеризації. Будуть згруповані як клієнтів, так і весь товар за подібними характеристиками з допомогою статистичних алгоритмів. Тобто першим кроком буде сегментування купівельної аудиторії залежно від характеристик попередніх покупок для складання унікальної комерційної пропозиції, а в майбутньому і для персоналізації маркетингових кампаній під конкретний сегмент клієнтів.

Розглянемо основні шість алгоритмів класифікації, які найчастіше застосовуються практично.

Метод k-найближчих сусідів (K-Nearest Neighbors) Цей метод працює за допомогою пошуку найкоротшої дистанції між об'єктом, що тестується, і найближчими до нього класифікованими об'єктами з навчального набору. Класифікований об'єкт відноситься до того класу, до якого належить найближчий об'єкт набору. На рисунку 16 графічно представлений цей метод класифікації.

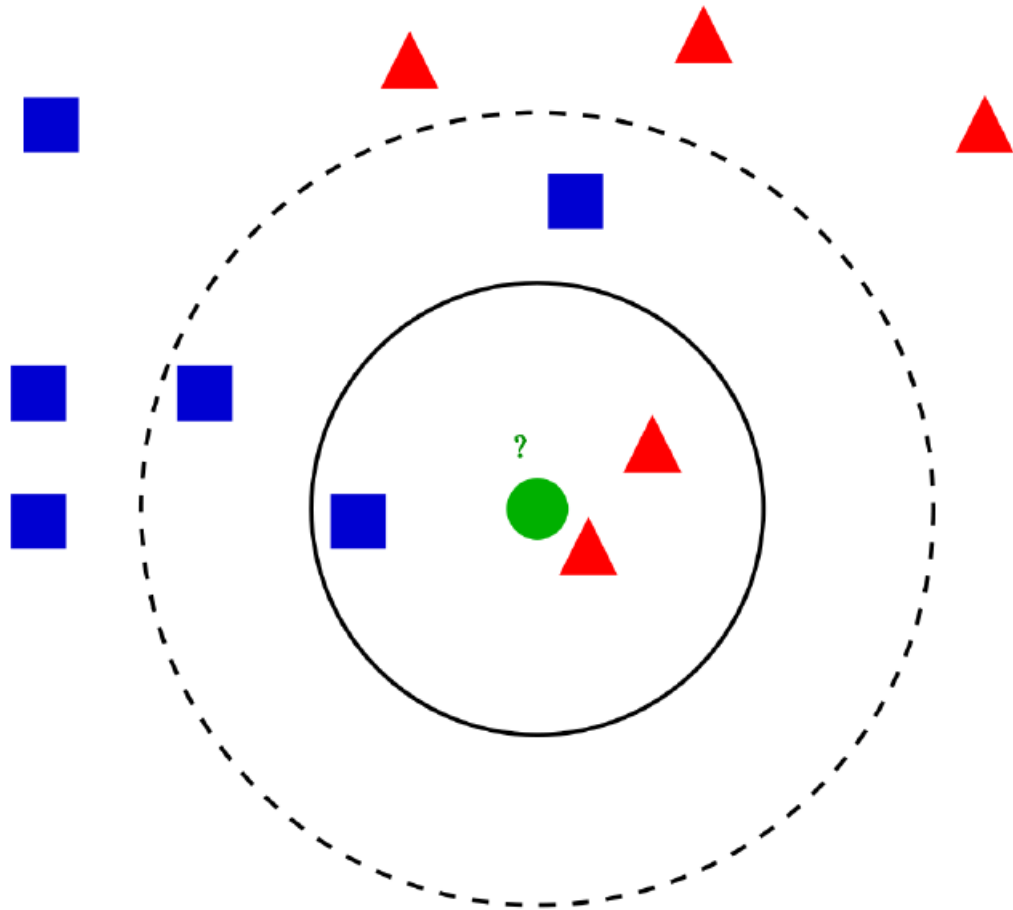


Рис 2.16 Метод k-найближчих сусідів.

Класифікатор дерева рішень (*Decision Tree Classifier*) - цей класифікатор розбиває дані на менші і менші підмножини на основі різних критеріїв, тобто у кожного підмножини своя категорія, що сортує. З кожним розподілом кількість об'єктів певного критерію зменшується. Класифікація підійде до кінця, коли мережа дійде до підмножини лише з одним об'єктом. Якщо поєднати кілька подібних дерев рішень, то вийде так званий Випадковий Ліс (англ. Random Forest).

Наївний байесовський класифікатор (Naive Bayes) - цей класифікатор обчислює можливість приналежності об'єкта до якогось класу. Ця ймовірність обчислюється з шансу, що якась подія відбудеться, з опорою на події, що вже відбулися. Кожен параметр об'єкта, що класифікується, вважається незалежним від інших параметрів

Лінійний дискримінантний аналіз (*Linear Discriminant Analysis*) - цей метод працює шляхом зменшення розмірності набору даних, проектуючи всі точки даних на лінію. Потім він комбінує ці точки у класи, базуючись з їхньої відстані від центральної точки. Цей метод, як можна вже здогадатися, відноситься до лінійних алгоритмів класифікації, тобто добре підходить для даних з лінійною залежністю.

Метод опорних векторів (*Support Vector Machines*) - робота методу опорних векторів полягає у малюванні лінії між різними кластерами точок, які потрібно згрупувати у класи. З одного боку лінії будуть точки, що належать до одного класу, з іншого боку — до іншого класу.

Класифікатор намагатиметься збільшити відстань між лініями, що рисуються, і точками на різних сторонах, щоб збільшити свою «впевненість» визначення класу. Коли всі точки побудовані, сторона, яку вони падають — це клас, якому ці точки належать. На рисунку 17 графічно представлений цей спосіб класифікації.

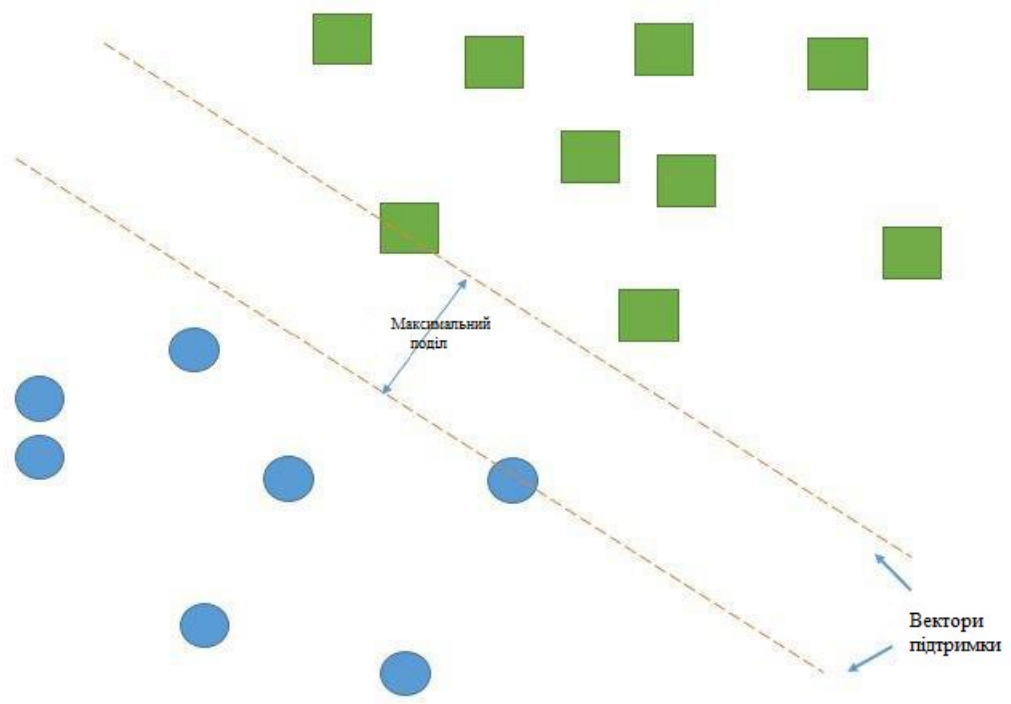


Рис 2.17 Метод опорних векторів.

Логістична регресія виводитиме прогнози щодо точок у бінарному масштабі — нульовому чи одиничному. Якщо значення чогось дорівнює чи більше 0.5, то об'єкт класифікується у бік (до одиниці). Якщо значення менше 0.5 – у меншу (на нуль).

Кожна ознака має свою мітку, рівну лише 0 або лише 1. Логістична регресія є лінійним класифікатором і тому використовується, коли в даних простежується якась лінійна залежність.

Так, залежно від завдання класифікації, необхідно використовувати різні типи класифікаторів. Наприклад, якщо класифікація містить якусь бінарну логіку, то в цьому випадку найкраще підійде логістична регресія. У міру накопичення досвіду простіше вибирати відповідний тип класифікатора. Проте гарною практикою є реалізація кількох відповідних класифікаторів та вибір найбільш оптимального та продуктивного.

Отримуючи кластерну структуру, необхідно бути впевненим у тому, що ця структура є коректною, відповідає всім поставленим вимогам і підходить для подальшого використання.

Основне завдання кластеризації зазвичай формулюється так: розділити об'єкти на групи таким чином, щоб подібність між об'єктами однієї групи була великою, а подібність між об'єктами різних груп – мало. Виходячи з цього, поняття якості кластеризації складається з наступних пунктів:

- **Компактність:** елементи кластера повинні бути якомога ближче один до одного. Це властивість можна висловити через відстань між елементами в кластері, щільністю всередині кластера або об'єктом, що займається кластером у багатовимірному просторі.
- **Відділення:** відстань між різними кластерами має бути, як можна. Це застосовується як чіткої, так нечіткою кластеризації. У першому випадку відстань між кластерами зазвичай вимірюється одним із трьох наступних способів: як відстань між найближчими елементами кластерів; як відстань між найбільш віддаленими

один від одного елементами кластерів; як відстань між кластерними центрами.

- Концентрація: елементи кластера повинні бути сконцентровані навколо центру кластеру.

На другому кроці використовується метод пошуку асоціативних правил. Будуть побудовані відносини між змінними (клієнтами та товарами). Вивчивши купівельну поведінку шляхом пошуку асоціативних правил, можна пропонувати спеціальні товари такі, як сезонні, які не використовуються цілий рік. Асоціативні правила дозволять з'ясувати, коли і за яких обставин покупці набували раніше ті чи інші поєднання товарів. Використовуючи відомості про минулі покупки та час їх здійснення, можна скласти програму знижок та сформувані індивідуальні пропозиції з розрахунком на підвищення продажів.

Третім кроком буде аналіз отриманих кластерів та коригування їх за допомогою додаткового навчання.

Після проведення всіх перерахованих вище етапів будуть сформовані пропозиції для кожного клієнта. Дані кожного клієнта на певну дату будуть зберігатися і використовуватися в таблицях Excel, як представлено на рисунку 18.

	A	B	C	D	E	F	G	H	I	J	K
1											
2	id клієнта	Прізвище	Ім'я	По батькові	Email	Товар 1	Товар 2	Товар 3	Товар 4	Товар 5	Товар 6
3	1	Максименко	Руслан	Сергійович	<a href="mailto:alan_waik.1998@gmail.com">alan_waik.1998@gmail.com</a>	228322	133221	215562	219321	412911	826123
4	2	Сасько	Ілля	Охримович	<a href="mailto:lecruddassannei-3603@yopmail.com">lecruddassannei-3603@yopmail.com</a>	21213	121287	111747	162995	106409	121287
5	3	Загайкевич	Єрмей	Найденович	<a href="mailto:WinstonHoward@gmail.com">WinstonHoward@gmail.com</a>	231222	126278	112199	285624	109482	126278
6	4	Коцур	Олександр	Августиневич	<a href="mailto:QuangWilson@gmail.com">QuangWilson@gmail.com</a>	121341	169802	208050	292428	124077	169802
7	5	Погорелов	Явір	Леонідович	<a href="mailto:WellsCook@gmail.com">WellsCook@gmail.com</a>	588682	309603	229674	318264	210194	309603
8	6	Плющ	Троян	Антонович	<a href="mailto:OrlandoNelson@gmail.com">OrlandoNelson@gmail.com</a>	152182	331856	247390	322159	284092	331856
9	7	Вдовенко	Флор	Чеславович	<a href="mailto:HassanFlores@gmail.com">HassanFlores@gmail.com</a>	107108	374168	299388	421288	337664	374168
10	8	Багмут	Ігор	Мстиславович	<a href="mailto:YoshinobuRivera@gmail.com">YoshinobuRivera@gmail.com</a>	164029	376241	404416	635716	338648	376241
11	9	Ковалик	Русан	Жданович	<a href="mailto:NormanWood@gmail.com">NormanWood@gmail.com</a>	189795	402723	429763	658919	370147	402723
12	10	Фастенко	Євген	Любомирович	<a href="mailto:NashRichardson@gmail.com">NashRichardson@gmail.com</a>	234614	455497	435604	665404	513831	455497
13	11	Сітько	Мирослав	Орестович	<a href="mailto:RobertoSmith@gmail.com">RobertoSmith@gmail.com</a>	242652	511892	442407	674828	598354	511892
14	12	Мандзюк	Божен	Арсенович	<a href="mailto:ZevWilson@gmail.com">ZevWilson@gmail.com</a>	281306	647121	457720	708803	646996	647121
15	13	Левчук	Царук	Драганович	<a href="mailto:IsaiasGarcia@gmail.com">IsaiasGarcia@gmail.com</a>	291026	656977	651970	752913	652122	656977
16	14	Супруненко	Радован	Юхимович	<a href="mailto:OlinHoward@gmail.com">OlinHoward@gmail.com</a>	400254	682095	817212	801497	717835	682095
17	15	Білоножченко	Милодуч	Макарович	<a href="mailto:UlysesMurphy@gmail.com">UlysesMurphy@gmail.com</a>	444166	724118	881204	824710	718463	724118
18	16	Гребінь	Єрмей	Сарматович	<a href="mailto:ZanderSanchez@gmail.com">ZanderSanchez@gmail.com</a>	451653	728541	886461	830538	724990	728541
19	17	Онщенко	Ведан	Охримович	<a href="mailto:EmersonNelson@gmail.com">EmersonNelson@gmail.com</a>	471237	774929	892443	832865	799163	774929
20	18	Матяш	Афанасій	Азарович	<a href="mailto:YosefWard@gmail.com">YosefWard@gmail.com</a>	506532	818908	895434	844678	881774	818908
21	19	Шаповал	Йошка	Адріанович	<a href="mailto:LandynMartin@gmail.com">LandynMartin@gmail.com</a>	650643	915295	900058	868029	927063	915295
22	20	Капуста	Устим	Сарматович	<a href="mailto:IzaacHayes@gmail.com">IzaacHayes@gmail.com</a>	691760	929072	935447	879112	960882	929072
23	21	Гринько	Божен	Сарматович	<a href="mailto:FloydWatson@gmail.com">FloydWatson@gmail.com</a>	834784	964735	950250	978179	980587	964735

Рис 2.18 Дані за пропозиціями в Excel.



Процес формування пропозицій та розсилки даних пропозицій на електронну пошту клієнтів проходитиме у напівавтоматичному режимі через картку клієнта в бек-офісі за натисканням кнопки «Надіслати листа». Для цього необхідно всі запропоновані товари (їх буде 6 шт.) перенести через кому без прогалін у полі «Запропонувати речі», а потім натиснути кнопку «Надіслати листа». Це вікно представлено рисунку 19.

Клієнт № 322

Завантажити договір   Створити товарну накладку   Створити оголошення   Відправити лист   Запропонувати товар

Велосипедист

Сегменти

- A1 (бв бренд)
- A2 (бв дешеві)
- A3 (тільки нові)
- A4 (токовсі бв)
- A5 (пагенти)
- A6 (токовсі лев)
- Z1 (брандові)
- Z2 (пагалині)
- Z3 (дешеві)
- Z4 (браншові торгуються)
- Z5 (токовсі)

Тариф: Клієнти 4

Дата створення: 03.12.2021

Код для входу в особистий кабінет: 1211-2121-9211-0910 [Перегенерувати](#) [Надіслати клієнтові на e-mail](#) Останній раз надсилався: 3.12.2021 13:51:21

Прізвище клієнта: Максименко

Ім'я клієнта: Руслан

По батькові клієнта: Сергійович

Телефон: +380958780373

Місто: Київ

Адреса: Київ, Волгоградская 18

Email: alan.waik.1998@gmail.com

Рис 2.19 Картка клієнта

Пропозиції надсилаються на електронну пошту (це поле є обов'язковим для кожного клієнта) через сервіс e-mail розсилки. Макет цього листа представлений рисунку 20.

Головне завдання листа – це привернути увагу клієнта до цікавої пропозиції, яка підібрана індивідуально для нього. У листі має бути таке:

- ім'я клієнта;
- промокод на додаткову знижку на ці товари протягом 7 днів з моменту сформування пропозиції;
- 6 спеціально підібраних товарів з картинкою, назвою товару та ціною;

- посилання на повний опис товару на сайті (під кожним товаром);
- інформація щодо способів доставки по Києву та по Україні;
- контактна інформація.

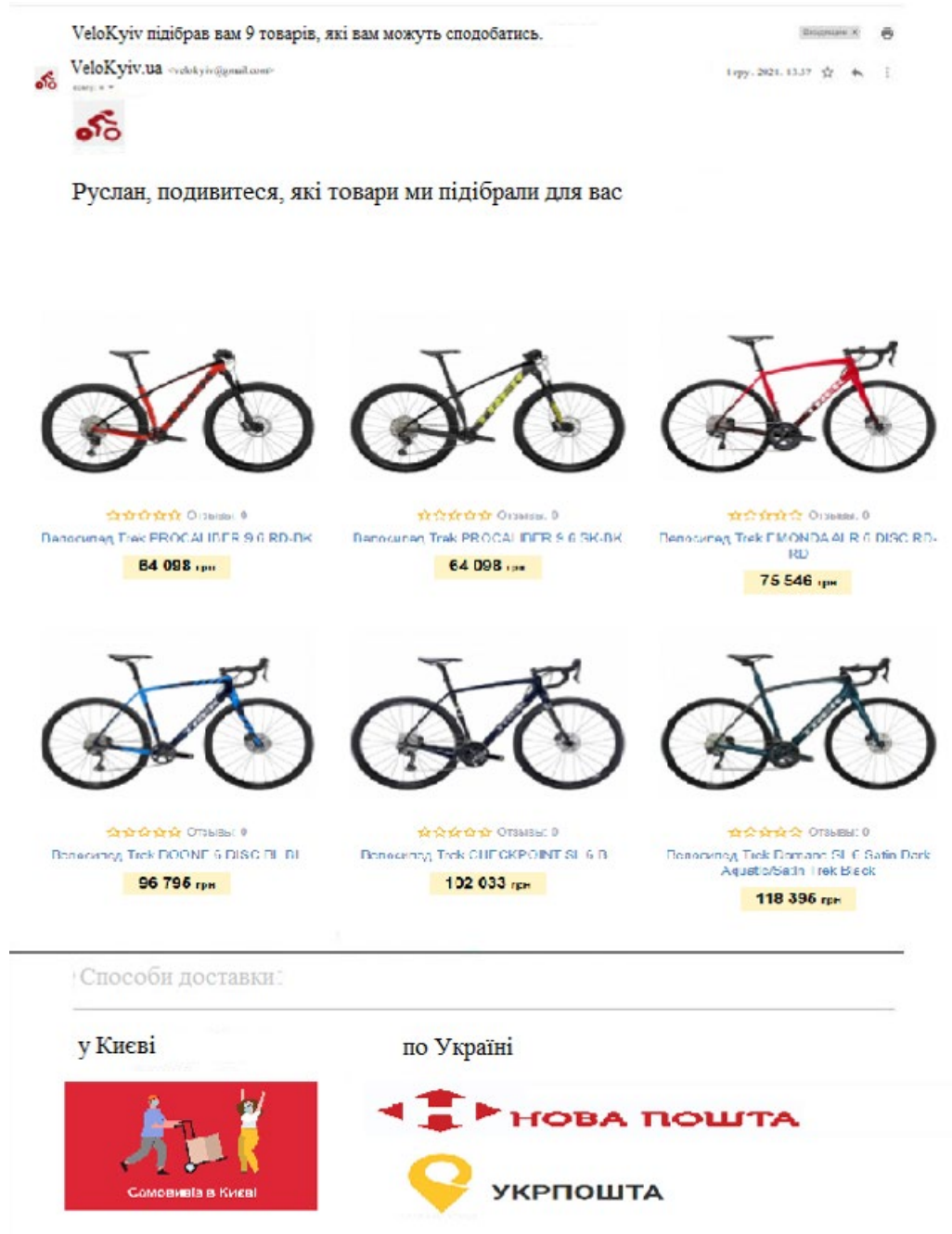


Рис 2.20 Макет листа е-маїл розсилки

Наступним кроком роботи з даною моделлю буде аналіз ефективності комерційних пропозицій (рис 2.21). Величезний плюс даного способу залучення клієнтів полягає у чотирьох основних моментах:

- швидкість отримання зворотного зв'язку;
- відстеження досяжності пропозиції до користувача (сервіс з email розсилки дозволяє відслідковувати статус листів: лист отриманий, лист відкритий, лист не був відкритий, перехід за посиланням з листа);
- чіткий вимірний результат (конверсія);
- аналітика під час роботи з сайтом (за якими посиланнями були здійснені переходи на сайт).

Таким чином, кожна аналітика відкриває багато корисної та унікальної інформації, яку може надати та відкрити нові можливості для роботи з цільовою аудиторією. Чим більше даних, тим менша ймовірність невдачі, тобто існує чітка необхідність постійної актуалізації інформації за своєю аудиторією, проведення А/В тестування з пропозиціями або макетами e-mail розсилок з паралельним відстеженням та постійним коригуванням за допомогою засобів статистики, а також важливо вміти за допомогою аналітики відстежувати кожну конверсію.

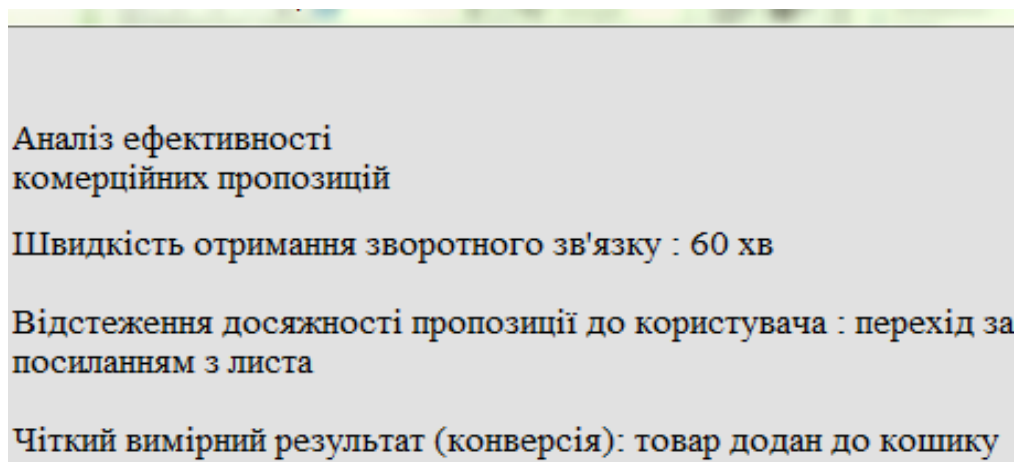


Рис 2.21 Аналіз ефективності комерційних пропозицій

## 2.4 Оцінка конкурентоспроможності підприємства за допомогою методів машинного навчання

Аналіз та дослідження поняття «конкурентоспроможність» призводять до висновку, що поки що не склалося загальноприйнятого визначення цього терміна. Будь-який аналітик звертає увагу різні аспекти конкуренції, розглядає різні трактування даного поняття. Усе це призводить до наявності різноманітних визначень однієї й тієї ж поняття. У цілому нині конкурентоспроможність розуміють, передусім, як здатність конкурувати, утримувати позиції над ринком, як властивість суб'єктів результатів конкурентних відносин. Оцінюючи класифікації конкурентоспроможності підприємства на рівні країн, галузей, підприємств, продукції, слід врахувати, що країни, галузі, товари не конкурують власними силами, в основі лежить діяльність підприємств з їх економічними інтересами.

Індекс глобальної конкурентоспроможності складено з великої кількості змінних, які детально характеризують класифікації конкурентоспроможності країн світу, що знаходяться на різних рівнях економічного розвитку. Усі змінні об'єднані у 12 контрольних показників, що визначають національну класифікацію конкурентоспроможності, у тому числі якість інститутів, інфраструктуру, макроекономічну стабільність, здоров'я та освіту, ефективність ринку товарів та послуг тощо.

Зовсім інакша справа з класифікацією конкурентоспроможності підприємств. Велика кількість визначень цього терміну класифікації конкурентоспроможності, роз'яснень його сутності неможливо однозначно виявити фактори, що мають найбільший вплив на класифікацію конкурентоспроможності підприємств. Наприклад, серед факторів класифікації конкурентоспроможності, контрольованих організацією,

зазначимо такі: рівень якості продукції; передпродажне та після продажне обслуговування, наявність сервісних центрів; кваліфікація персоналу підприємства; імідж компанії тощо.

Машинне навчання використовує теорію статистики під час побудови математичних моделей, оскільки основним завданням є формування виведення з вибірки даних. Роль комп'ютерних наук двоїста. По-перше, у навчанні потрібні ефективні алгоритми для вирішення задачі оптимізації, а також для зберігання та обробки величезної кількості даних, які зазвичай є. По-друге, як тільки модель вивчена, її подання та алгоритмічне рішення для виведення також мають бути ефективними. У деяких додатках ефективність алгоритму навчання або логічного висновку, а саме його просторова та тимчасова складність, може бути такою ж важливою, як і його точність прогнозування.

Завдяки нейроним мережам можна проаналізувати можливості формування кількісної оцінки конкурентоспроможності підприємств.

Як фактори, що впливають на класифікації конкурентоспроможності підприємства, приймемо такі:

- якість продукції;
- товарно-збутова діяльність підприємства;
- організаційно-технічна досконалість виробництва;
- фінансовий стан;
- імідж підприємства.

Такий набір факторів, на думку автора, є достатнім для цілей дослідження, оскільки він охоплює основні аспекти класифікації конкурентоспроможності підприємства та не суперечить загальноприйнятим тенденціям

Класи класифікації конкурентоспроможності підприємства встановимо такими

- правильний – результати конкретних взаємодій збігаються з очікуваннями зацікавлених сторін;
- нижче за правильний – фактичні результати не відповідають очікуванням. Потім зацікавлені сторони, залучені до цієї ситуації, зроблять дії, щоб відмовитися від взаємодії з компанією та перейти до іншої, більш привабливої;
- вище за правильний – фактичні результати вищі, ніж очіувалося. Зацікавлені сторони, які мають підстави для таких оцінок, прагнуть зміцнити свої відносини з компанією.

Для створення бази прикладів, що використовується в машинному навчанні для формування моделі, можливе використання двох підходів:

- Застосування реальних даних;
- Використання «іграшкових» наборів даних.

У дослідженні реальних даних ще немає, тому звернемося до другого прийому. У машинному навчанні важливо навчитися правильно застосовувати іграшкові набори даних, оскільки навчання алгоритму реальних даних пов'язані з труднощами може закінчитися невдачею. Іграшкові набори даних грають вирішальну роль розуміння роботи алгоритмів. За наявності простої синтетичної вибірки даних досить просто оцінити, навчився алгоритм потрібне правило чи ні. На реальних даних отримати таку оцінку складно. Тут скористаємося методом Монте-Карло на формування синтетичної вибірки даних. При розіграві приймемо для кожного фактора діапазон від 0 до 10 балів з урахуванням класів та доповнимо результати моделювання випадковим шумом. Кожен розігравий клас має по 10 рядків, і вся таблиця є матрицею розміром 30\*5. Фрагмент розіграних даних показано у таблиці 5.

Таблиця 5 Фрагмент розіграних даних

X1	X2	X3	X4	X5	Клас
2,8	2,9	0,4	2,3	2,5	Менше
3	1	3,9	3,7	3,7	Менше
4,5	1,4	0,9	0,8	3,6	Менше
3,8	0,7	1,9	3,2	0,8	Менше
2,2	1,4	4,5	3,3	1	Менше

На рисунку 2.22 у площині перших двох змінних (якість продукції та товарно-збутова діяльність) представлені всі 30 змодельованих об'єктів, кожен з яких характеризується певною величиною класифікації конкурентоспроможності підприємства.

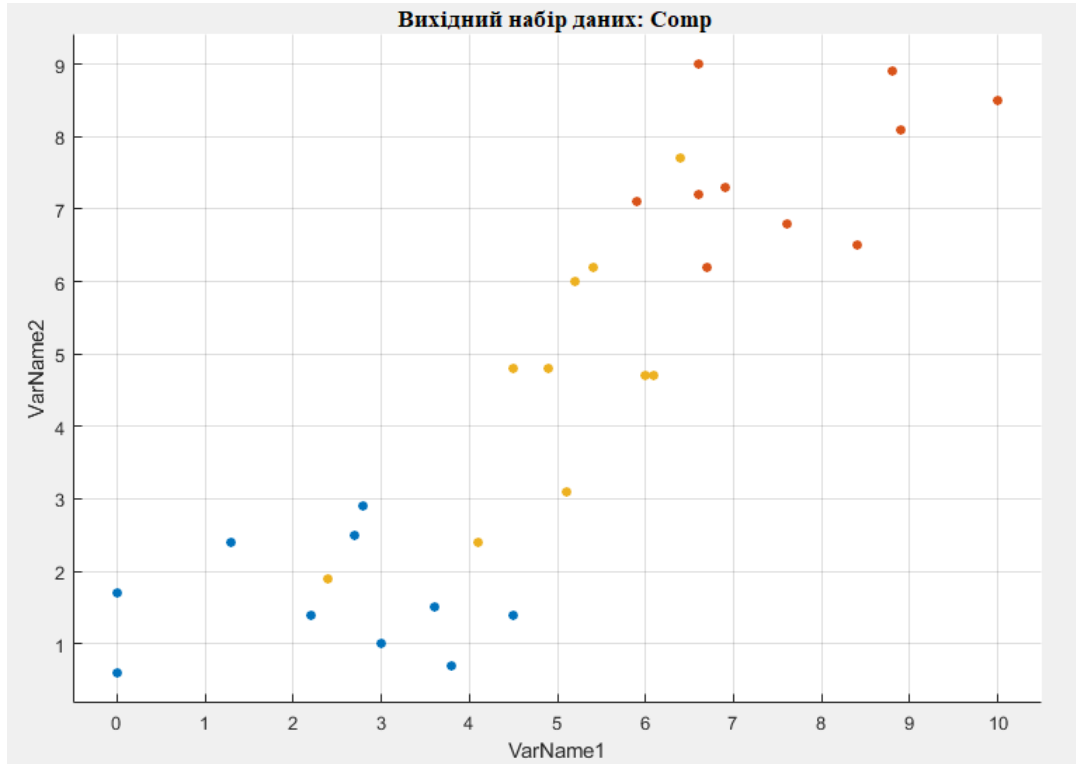


Рисунок 2.22 Вихідні дані

Далі на основі такої таблиці необхідно проаналізувати різні методи класифікації та вибрати той, який покаже найкращу точність. Для вирішення цього завдання скористаємося програмним продуктом MatLab . У цієї програми є спеціальний модуль класифікація учня, який відкриває більш широкі можливості побудови класифікаторів, порівняно з попередніми версіями. Цей модуль може працювати в ручному режимі , коли по черзі перевіряється той чи інший метод, та автоматичному , коли програма переглядає всі можливі розв'язки цієї задачі та вказує найкращі результати. Нагадаємо, що методи машинного навчання на основі наявних даних формують модель, яка може використовуватись для класифікації нових спостережень. Тут використовується автоматичний режим, в якому виберіть опцію “усі швидко тренуються”. Ця опція для нашого набору даних навчить усі найшвидші моделі.

Результати цієї процедури показано на рисунку 2.23, де серед 9 проаналізованих методів найвищу точність продемонстрував середній метод найближчих сусідів.

▼ History		
1.1	☆ Tree Last change: Fine Tree	Accuracy: 70.0% 5/5 features
1.2	☆ Tree Last change: Medium Tree	Accuracy: 70.0% 5/5 features
1.3	☆ Tree Last change: Coarse Tree	Accuracy: 70.0% 5/5 features
1.4	☆ KNN Last change: Fine KNN	Accuracy: 83.3% 5/5 features
1.5	☆ KNN Last change: Medium KNN	Accuracy: 90.0% 5/5 features
1.6	☆ KNN Last change: Coarse KNN	Accuracy: 33.3% 5/5 features
1.7	☆ KNN Last change: Cosine KNN	Accuracy: 66.7% 5/5 features
1.8	☆ KNN Last change: Cubic KNN	Accuracy: 86.7% 5/5 features
1.9	☆ KNN	Accuracy: 86.7%

Рисунок 2.23. Результати підбору моделі



Наведемо результати для обраного програмою методу. На рисунку 2.24 представлена матриця помилок методу найближчих сусідів. Рядками матриці помилок є справжні класи, стовпцями – передбачені. Діагональні елементи показують, де істинний та передбачений класи збігаються. Сума всіх чисел у осередках матриці становить 30, тобто. дорівнює кількості об'єктів. Як видно з матриці, навчений класифікатор помилявся на об'єктах класу правильний: переплутав його із класом більше (1 раз) та класом менше (2 рази).

		Model 1.5		
		Менше	Більше	Правильний
Справжній клас	Менше	10		
	Більше		10	
	Правильний	2	1	7
		Передбачені класи		

Рисунок 2.24 Матриця помилок.

На рисунку 2.25 показано матрицю помилок з додаванням "Долі вірних позитивних класифікацій" і "Долі хибних негативних класифікацій" розміщених у правій частині рисунка.

В останніх двох стовпцях праворуч показані зведені дані щодо кожного класу. Наприклад, у третьому рядку для класу правильний стоять значення 70% та 30%, що відповідають "Долі вірних позитивних класифікацій" та "Долі хибних негативних класифікацій". Іншими словами, для цього класу частка правильно розкласифікованих об'єктів становить 70%, а частка хибних – 30%. Також видно, що об'єкти першого та другого класів розділені безпомилково: для них значення долі вірних позитивних класифікацій = 100%.

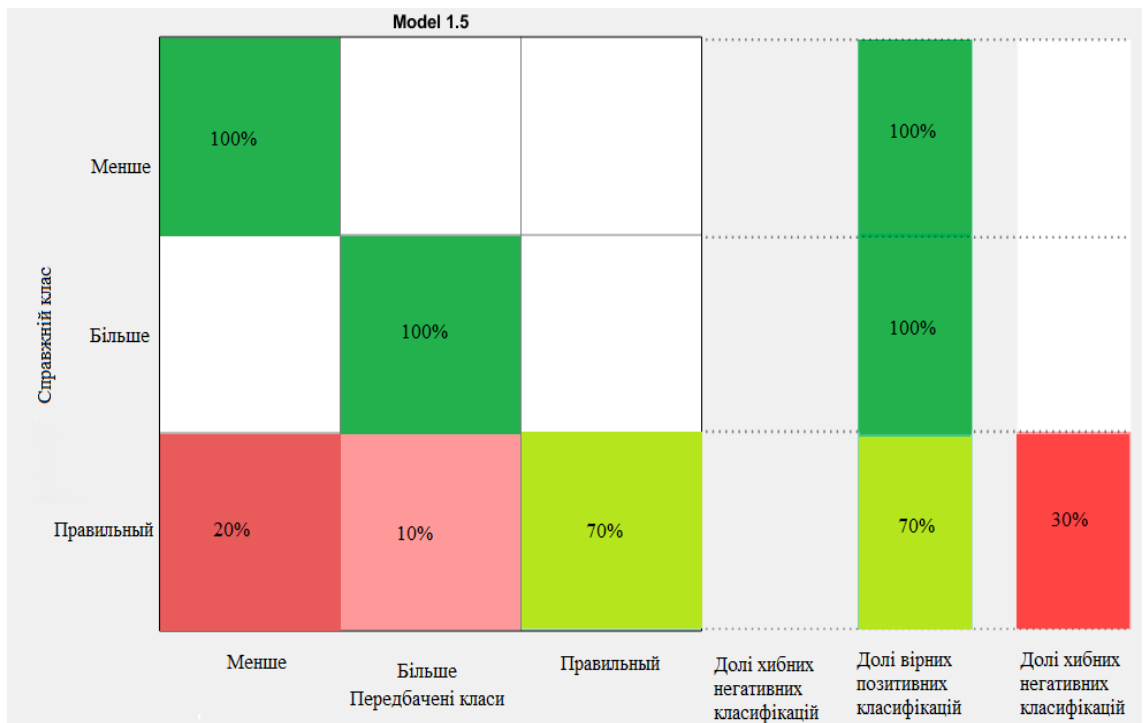


Рисунок 2.25. Позитивні та негативні класифікації

На рисунку 2.26 показана крива операційної характеристики (робоча характеристика приймача), яка показує частку вірних позитивних класифікацій у функції частки хибних позитивних класифікацій для обраного класифікатора.

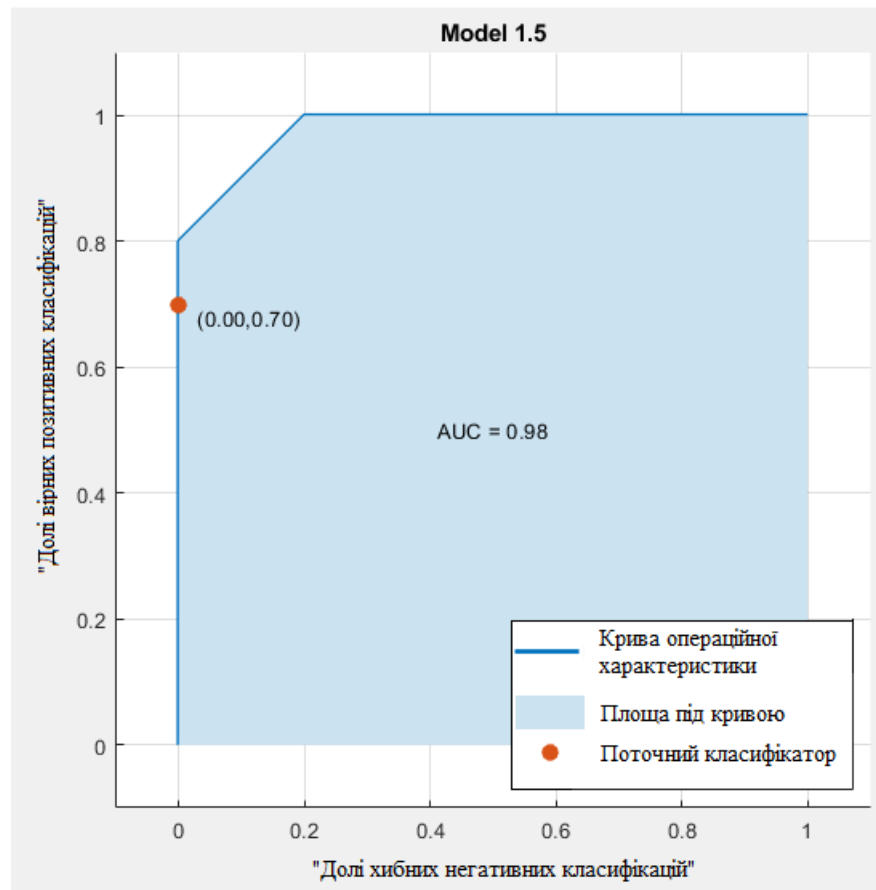


Рисунок 2.26. Крива операційної характеристики

На наведеному графіку ідеальний результат безпомилкової класифікації - це лінія, що утворює прямий кут у лівому верхньому куті графіка. Поганий результат, який не кращий за випадковий, відображається лінією, що проходить під кутом 45 градусів з лівого нижнього кута графіка в правий верхній. Число, що характеризує площу під кривою, є показник загальної якості класифікатора. Чим більше значення площини під кривою, тим краща якість класифікатора.

Для передбачення класу нового спостереження необхідно після створення моделі експортувати її у робочу область. У результаті формується структура `trainModel`, яка використовується для прогнозування з використанням нових даних. Як приклад нижче наведено результат класифікації нового вектора  $mj1 = (2.8 \ 2.9 \ 0.4 \ 2.3 \ 2.5)$ :

- `>> yfit = trainedModel.predictFcn (mj1)`
- `yfit =`
- `categorical`
- `Less`

Таким чином, обраний метод метод найближчих сусідів відніс класифікацію конкурентоспроможності підприємства даного об'єкта до класу, який позначений як «нижче за правильну класифікацію конкурентоспроможності підприємства».

Звернемося до нейронної мережі для класифікації конкурентоспроможності підприємства розіграних об'єктів. Хоча нейронна мережа не обрана в автоматичному режимі як прийнятний класифікатор (була обрана опція “Усі швидко тренуються”, націлена перш за все на швидкість обчислень). Створена у програмі MatLab двошарова нейронна мережа з обраною кількістю нейронів, що відповідають задачі, наведено на рисунку 2.27.

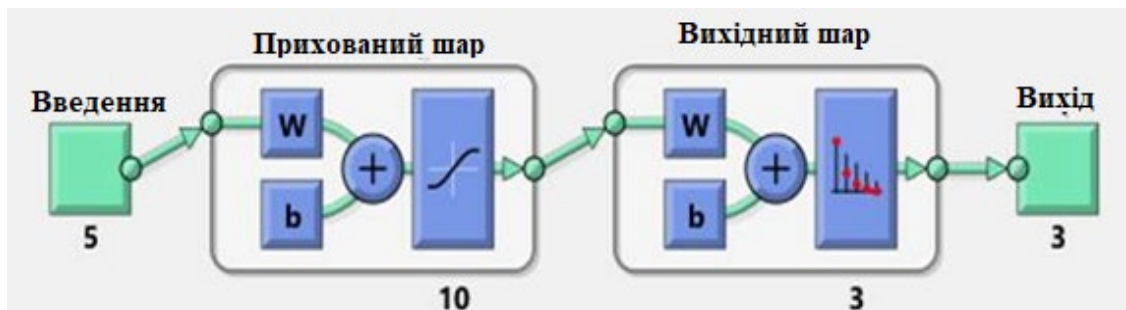


Рисунок 2.27. Нейронна мережа

Вихідна сукупність даних з 30 об'єктів ділиться на 3 частини у пропорції 70%, 15% та 15%:

- навчальна вибірка (20 об'єктів), що використовується для навчання мережі;
- валідаційна вибірка (5 об'єктів), що застосовується для оцінювання узагальнюючої здатності мережі та зупинення

процесу навчання у разі припинення покращення узагальнення;

- Тестова вибірка (5 об'єктів), призначена для незалежної міри характеристики мережі під час та після навчання.

Частина результатів навчання представлена нижче. На рисунку 2.28 наведено три криві навчання, що відповідають навчальній, валідаційній та тестовій вибіркам.

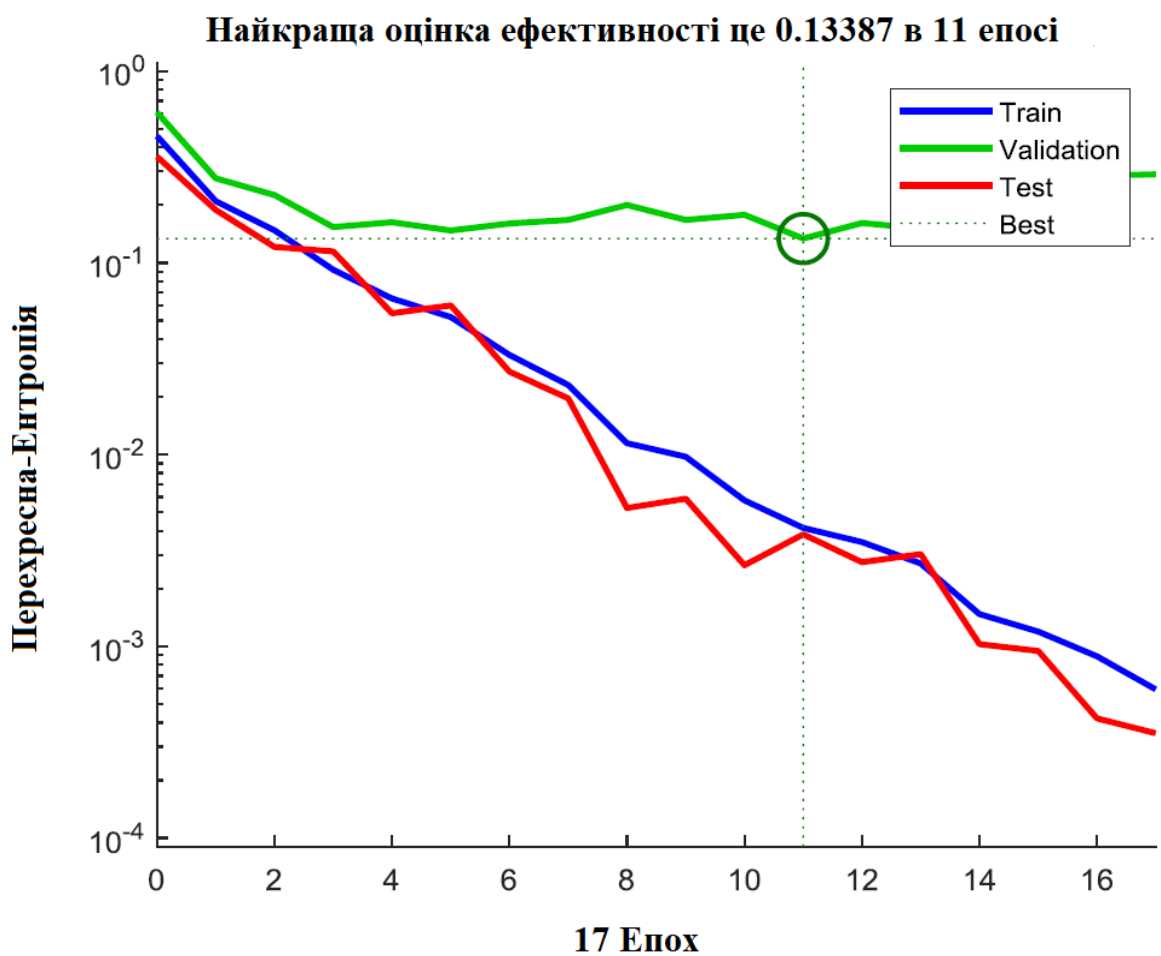


Рисунок 2.28. Криві навчання мережі

На вертикальній осі графіка рисунка 8 відкладено величину крос-ентропії, яка оцінює поведінку мережі з урахуванням цільових та вихідних

значень. Мінімізація крос-ентропії призводить до класифікаторів хорошої якості. На горизонтальній осі зазначено кількість епох навчання, звідки видно, що на 11-й епосі відбулося погіршення узагальнення мережі, тому мережа навчилася за цей час.

Для визначення приналежності підприємства-об'єкта з конкретного класу з 3 виділених збережемо нейронну мережу в робочому просторі MatLab, а потім подаємо на її вхід через командне вікно параметри об'єкта, наприклад,  $mj3 = [5.5; 6.5; 5; 5; 6]$ .

Мережа видає у вигляді вектора стовпця, найбільше значення якого визначає клас:

- `>> fit = net (mj3)`
- `fit =`
- 0.0001
- 0.9990
- 0.0009.

Таким чином, за допомогою нейронної мережі для конкретного прикладу отримали відповідь щодо належності об'єкта з такими параметрами до другого класу (у наших позначеннях – правильний).

### **3 МЕТОДОЛОГІЯ ВПРОВАДЖЕННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ АВТОМАТИЗАЦІЇ АНАЛІТИЧНОЇ ДІЯЛЬНОСТІ НА РИНКУ ВЕЛОСИПЕДНОГО РІТЕЙЛУ**

#### **3.1 Розробка методики впровадження методів машинного навчання для автоматизації аналітичної діяльності**

Значення машинного навчання для бізнесу стало дуже велике, розумні алгоритми замінюють людську працю в багатьох бізнес-процесах, тим самим скорочуючи витрати, зменшуючи витрати часу, сил, людино-годин. Перерахуємо переваги застосування машинного навчання у бізнесі:

- більший обсяг даних обробляється за менший час;
- висока швидкість часу реагування;
- процеси автоматизуються та не вимагають постійного втручання з боку людини. Більше того, чим довше машина працює над конкретним завданням, тим успішніше стають її рішення і тим вища конверсія;
- машина враховує безліч факторів, на основі яких приймає рішення;
- технології машинного навчання – це гнучкий інструмент, який підійде для вирішення проблем конкретного бізнесу, при цьому модель працює індивідуально для кожного випадку;
- використання машинного навчання скорочує витрати на персонал та на залучення клієнтів .

Найпоширеніші типи проектів з урахуванням технологій машинного навчання, націлених отримання підприємством додаткової виручки чи скорочення витрат такі:

- Підвищення продуктивності технологічного процесу за рахунок підбору оптимальних режимів роботи обладнання, завантажень сировини;

- підвищення якості продукції шляхом виявлення критичних факторів у виробничому процесі, що впливають на кінцевий результат;
- оптимізація технологічного обслуговування та ремонту (ТОiP) дорогого виробничого обладнання, прогноз поломок та зносу обладнання;
- управління ціноутворенням та ланцюжками поставок — оптимізація та прогнозування щодо процесів закупівель, доставки, зберігання, попиту та пропозиції;
- комплексне поліпшення виробничих показників за рахунок виявлення прихованих факторів, що впливають на виробничі процеси, та застосування моделювання ситуацій у цифрових середовищах.

На старті будь-якого проекту дуже важливо закласти фундамент під масштабування проекту у майбутньому. Це, звичайно, сильно збільшує терміни випуску першої робочої версії продукту з мінімальним функціоналом (mvp), але надалі буде можливість безболісно для компанії нарощувати функціональність проекту, а також розгортати його на інших майданчиках.

Даний проект можна застосувати на будь-якому інтернет-магазині, починаючи від маленького бізнесу, до великих лідерів в інтернет-торгівлі. Проект був розроблений з урахуванням його можливого масштабування, зі збільшенням обсягів даних завдання стоїть лише збільшення обчислювальних потужностей.

Розглянемо шість етапів впровадження технології машинного навчання у бізнесі:

1. Постановка задачі.

Будь-яка технологія має вирішувати конкретні завдання бізнесу. Для успішного впровадження необхідно визначити, які КРІ необхідно покращити,



як і за якими метриками вимірюється отриманий результат, досягнуто мети роботи чи ні.

## 2. Збір, зберігання та попередня обробка даних.

Коли завдання поставлено, необхідно створити навчальну вибірку (на жаль, більшість бізнес-завдань вирішуються шляхом «навчання з учителем», тобто швидкість впровадження буде нижчою, але підвищується якість отриманих даних). Формування вибірки є найтривалішим етапом. Для його скорочення в компанії має бути налагоджена культура роботи з даними. Крім збору даних, необхідно їх очистити та визначити особливості, що впливають на підсумковий результат.

## 3. Навчання алгоритму.

Розробка алгоритмічної частини найцікавіший, але й не найшвидший етап. На нього зазвичай йде від кількох годин до кількох тижнів роботи.

## 4. Розробка високорівневої обв'язки.

Рішення має бути зрозумілим не тільки фахівцю в аналізі даних, але й програмісту або адміністратору, який впроваджуватиме це рішення.

## 5. Інтеграція.

Як правило, займає багато часу через необхідність додаткових комунікацій та погоджень. Цей етап краще виконувати внутрішніми силами шаблону замовника.

## 6. Збір зворотного зв'язку, коригування моделі.

Світ постійно змінюється, в повному обсязі особливості можна враховувати на початку розробки. Збір зворотного зв'язку допомагає своєчасно перенавчати моделі. В ідеалі цьому етапі цикл запускається заново, але з меншими тимчасовими витратами.

Успіх проектів машинного навчання визначається безліччю факторів: злагодженістю та досвідом команди, розумінням даних та технологічних процесів, оптимальним розподілом ресурсів у ході проекту. Чітке проходження бізнес-мети проекту машинного навчання дозволяє

досліджувати дані, будувати гіпотези та перевіряти їх, обравши оптимальні прототипи моделей та спланувавши поетапне тестування. Такий підхід дозволяє заощадити ресурси та на ранніх етапах виключити проекти, що не приносять економічної вигоди.

Технології машинного навчання підвищують ефективність бізнесу, але треба пам'ятати, щоб зробити закінчене рішення недостатньо лише навчити алгоритм, а необхідно підготувати дані та інтегрувати рішення з внутрішніми системами. І пам'ятати, що результат залежний від якості навчальної вибірки.

### **3.2 Конверсія продажів**

Конверсія сайту – це один із показників ефективності, який вимірюється у відсотках. Коефіцієнт конверсії дорівнює відношенню покупців до відвідувачів сайту, помноженого на 100%. Прийнято вважати конверсію для кожного етапу воронки та джерела трафіку окремо.

Слід зазначити, що загальноприйнятої середньої конверсії сайту немає жодної ніші. Кожен власник сайту визначає для себе мінімальний рівень конверсії, щоб швидко зорієнтуватися та виправити помилки при збої (наприклад, не завантажується каталог). Для цього намагайтеся відслідковувати конверсію щонайменше один раз на тиждень.

По цільовій дії розрізняють конверсію на купівлю (зробили замовлення), додавання товару до кошику та коли відвідувач відкрив товар.

Після удосконалення інформаційної моделі, а саме удосконалення обробки даних при формуванні цільової аудиторії компанії та удосконалення аналізу ефективності комерційної пропозицій, можна побачити значне зростання конверсії продажів, його можна побачити на рисунку 3.1

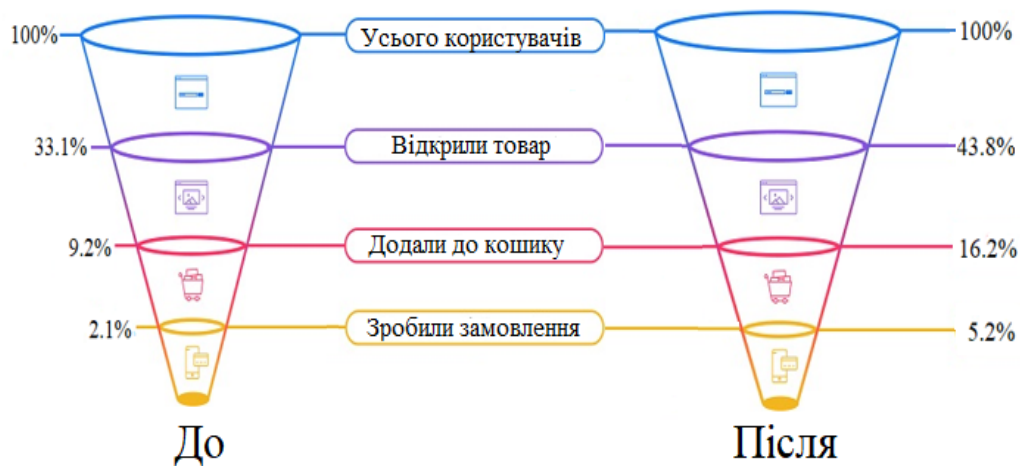


Рис 3.1 Конверсія продажів до та після вдосконалення інформаційної системи

Конверсія розраховується за формулою (3.1)

$$\text{Конверсія} = \frac{\text{Кількість покупок}}{\text{Кількість відвідувачів}} \times 100\% \quad 3.1$$

## ВИСНОВОК

Сфера велосипедного ритейлу на сьогоднішній день є однією з найбільш стабільних галузей, зважаючи на широкий попит на велосипеди та аксесуарів до нього .

Дедалі більше людей звертається до велосипедного інтернет-магазину, намагаючись, з одного боку, заощадити час та гроші при покупці нового велосипеда , а з іншого, скориставшись сервісом здати свій старий велосипед та повернути частину витрачених коштів.

У ході виконання випускної кваліфікаційної роботи для досягнення мети було вирішено ряд завдань.

- застосував машинне навчання як інструмент обробки даних;
- вивчив діяльність організації та побудувати її повну модель
- удосконалив обробку даних при формуванні цільової аудиторії компанії;
- удосконалив аналіз ефективності комерційної пропозицій.
- Оцінив конкурентоспроможність підприємства
- Підвищив конверсію продажів

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- 1) «Real-World Machine Learning» (Машинне навчання у реальному світі).  
Хенрик Бринк, Джозеф Ричардс, Марк Феверолф - Нью-Йорк , 2019 - 264 с.
- 2) «Advances in Financial Machine Learning» (Достижения в области  
финансового машинного обучения). Маркос Лопез де Прадо - Нью-Йорк ,  
2018 - 400 с.
- 3) «Pragmatic AI»(Прагматичний ІІ). Ной Гифт - Бостон , 2018 - 302 с.
- 4) «Relevant Search: With applications for Solr and Elasticsearch» (Релевантний  
пошук з використанням Elasticsearch та Solr). Даг Тарнбулл, Джон  
Берримен - - Нью-Йорк , 2016 - 360 с.
- 5) «The Master Algorithm» (Верховний алгоритм) Педро Домингос. - Нью-  
Йорк , 2016 - 352 с.
- 6) YouControl [Електронний ресурс] : [Веб-сайт]. – Електронні дані. –  
Аналітична онлайн-система Режим доступу:  
[https://youcontrol.com.ua/ru/catalog/company\\_details/34880125/](https://youcontrol.com.ua/ru/catalog/company_details/34880125/) (дата  
звернення 11.11.2021);
- 7) YouControl [Електронний ресурс] : [Веб-сайт]. – Електронні дані. –  
Аналітична онлайн-система Режим доступу:  
[https://youcontrol.com.ua/ru/catalog/company\\_details/30036332/](https://youcontrol.com.ua/ru/catalog/company_details/30036332/) (дата  
звернення 11.11.2021);
- 8) YouControl [Електронний ресурс] : [Веб-сайт]. – Електронні дані. –  
Аналітична онлайн-система Режим доступу:  
[https://youcontrol.com.ua/ru/catalog/for\\_details/32224952/](https://youcontrol.com.ua/ru/catalog/for_details/32224952/) (дата звернення  
11.11.2021);

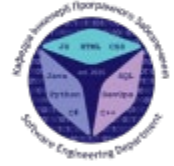
- 9) YouControl [Електронний ресурс] : [Веб-сайт]. – Електронні дані. – Аналітична онлайн-система Режим доступу:  
[https://youcontrol.com.ua/ru/catalog/company\\_details/44129937/](https://youcontrol.com.ua/ru/catalog/company_details/44129937/) (дата звернення 11.11.2021);

# ДОДАТОК

1 СЛАЙД



ДЕРЖАВНИЙ УНІВЕРСИТЕТ ТЕЛЕКОМУНІКАЦІЙ  
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ  
ТЕХНОЛОГІЙ



Кафедра інженерії програмного забезпечення

## МАГІСТЕРСЬКА РОБОТА «УДОСКОНАЛЕННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ ДЛЯ АВТОМАТИЗАЦІЇ АНАЛІТИЧНОЇ ДІЯЛЬНОСТІ КОМПАНІЇ VELO KYIV ІЗ ЗАСТОСУВАННЯМ МАШИННОГО НАВЧАННЯ»

Виконав: студент групи ПДМ-61 Максименко Руслан Сергійович  
Керівник: к.т.н., доц. кафедри ІІЗ Золотухіна О.А.

Київ - 2022

2 СЛАЙД

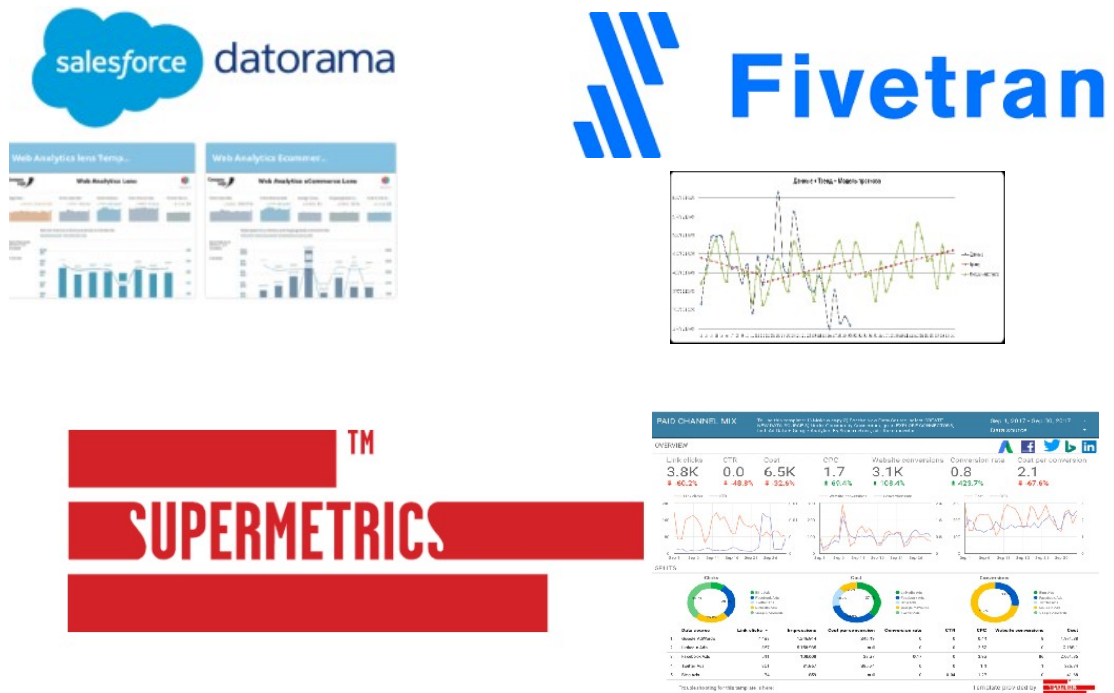
### МЕТА, ОБ'ЄКТА ТА ПРЕДМЕТ ДОСЛІДЖЕННЯ

**Мета роботи:** підвищення якості автоматизованої аналітичної діяльності інтернет-магазину VELO KYIV за рахунок удосконалення інформаційної системи на основі методів машинного навчання.

**Об'єкт дослідження:** Процес автоматизації аналітичної діяльності велосипедного інтернет-магазину

**Предмет дослідження:** Методи машинного навчання для автоматизації аналітичної діяльності інтернет-магазину

## АКТУАЛЬНІСТЬ РОБОТИ



## Кластеризація

**Кластеризація** – дані будуть згруповані як клієнти, і весь товар за подібними характеристиками з допомогою статистичних алгоритмів.

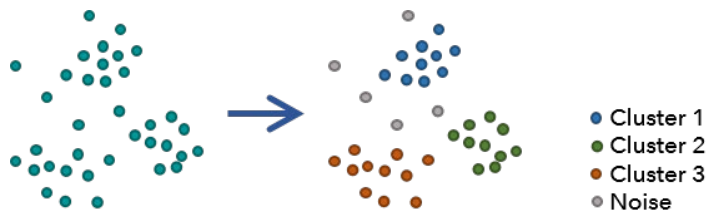
Властивості

- Розуміння даних шляхом виявлення кластерної структури
- Розбиття вибірки на групи подібних об'єктів
- Стиснення даних.

**Математичний опис**  $X$  — множина об'єктів,  $Y$  — множина номерів кластерів.

Задано функцію відстані між об'єктами  $\rho(x, x')$  та вибірка об'єктів

$X^m = \{x_1, \dots, x_m\} \subset X$ . Потрібно розбити вибірку на непересічні підмножини, що називаються кластерами, так, щоб кожен кластер складався з об'єктів, близьких по метриці  $\rho$  а об'єкти різних кластерів істотно відрізнялися. При цьому кожному об'єкту  $x_i \in X^m$  приписується номер кластеру  $y_i$





## Пошук асоціативних правил

**Пошук асоціативних правил**- дозволяють знаходити закономірності між пов'язаними подіями.

Властивості

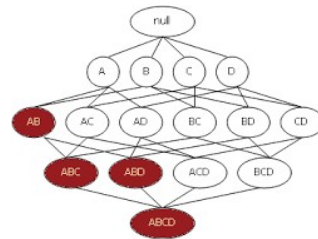
- виявлення зв'язків, що цікавлять нас, між змінними у великій базі дані
- допомагає машині імітувати виділення ознак людським та створити можливість знаходження абстрактних асоціацій

**Математичний опис**  $I = \{i_1, i_2, i_3, \dots, i_n\}$  - безліч (набір) товарів, які називають елементами.  $D$  Безліч транзакцій, де кожна транзакція  $T$  - це набір елементів з  $I, T \subseteq I$ . Кожна транзакція є бінарним вектором, де  $t(k) = 1$  якщо  $i_k$

Елемент присутній у транзакції, інакше  $t(k) = 0$ .

Асоціативним правилом називається -

$X \Rightarrow Y$ , де  $X \subset I, Y \subset I$  и  $X \cap Y = \emptyset$ .



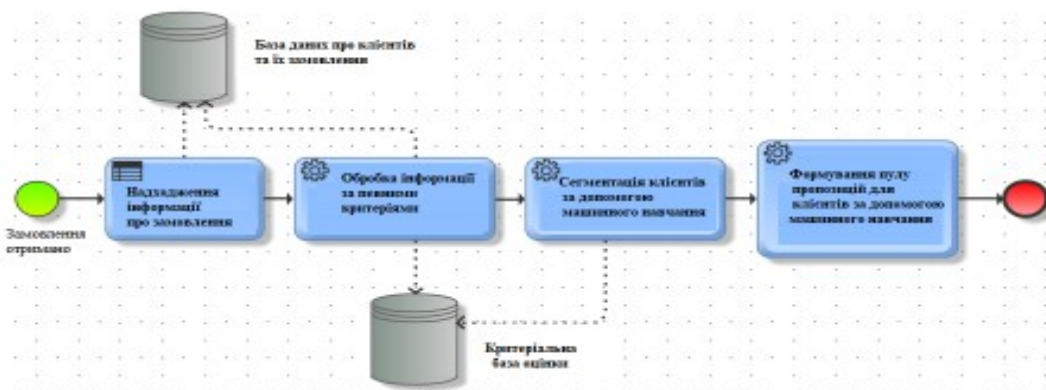
5

## КРИТЕРІАЛЬНА БАЗА КЛІЄНТА

Параметр	Тип	Значення
Стать	Текстовий	Чоловік/жінка
Зріст	Числової	Від 140 до 185 см (крок – 15 см)
Стан речі	Текстовий	Нове/бв
Сезонність	Текстовий	Літо, Весна, Осінь
Рівень бренду	Текстовий	Дешеві, Середні, Преміум
Лояльність клієнта	Логічний	Постійні/непостійні

6

## ПРОЦЕС ФОРМУВАННЯ ІНДИВІДУАЛЬНИХ ПРОПОЗИЦІЙ ДЛЯ КЛІЄНТІВ









## Приклад згенерованого повідомлення


Volixy вибрав вам 9 товарів, які вам можуть сподобатися.



Volixy

Руська, подивитесь, які товари ми підбрали для вас

 5.0/5.0 (10) Станція 4 Підсилювач Trek PROXIM 475 9 10 80 200 <b>84 000 грн</b>	 5.0/5.0 (10) Станція 3 Підсилювач Trek PROXIM 475 9 10 80 200 <b>64 300 грн</b>	 5.0/5.0 (10) Станція 3 Підсилювач Trek PROXIM 475 9 10 80 200 <b>75 546 грн</b>
 5.0/5.0 (10) Станція 4 Підсилювач Trek PROXIM 475 9 10 80 200 <b>56 700 грн</b>	 5.0/5.0 (10) Станція 3 Підсилювач Trek PROXIM 475 9 10 80 200 <b>132 032 грн</b>	 5.0/5.0 (10) Станція 3 Підсилювач Trek PROXIM 475 9 10 80 200 <b>110 306 грн</b>

Способи доставки:

у Києві 

по Україні  

### Аналіз ефективності комерційної пропозиції

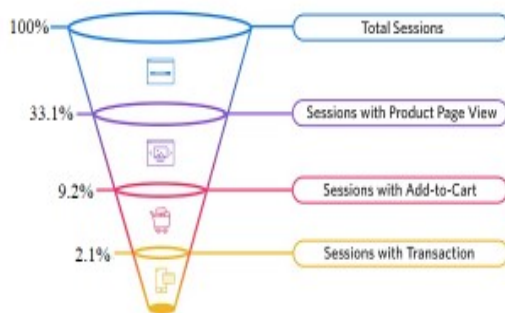
Аналіз ефективності  
комерційних пропозицій

Швидкість отримання зворотного зв'язку : 60 хв

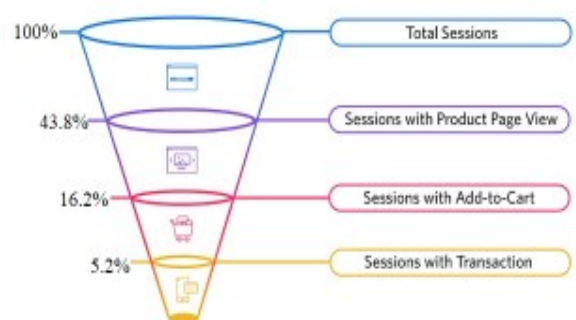
Відстеження досяжності пропозиції до користувача : перехід за  
посиланням з листа

Чіткий вимірний результат (конверсія): товар додан до кошику

### ПОРІВНЯЛЬНИЙ АНАЛІЗ КОНВЕРСІИ ПРОДАЖІВ



До вдосконалення  
інформаційної системи



Після вдосконалення  
інформаційної системи

## ВИСНОВКИ

1. Удосконалення інформаційної системи для автоматизації аналітичної діяльності компанії із застосуванням машинного навчання
2. Застосував машинне навчання як інструментів обробки даних
3. Побудував повну модель організації ТОВ «VeloKyiv»;
4. Удосконалив обробку аналізу ефективності комерційної пропозиції
5. Удосконалив обробку даних при формуванні цільової аудиторії компанії

## АПРОБАЦІЯ РОБОТИ

### Стаття:

1. Максименко Р.С.Золотухіна О.А.,“Оцінка конкурентоспроможності підприємства за допомогою методів машинного навчання” Телекомунікаційні та інформаційні технології. №3, 2021.(Подана до друку)

### Теза доповідей:

- 1.Максименко Р.С. "ДЕ ШТУЧНИЙ ІНТЕЛЕКТУ ВЖЕ КОНКУРУЄ З ЛЮДЬМИ". // XIII Міжнародна науково-технічна конференція студентства та молоді «Світ телекомунікації та інформатизації». – Київ: ДУТ, 2021. – 151 с.

13 СЛАЙД

**ДЯКУЮ ЗА УВАГУ!**