

**Київський національний торговельно-економічний
університет**

Кафедра кібернетики та системного аналізу

ВИПУСКНА КВАЛІФІКАЦІЙНА РОБОТА

на тему:

**«Аналіз ринків збуту за допомогою технологій Data
Mining»**

Студента 2 курсу, 1м групи,

спеціальності
051 «Економіка»

спеціалізації
«Економічна кібернетика»

Науковий керівник
доктор економічних наук, професор

Гарант освітньої програми
доктор фізико-математичних наук,
професор

Хоцького Андрія
Ігоровича

_____ *підпис студента*

Роскладка Андрій
Анатолійович

_____ *підпис керівника*

Гамалій
Володимир
Федорович

_____ *підпис гаранта*

Київ 2018

Київський національний торговельно-економічний університет

Факультет обліку, аудиту та інформаційних систем

Кафедра кібернетики та системного аналізу

Спеціальність 051 «Економіка»

Спеціалізація «Економічна кібернетика»

Зав. кафедри _____

Затверджую

Роскладка А. А.

«05» листопада 2017р.

Завдання на випускню кваліфікаційну роботу (проект) студента

Хоцького Андрія Ігоровича

(прізвище, ім'я, по батькові)

1. Тема випускної кваліфікаційної роботи (проекту)

«Аналіз ринків збуту за допомогою технологій Data Mining»

Затверджена наказом ректора від «02» жовтня 2017 р. № 3035

2. Строк здачі студентом закінченої роботи 15 листопада 2018 року

3. Цільова установка та вихідні дані до роботи

Мета роботи: Дослідження та аналіз ринків збуту товарів та послуг за допомогою технологій Data Mining.

Об'єкт дослідження: ринки збуту товарів та послуг.

Предмет дослідження: моделі і методи Data Mining у дослідженні ринків збуту.

4. Перелік графічного матеріалу _____

5. Консультанти по роботі із зазначенням розділів, за якими здійснюється консультування:

Розділ	Консультант (прізвище, ініціали)	Підпис, дата	
		Завдання видав	Завдання прийняв
1	Роскладка А. А.	05.11.2017 р.	05.11.2017 р.
2	Роскладка А. А.	05.11.2017 р.	05.11.2017 р.
3	Роскладка А. А.	05.11.2017 р.	05.11.2017 р.

6. Зміст випускної кваліфікаційної роботи (проекту) (перелік питань за кожним розділом)

ВСТУП

РОЗДІЛ 1. АНАЛІЗ РИНКІВ ЗБУТУ ТОВАРІВ ТА ПОСЛУГ

1.1. Поняття та сутність ринків збуту

1.2. Особливості ринків збуту і проблеми їх аналізу

1.3 Огляд методів аналізу ринків збуту товарів та послуг

Висновки до розділу 1

РОЗДІЛ 2. МЕТОДИ DATA MINING ДЛЯ ВИВЧЕННЯ РИНКІВ ЗБУТУ

2.1. Сутність і особливості методів Data Mining

2.2. Вибір моделей Data Mining для сегментації ринку збуту товарів та послуг

2.3. Побудова та оптимізація моделей аналізу ринку збуту

Висновки до розділу 2

РОЗДІЛ 3. ПРОГРАМНА РЕАЛІЗАЦІЯ МЕТОДІВ DATA MINING НА РИНКАХ ЗБУТУ

3.1. Вибір інструментарію Data Mining для реалізації моделі аналізу ринку збуту

3.2. Проектування та програмна реалізація

3.3. Результати використання програмної розробки

Висновки до розділу 3

ВИСНОВКИ

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

ДОДАТКИ

7. Календарний план виконання роботи

№ пор.	Назва етапів випускної кваліфікаційної роботи	Строк виконання етапів роботи	
		за планом	фактично
1	2	3	4
1	<i>Вибір теми випускної кваліфікаційної роботи</i>	01.10.2017	01.10.2017
2	<i>Розробка та затвердження завдання на випускну кваліфікаційну роботу</i>	05.11.2017	05.11.2017
3	<i>Вступ</i>	01.04.2018	
4	<i>Розділ 1. Аналіз та огляд ринків збуту</i>	01.05.2018	
5	<i>Розділ 2. Data Mining на ринку збуту товарів та послуг</i>	20.06.2018	
6	<i>Підготовка статті у збірник наукових статей магістрів</i>	15.09.2018	
7	<i>Розділ 3. Практичне застосування data mining на ринку збуту</i>	01.10.2018	
8	<i>Висновки</i>	01.11.2018	
9	<i>Здача випускної кваліфікаційної роботи на кафедрі науковому керівнику</i>	15.11.2018	
10	<i>Попередній захист випускної кваліфікаційної роботи</i>	22.11.2018	
11	<i>Виправлення зауважень, зовнішнє рецензування випускної кваліфікаційної роботи</i>	25.11.2018	
12	<i>Представлення готової зшитої випускної кваліфікаційної роботи на кафедрі</i>	28.11.2018	
13	<i>Публічний захист випускної кваліфікаційної роботи</i>	За розкладом роботи ЕК	

8. Дата видачі завдання «05» листопада 2017 р.

9. Керівник випускної кваліфікаційної роботи (проекту)

Роскладка А. А.

(прізвище, ініціали, підпис)

10. Гарант освітньої програми

Гамалій В. Ф.

(прізвище, ініціали, підпис)

11. Завдання прийняв до виконання студент-дипломник

Хоцький А. І.

(прізвище, ініціали, підпис)

Анотація

В даній роботі розглянуто сутність, поняття ринку збуту його класифікацію та особливості. Розглянуто проблематику аналізу збутової діяльності на ринках товарів та послуг. Досліджено основні моделі Data Mining, розглянуто поняття та моделі пошуку цінної інформації серед масивів необроблених даних. Взявши за основу теоретичний матеріал особливостей ринків збуту товарів та послуг було обрано один об'єкт для створення та реалізації програмного інструменту. Проведено вибір моделі пошуку цінної інформації для ринку ігрової індустрії. Ґрунтуючись на дослуженому теоретичному матеріалі було виконано практичну реалізацію пошукової роботи для користувачів платформи Steam. Створений інструмент розроблено для підвищення продуктивності та прибутковості підприємницької діяльності.

Ключові слова: ринок збуту, data mining, Steam, web-crawler, брудні дані, цінна інформація, пошуковий робот, збір даних.

Anotation

The essence, marketing market concept, its classification and features are considered in this thesis. The sales activity analysis in the goods and services markets is considered. The main Data Mining models have been investigated; concepts and models of the search of valuable information among the arrays of raw data are considered. Based on the theoretical material of the markets features for the sale of goods and services, one object was selected for the software tool creation and implementation. The choice of the model of search for valuable information for the market of the gaming industry was conducted. Based on a theoretical material, a search work practical implementation for Steam platform users was performed. The created tool is designed to increase the productivity and profitability of entrepreneurial activity.

Key words: market for goods and services, data mining, Steam, web crawler, dirty data, valuable information.

ЗМІСТ

ВСТУП.....	3
РОЗДІЛ 1. АНАЛІЗ РИНКІВ ЗБУТУ ТОВАРІВ ТА ПОСЛУГ.....	6
1.1. Поняття та сутність ринків збуту.....	6
1.2. Особливості ринків збуту і проблеми їх аналізу.....	21
1.3. Огляд методів аналізу ринків збуту товарів та послуг.....	26
Висновки до розділу 1.....	32
РОЗДІЛ 2. МЕТОДИ DATA MINING ДЛЯ ВИВЧЕННЯ РИНКІВ ЗБУТУ.....	33
2.1. Сутність і особливості методів Data Mining.....	33
2.2. Вибір моделей Data Mining для сегментації ринку збуту товарі та послуг.....	37
2.3. Побудова та оптимізація моделей аналізу ринку збуту.....	47
Висновки до розділу 2.....	52
РОЗДІЛ 3. ПРОГРАМНА РЕАЛІЗАЦІЯ МЕТОДІВ DATA MINING НА РИНКАХ ЗБУТУ.....	54
3.1. Вибір інструментарію Data Mining для реалізації моделі аналізу ринку збуту.....	54
3.2. Проектування та програмна реалізація.....	58
3.3. Результати використання програмної розробки.....	62
Висновки до розділу 3.....	67
ВИСНОВКИ.....	68
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	70

ВСТУП

Актуальність теми дослідження. В результаті розвитку інформаційних технологій, кількість даних, накопичених людством в електронному вигляді, зростає швидкими темпами. Ці дані існують навколо нас в різних видах: тексти, зображення, аудіо, відео, гіпертекстові документи, реляційні бази даних і т.д. Величезна кількість даних з'явилася в результаті повсюдного використання мережі Інтернет яка значно полегшила доступ до інформації з графічно віддалених точок Землі. Однак, переважна частина доступної інформації не несе для конкретної людини будь - якої користі. Людина не в змозі переробити таку кількість відомостей. Виникає проблема вилучення корисної для користувача інформації з великого обсягу «сирих» даних.

Організації накопичують величезні масиви даних, однак не в змозі отримати від цієї роботи реальну віддачу. Маючи історичні дані, можна вирішити критично важливі для бізнесу завдання: оптимізувати процеси, управляти ризиками, підвищувати прибутковість, утримувати клієнтів. Обґрунтованість і професійний рівень прийняття рішень визначає ефективність діяльності фірми. В цьому відношенні істотну допомогу керівнику надають сучасні інформаційні системи. Аналіз інформації є невід'ємною частиною ведення бізнесу і одним з важливих чинників підвищення його конкурентоспроможності. В проведенні статистичного аналізу даних допомагають методи Data Mining. Існуючі статистичні системи через свою високу ціну ліцензії, складність у освоєнні та відсутність можливості працювати з портативних приладів стримують впровадження подібних систем у виробництво. Тому розробка та дослідження інформаційної системи статистичного аналізу даних на основі методів Data Mining є актуальним напрямком досліджень для подальшого впровадження на підприємствах.

У запропонованій роботі окреслено основні теоретичні особливості понять «Data Mining» та «Ринок збуту», проаналізуємо існуючі моделі та методи Data Mining, оберемо та створимо власну розробку виходячи з обраного методу та моделі, яка буде шукати та аналізувати великі масиви даних в обраній сфері

ринку збуту товарів. Проаналізуємо економічну важливість технологій пошуку корисної інформації серед великих масивів даних для підприємств різних рівнів.

Об'єктом дослідження виступають ринки збуту товарів та послуг.

Предметом дослідження є моделі і методи Data Mining у дослідженні ринків збуту.

Мета роботи полягає у дослідженні та аналізі ринків збуту товарів та послуг за допомогою технології Data Mining. Створення власного програмного інструменту для пошуку цінної інформації серед великої кількості доступних «сирих» даних для практичного застосування на підприємствах певної сфери ринку збуту, в цілях покращення економічної вигоди компанії.

Для досягнення мети необхідно проаналізувати поняття ринку збуту товарів та послуг, дізнатися про існуючі технології пошуку цінної інформації, виконати аналіз сфер застосування технологій та методів Data Mining. Виходячи з проаналізованих даних обрати ринок, збуту який можна доповнити цінною інформацією, для створення інструменту Data Mining з цілю віднайти потрібну інформацію для економічного прибутку підприємств які надають послуги чи продають власний товар у обраній сфері.

Для досягнення поставленої мети потрібно виконати такі **завдання**:

- розкрити економічну сутність поняття збутової діяльності як економічної категорії;
- визначити проблематику аналізу ринку збуту товарів та послуг у сучасних умовах;
- охарактеризувати методи аналізу ринку товарів та послуг;
- проаналізувати сутність і особливості методів Data Mining у різних сферах діяльності підприємств;
- визначити основні моделі Data Mining для сегментації ринку товарів та послуг;
- побудова та оптимізація моделей для аналізу ринку збуту товарів та послуг;

- виконати практичну реалізацію для створення моделі Data Mining на ринку збуту товарів та послуг;
- виконати практичне застосування розробленого інструменту та проаналізувати отримані дані.

Практичне значення. Розроблений програмний інструмент створено для пошуку цінної інформації серед великих масивів даних для збору потрібної інформації, яку можна використовувати підприємствам для підвищення продажів своїх товарів за рахунок більшої обізнаності про свою цільову аудиторію покупців, що залежить від економічної окупності та збільшенню прибутків на підприємстві.

Структура роботи. Робота складається зі вступу, трьох розділів, висновків, списку використаних джерел в обсязі 30 найменування. Повний обсяг роботи становить 72 сторінок, із яких основна частина роботи займає 62 сторінок.

РОЗДІЛ 1

АНАЛІЗ РИНКІВ ЗБУТУ ТОВАРІВ ТА ПОСЛУГ

1.1. Поняття та сутність ринків збуту

Щоб зрозуміти, які існують типи ринків збуту товарів та послуг ми оглянемо поняття та їх визначення у даному підрозділі, та дамо характеристику описаним варіантам.

Ринок збуту — в широкому сенсі, будь-який економічний простір для пропозиції товарів і послуг, що є кінцевою метою господарської діяльності підприємства. [7]

Ринок збуту товарів (послуг) — аналітичний розділ бізнес-плану, який дозволяє на основі результатів аналізу можливостей діючих ринків, попиту на продукцію (послуги) фірми визначити сегменти ринку, прийнятні для продукції фірми, зумовити можливі ніші, оцінити потенційну ємність ринків, плановану і фактичну величину обсягу реалізації та виручку. [7]

Ринок реального товару — ринок, на якому ведеться торгівля реальним товаром на відміну від ф'ючерсної біржі. На ринку реальних товарів укладають угоди як з негайним постачанням («спот»), так і з постачанням на термін, у майбутньому. Форми організації ринку реальних товарів — аукціон, довгостроковий контракт, «телефонно-телексний ринок», разова угода, тендер, біржа реального товару і т. п.

Потенційний ринок — сукупність покупців, які виявляють зацікавленість до певного товару або послуги.[3]

Ринок внутрішній — сфера товарного обміну, обмежена державними кордонами; сукупність економічних відносин товаровиробників і споживачів у частині реалізації товару.[3]

Система збуту товарів — ключова ланка маркетингу, завершальне комплекс заходів щодо створення продукції, виробництву і доведенню її до споживача. На цьому етапі покупець визнає чи ні зусилля фірми корисними і

потрібними для себе, а значить, купує або не купує товар. У поняття збут слід включити ряд функцій: транспортування, складування, зберігання, доробку, просування оптовим і роздрібним торговельним ланкам, передпродажну підготовку і власне продаж товару.[10]

Необхідність створення збутової мережі підприємства обумовлена тим, що виробник далеко не завжди готовий взяти на себе всі обов'язки та функції, що впливають з вимог вільного обміну відповідно до очікувань потенційних споживачів. Значення збутової мережі особливо зростає в умовах конкуренції, глобалізації ринку, електронних методів поширення інформації та скорочення життєвого циклу продукту або послуги.

Збут продукції будь-якої компанії здійснюється через систему розподілу, ефективність якої гарантує успіх бізнесу.

Канал товароруку (розподілу) — це шлях, по якому товари рухаються від виробника до споживача. Канали розподілу можна визначити як сукупність незалежних посередницьких організацій, що забезпечують передачу права власності на товар (послугу) будь-кому іншому на шляху його руху від виробника до споживача. Канали розподілу - це більше, ніж прості об'єднання фірм. Вони являють собою складні поведінкові системи, в яких люди і компанії взаємодіють для реалізації своїх цілей. [16]

Основним управлінським рішенням виробника в області товароруку є вибір числа учасників каналу, при якому розподіл функцій між ними дозволить мінімізувати витрати обігу при своєчасному і якісному реченні цільовому ринку необхідного асортименту товарів та забезпеченні максимуму комфорту покупцям. Безпосередні поставки товарів від виробника до споживача називаються фізичним розміщенням, яке включає в себе три основні елементи: створення товарного складу, транспортування, управління матеріально-технічним постачанням.

Використання посередницьких ланок вигідно самим виробникам, оскільки забезпечує скорочення числа зв'язків і контактів, які необхідні при реалізації продукції. Канали розподілу можна охарактеризувати за кількістю

складових рівнів.

Рівень каналу розподілу - це будь-який посередник, який виконує ту чи іншу роботу по наближенню товару і вдачі власності на нього до кінцевого покупця. Оскільки певну роботу виконують і сам виробник, і кінцевий споживач, вони теж входять до складу будь-якого капала.

Існуючі канали розподілу припускають використання наступних методів збуту.[10]

Прямий збут (канал товароруку нульового рівня) не передбачає наявності посередників, продаж здійснюється безпосередньо споживачам на основі прямих контактів - через власну торговельну мережу, по оголошеннях в газетах і журналах з купонами для відповідей, через Інтернет, телемаркетинг, директ-мейл. Продаж через Інтернет набирає обертів і незабаром може витіснити інші методи дистрибуції. Подібний канал використовується для реалізації товарів, що вимагають демонстрації і переговорів для того, щоб переконати клієнта їх придбати: косметика, сировина, обладнання та ін.

До переваг прямих каналів можна віднести можливість повного контролю з боку виробника за рухом товару, до недоліків - незначне охоплення ринку збуту. Якщо прямий збут носить постійний, а не разовий характер, підприємство повинно мати власні регіональні склади. Канал товароруку, що містить нульовий рівень, відноситься до прямого каналу, що містить більше число проміжних рівнів - до непрямих каналам. З позицій виробника, чим більше рівнів має канал розподілу, тим менше можливостей його контролювати. [7]

Непрямий збут (багаторівневий канал товароруку) передбачає продаж продукції через торгові організації, незалежні від виробника, і використовується для досягнення географічно розосереджених ринків. Подібний вид збуту практикується при реалізації товарів широкого вжитку. Товар від виробника спочатку йде до посередника, а від нього або до кінцевого споживача, або до іншому посереднику. Перевагами непрямих

каналів є розширення можливостей збуту, меж ринку і т.д. Недоліки пов'язані з ускладненням контролю за просуванням товару на ринок. Кожен посередник у ланцюжку товароруху являє собою окремий рівень каналу розподілу. Виділяють одно-, дво- і тривірневі канали.

Однорівневий канал включає в себе одного посередника. На споживчих ринках цим посередником зазвичай буває роздрібний торговець, а на ринках товарів промислового призначення - агент по збуту або брокер. Велика роздрібна мережа, закупаючи товари у багатьох виробників, продає їх у своїх магазинах кінцевим споживачам.

Дворівневий канал містить двох посередників. На споживчих ринках такими посередниками зазвичай стають оптовий і роздрібний торговці, на ринках товарів промислового призначення - промисловий дистриб'ютор і дилер. Через дистриб'юторів поширюється офісна оргтехніка, програмне забезпечення. Наприклад, група компаній "Формоза" - найбільший російський виробник комп'ютерної техніки, має у своєму складі дистриб'юторську фірму, що реалізовує продукцію через регіональних дилерів.

Тривірневий канал охоплює трьох посередників - оптового, дрібнооптового посередників і роздрібногo торговця. Наприклад, компанія "Вімм-Білл-Данн" реалізує свою молочну продукцію через вісім дистриб'юторських фірм, які поставляють її дилерам, налагоджувати прямі контакти з роздрібними мережами. [7]

Багато компаній використовують змішані канали розподілу. Ці канали об'єднують в собі ознаки і прямих, і непрямих каналів. Продукція реалізується як прямими поставками, так і через посередницькі ланки, які часто називаються рівнями розподілу.

Канал збуту можна оцінити двома параметрами - довжиною капала розподілу, тобто протяжністю, і шириною. Довжина каналу розподілу визначається числом проміжних ланок і пов'язана з передачею функцій посередникам. З погляду виробника, велике число рівнів означає втрату

контролю над процесом продажу і складну схему каналів товароруху. Загально визнано, що чим канал коротший, тим він дешевше. Тому продаж компанією продукції в прямому режимі може забезпечити велике охоплення ринку, але призведе до серйозних витрат на транспортування і складування. Це може компенсуватися більш високою маржею прибутку, одержуваної за рахунок усунення посередників у процесі дистрибуції, і тим, що маржа з ними не ділиться. Крім фінансових критеріїв короткі канали володіють перевагою більш оперативного виходу на кінцевих споживачів.

В останні роки виробники намагаються використовувати більш короткі канали, щоб контролювати дистрибуцію своєї продукції. Особливо це стосується товарів, які вимагають попереднього знайомства і вибору, а тому для їх просування використовується дорога реклама. Практика показує, що дешеві товари з більш низьким рівнем використовуваних технологій краще пристосовані до довгих каналів. Більш складні товари, часто вимагають післяпродажних послуг, продаються через більш короткі канали. Більшість виробів промислового призначення реалізуються виробниками безпосередньо користувачам. Кількісною характеристикою каналу, поряд з довжиною, є і його ширина - число посередників (оптових і роздрібних) на будь-якому етапі реалізації продукції підприємства (наприклад, число всіх оптових фірм, що закупають продукцію у виробника). Ширина каналу - важливий фактор, що дозволяє скорочувати терміни збуту. Зазвичай великі оптові фірми розташовуються в безпосередній близькості від виробника або у великих регіональних центрах; дрібнооптові - в більш віддалених і менших за чисельністю населених пунктах; роздрібні торговці - в місцях проживання кінцевих споживачів.[16]

При вузькому каналі виробник продає свою продукцію через декількох учасників збуту, при широкому - через багатьох. Бажаючи зміцнити своє становище на певному етапі каналу, компанія може реалізувати горизонтальну інтеграцію або експансію, придбавши підприємства аналогічної спеціалізації. Це дозволяє компанії збільшити свій розмір, частку

ринку, ефективніше використовувати засоби масової інформації та методи розподілу та збуту.

Виробники послуг і ідей стикаються з проблемою забезпечення доступності своїх пропозицій для цільових аудиторій. Тому компанії, що діють в сфері надання послуг, створюють власні системи розподілу, що відповідають особливостям їх пропозиції. Наприклад, лікарні необхідно розташувати таким чином, щоб вони охоплювали своїми послугами жителів району і забезпечували їм повне медичне обслуговування.

Рішення про вибір каналів розподілу є одним з ключових у діяльності компанії. Кожен канал пов'язаний з певним рівнем збуту і витрат, повинен формуватися на тривалий період часу. У зв'язку з цим компанія може встановити кілька варіантів дистрибутивних ланцюжків, що включають в себе різну кількість посередників. Управління каналами розподілу тісно пов'язане з проблемами відбору та мотивування партнерів по каналу. Періодична оцінка ефективності роботи кожної ланки ланцюжка товароруку на основі аналізу прибутковості є обов'язковою процедурою, що забезпечує життєздатність каналу. Економічна роль каналів товароруку

Економічна роль каналів товароруку пов'язана з тим, що, вступаючи в певні відносини з партнерами, компанія прагне мінімізувати витрати часу та коштів. Це проявляється в отриманні ряду переваг. Розподіляючи витрати між декількома партнерами, фірма економить на масштабі. Наприклад, багато авіакомпаній заохочують використання пасажирами послуг туристичних агентств, які оформляють і переоформляють квитки, приймають оплату, здійснюють доставку квитків клієнтам. Таким чином, можлива економія на скороченні числа контактів виробників з кінцевими споживачами. Зменшуючи число рівнів каналів збуту, ресторани часто безпосередньо закупають продукти у сільськогосподарських виробників. Багато компаній розраховують на більш високий рівень обслуговування клієнтів з боку посередників, оскільки, як правило, вони ближче до кінцевого споживача і краще знають його потреби і переваги.

Характер ринку є важливою складовою ринку збуту товарів та послуг, якому варто приділити увагу надалі.

Розмір ринку визначається числом потенційних покупців. Якщо ринок великий, залучення посередників обов'язково, якщо малий, то компанія сама може забезпечити збут своїх товарів і послуг. Якщо покупки носять нерегулярний або сезонний характер, доцільно використовувати довгий канал. Більшість кінцевих споживачів робить покупки в роздрібній мережі, ділові споживачі мають справу з виробником. Продукція, призначена як для споживчого, так і для ділового ринку, реалізується більш ніж через один канал (авіаквитки, програмне забезпечення, комп'ютери). Для охоплення ринку з малим числом споживачів і великим обсягом середньої закупівлі використовується прямий канал. На географічно розосередженому ринку, де купуються дрібні партії, залучаються посередники (ринок споживчих товарів). При роботі з різними сегментами поєднують прямі і непрямі канали. На ринках з великим обсягом продажів використовуються прямі канали, з меншим обсягом реалізації - незалежні торгові представники. Інтернет і діалогові бази даних для споживачів відкрили нові можливості збуту товарів і послуг, що дозволило створити прямі канали збуту.

Гуртова торгівля є окремою ланкою ринку збуту та має свої особливості описані далі. [3]

Оптова торгівля — це підприємницька діяльність з продажу товарів або послуг тим, хто купує їх з метою перепродажу організаціям роздрібною торгівлі або іншим оптовим організаціям, але не кінцевим споживачам-індивідуумам. Оптова торгівля є важливою ланкою дистрибуції та вирішує багато маркетингові завдання. Роль оптової торгівлі полягає у максимальному задоволенні потреб роздрібних мереж на основі поставки їм необхідних товарів у визначених обсягах і у встановлені терміни. Розташовуючись зазвичай у великих містах, оптові компанії добре знають потреби кінцевих покупців, можуть організувати маркетингову підтримку роздрібною торгівлі. Досвід показує, що оптові фірми здійснюють збутові

функції краще виробників, оскільки мають постійні контакти з роздрібною мережею, розташовують розвиненим складським і транспортним господарством. Оптова торгівля є важливим важелем маневрування матеріальними ресурсами, сприяє скороченню зайвих запасів продукції на всіх рівнях і усуненню товарного дефіциту, бере участь у формуванні регіональних і галузевих товарних ринків. Через оптову торгівлю посилюється вплив споживачів на виробників, у свою чергу виробник сам вибирає споживачів. Оптовики надають партнерам не тільки товари, але й широкий спектр послуг: рекламу на місці продажу; доставку товару; передпродажну підготовку, в тому числі розфасовку і упаковку; організацію заходів щодо стимулювання збуту. На ринку технічно складних товарів оптові фірми організують сервісні центри за підтримки виробників. Оптові торговці отримують право власності на товар і фізичне володіння їм; мають склади для зберігання товарів часто кількох чи багатьох виробників; займаються просуванням товарів; регулюють проблеми фінансування, замовлень і здійснення платежів зі своїми покупцями. [3]

Оптові посередники отримують прибуток, купуючи товари за оптовою ціною і продаючи з націнкою, віднімаючи при цьому витрати на дистрибуцію. Діяльність будь-якого посередника збільшує вартість товару. Тому завдання оптової ланки системи збуту полягає у формуванні мінімальної оптової націнки (за рахунок раціоналізації торгово-логістичних операцій) або в доданні товару додаткової цінності для покупця, що сприймає встановлену ціну як справедливую.

Оскільки оптовий посередник повинен задовольняти інтереси виробників, роздрібних торговців і кінцевих споживачів, виникло безліч методів і форм оптової торгівлі.

Класифікація оптової торгівлі. За широтою асортимент може бути широким (1-100 тис. Найменувань), обмеженим (менше 1000 найменувань), вузьким (менше 200 найменувань) і спеціалізованим. [3]

За способом доставки оптової торгівлю поділяють на такі види:

доставка своїм транспортом, продаж зі складу (самовивіз).

За ступенем кооперації виділяють: горизонтальну кооперацію для спільних закупівель і організації оптових ринків; вертикальну кооперацію для цілей збуту і конкуренції з роздрібною торгівлею за ринки кінцевих споживачів.

За розміром обороту оптовиків ділять на великих, середніх і дрібних.

З точки зору організації оптової торгівлі існують три загальні категорії: оптова торгівля виробника, підприємств-посередників, здійснювана агентами і брокерами.

Торгівля, здійснювана виробниками за допомогою власних органів збуту, вимагає створення дочірньої оптової компанії. Діяльність такої компанії виправдана, якщо асортимент та обсяги виробленої продукції будуть достатніми для прибуткової їх реалізації. В іншому випадку функції оптової ланки доцільно передати незалежним компаніям. Торгові представники виробника, агенти і брокери зазвичай не отримують ні право власності, ні фізичне володіння товаром. Вони займаються просуванням товару і переговорами з приводу умов продажу.[16]

Вибір виду торгівлі для оптовика визначається з урахуванням бажання і розміру (потужності) роздрібною організації. Оптова торгівля з особистим відбором доцільна, коли ритейлеру необхідно зробити термінову закупівлю (закінчуються складські запаси), вибрати "ходової" асортимент, отримати для реалізації новинку, знижки за самовивіз. Особистий відбір використовується при закупівлі швейних виробів, тканин, хутряних виробів тощо

Для демонстрації товарів ринкової новизни оптовик часто обладнає демонстраційний або виставковий зал. При особистому відборі може бути проведено опитування відвідувачів, тестування товарів. Результати мають бути враховані в процесі закупівлі та реалізації демонстрованих виробів. Для багатьох стандартних товарів особистий відбір організують з використанням самообслуговування. Переміщення відібраних виробів здійснюється за допомогою засобів малої механізації: штабелери,

вантажних візків, транспортних рольгангів і т.д.

Оптова торгівля за письмовою заявкою або телефону проводиться на підставі попередньо підписаного договору між продавцем і покупцем, в якому обумовлюються умови оплати замовлення. Доставка товару в магазин може бути здійснена транспортом оптовика або ритейлера. Від надійності системи доставки оптової ланки буде залежати, заощадить Чи роздрібний торговець час або швидкість доставки його не влаштує.

Оптова торгівля за допомогою роз'їзних збутових агентів і менеджерів набула широкого поширення, будучи найбільш активною формою розподілу. Оптова компанія організовує агентську мережу для пошуку покупців - більш дрібних оптовиків і магазинів. Роз'їзні агенти підтримують контакти з клієнтами, контролюють наявність товарів у торговельному залі магазину, їх викладку, своєчасність розрахунків за товар і т.д. Збутових агентів звичайно закріплюють за певною територією, за групою клієнтів або асортиментом товару.

Оптова торгівля з використанням телемаркетингу (дзвінків з офісу або відділу збуту) вимагає створення диспетчерської служби, навчання персоналу навичкам ведення телефонних переговорів. Отриману про потенційних покупців інформацію диспетчери передають у відділ збуту. Диспетчерська служба може приймати замовлення від покупців, проводити телефонні опитування, вести статистику збуту.

Оптова торгівля на виставках і ярмарках дозволяє укладати договори на поставку під час відвідування виставки представниками покупця або провести попередні переговори. Присутність на подібних заходах великої кількості професіоналів (виробників, посередників, споживачів) дозволяє отримати останню інформацію про кон'юнктуру ринку, новинки, досягнення конкурентів та ін. На виставках доцільно проводити промоушен-кампанії, орієнтовані на споживачів.

Оптові посередники виконую наступні функції:

- закупівля для споживачів - прогнозування попиту і на основі аналізу

результатів формування асортименту для споживачів;

- продажу та просування для виробників - надання виробникам штату продавців для звернення до роздрібним торговцям і діловим користувачам. Роздрібні оператори та ділові покупці найчастіше контактують з оптовими посередниками, ніж з виробником, більше довіряють їм;

- зберігання запасів за низькими цінами - скорочення запасів, складських інвестицій і ризику постачальників і споживачів;

- транспортування - забезпечення найбільш швидкої та ефективної доставки за рахунок близькості до виробника;

- розбивка великих партій - закупівля в економічно вигідному масштабі для подальшої передачі більш дрібними обсягами роздрібним операторам і діловим споживачам;

- надання маркетингової інформації - виробникам про запити споживачів, діяльності конкурентів, галузевих тенденціях; споживачам - про нові товари;

- фінансування - надання кредиту виробнику або споживачам, купуючи продукцію до того, як продав сам;

- прийняття ризику - транспортуючи і зберігаючи продукцію, посередник бере на себе ризик псування, крадіжки або застарівання товару;

- управлінський, методичний та технічний сервіс - оптовий посередник може проводити тренінги для роздрібних партнерів, надавати методичну та технічну допомогу в оформленні точок продажу.

Типи оптових посередників. Всіх оптових посередників можна класифікувати за наступними критеріями: форма власності па посередницькі структури; право власності на просування товари та послуги.

За формою власності на посередницькі структури оптові посередники діляться на три групи: оптовики у власності виробника (служба збуту або продажів, магазини виробника); незалежні оптові посередники; кооперативи та закупівельні офіси у власності роздрібних торговців.

Виробник може розподіляти продукцію через власні підрозділи, що

займаються збутом, що забезпечує прямий контроль за постачанням. Такий контроль важливий для технічних систем, що вимагають установки та обслуговування; швидкопсувних товарів; консалтингових послуг. Крім того, реалізація товарів високої вартості дозволяє виробникові прибутково продавати їх безпосередньо кінцевому споживачеві.

До власних структурам оптових продажів виробника відносяться філії продажів (збуту), офіс продажів, торговельні ярмарки, торгові центри.

Філія продажів зберігає запаси і обробляє замовлення на товари, виконує функції складування, служить офісом продажів для торгових представників на своїх територіях. Вони переважають в маркетингових каналах, що реалізують машини й устаткування.

Офіс продажів не має запасів продукції, але служить регіональним офісом для персоналу продажів виробника. Близьке розташування до споживача дозволяє знизити витрати на продажі і здійснювати ефективне обслуговування споживачів.

Торговий ярмарок — періодичний показ виробником своїх товарів оптовим і роздрібним покупцям.

Торговий центр — надання виробником простору для постійних експозицій товарів.

Право власності на просування товари та послуги може залишатися у виробника або передаватися незалежним посередникам.

Види посередників. По відношенню до виробників посередників слід поділяти на залежних і незалежних.

Незалежні посередники купують товари у виробників у свою власність для подальшого перепродажу з прийнятною націнкою, яка покриває їх витрати і формує прибуток (Рисунок 1.1)

Зовсім посередники не купують товари у свою власність, а отримують від виробників комісійні винагороди за реалізацію продукції (Рисунок 1.2).

Більшість оптових фірм просувають продукцію в регіони за рахунок створення дилерської мережі. Дилери - незалежні торгові посередники,

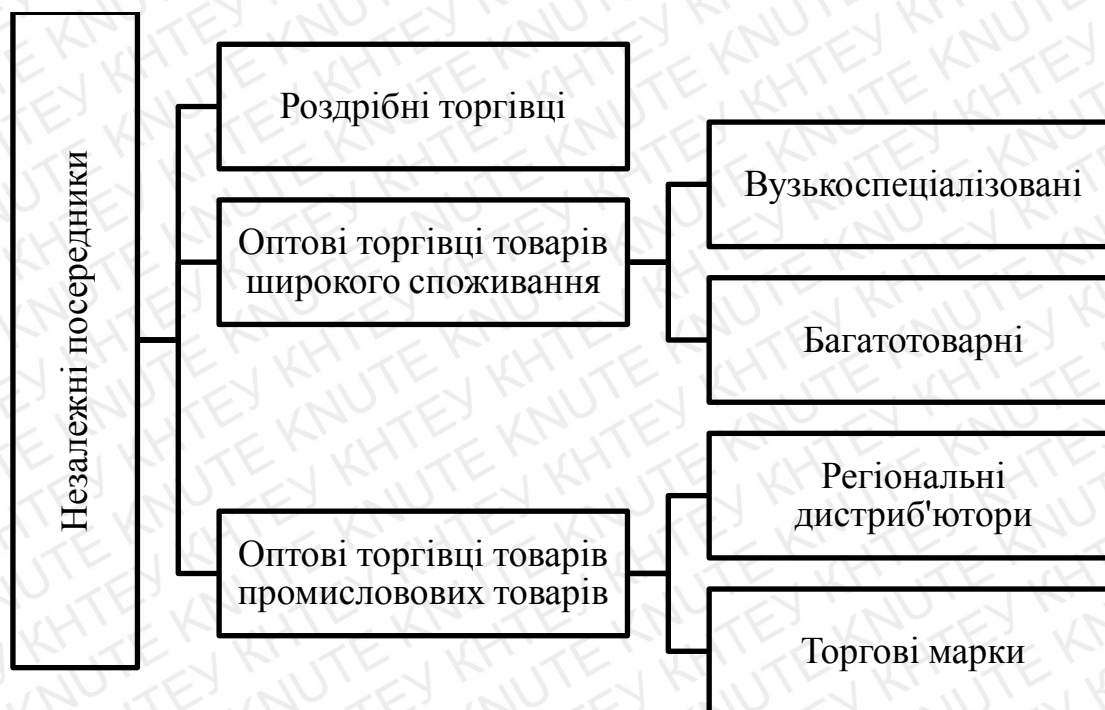


Рис. 1.1. Незалежні посередники

Джерело: Розроблено автором на основі даних [7]

уповноважені одним або обмеженим числом компаній-продавців надавати підтримку кінцевим споживачам. Регіональний дилер забезпечує регулярні закупівлі товарів, здійснює планування закупівель і продажів. Скорочення числа дилерів в період дефолту призвело до того, що на ринку залишилися найбільш ініціативні і заповзятливі. Наприклад, московська компанія-дилер, що представляє велику зарубіжну мережу з продажу взуття, відкрила в одному з міст Сибіру філія магазину, використовуючи напрацьований досвід. Оформлення нового магазину, торговельне обладнання, нестандартні рішення по презентації магазину, концепція продажів, застосовувана зарубіжним виробником, дозволили компанії привернути увагу місцевих споживачів, заявити про себе в місті як про представника передових технологій, значно збільшити продажі. Дилер зазвичай має паралельний бізнес, роблячи ставку на різні споживчі сегменти, несучи самостійні витрати за розвиток бізнесу і отримуючи весь прибуток від продажів.

Дистриб'ютор - фізична або юридична особа - оптовий посередник,

обслуговуючий різні галузі промисловості, що має склади і транспортні засоби та здійснює комерційну діяльність від свого імені і за свій рахунок. Дистриб'ютори широкого профілю забезпечують конкретний географічний регіон великою різноманітністю і кількістю товарів. Спеціалізовані дистриб'ютори займаються розповсюдженням вузького асортименту продукції, надають споживачам необхідну інформацію. Дистриб'юторів широкого профілю називають ще оптовиками, дистриб'юторами багато товарними або з широким асортиментом, дистриб'юторами з власним товарним складом. Їх перевагами є: широкий асортимент товарів, його доступність; конкурентоспроможне ціноутворення; налагоджені взаємини з клієнтами. [7]

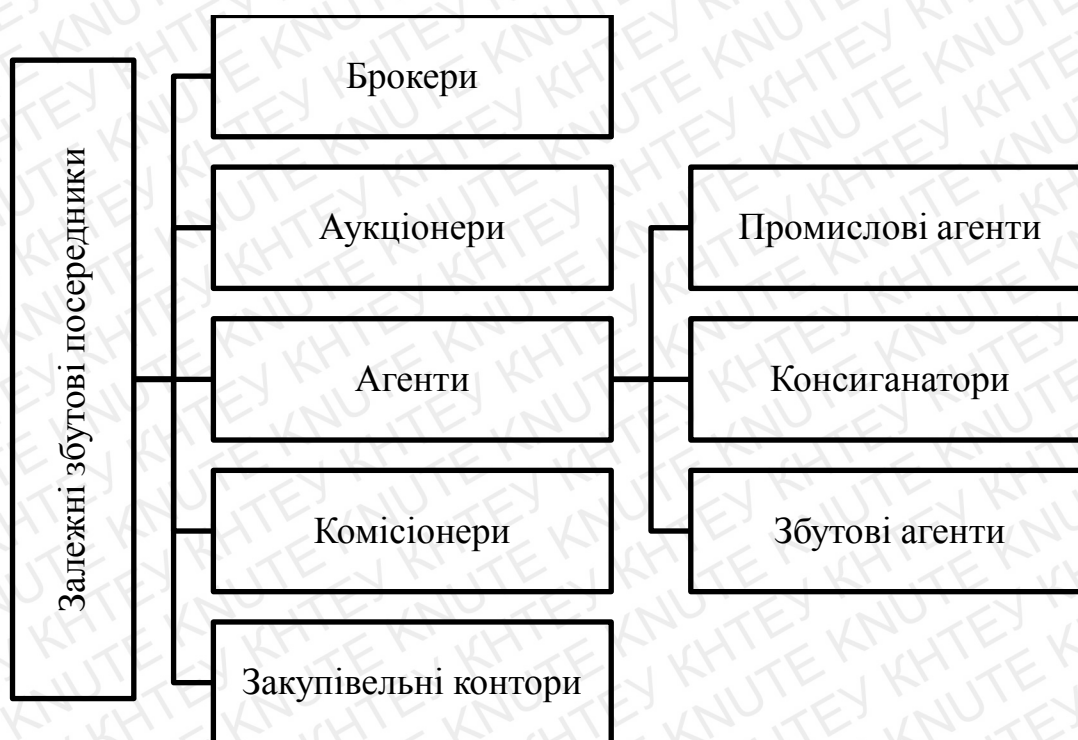


Рис. 1.2. Залежні посередники

Джерело: Розроблено автором на основі даних [7]

Причини швидких темпів змін у діяльності посередників пов'язані із збільшеним об'ємом інформації та швидкістю її передачі через Інтернет та електронну пошту. Сьогодні необхідно дуже швидко пристосовуватися до

нових умов ринку, новому типу споживачів. Економічна нестабільність також ускладнює діяльність посередників, змушуючи їх змінювати стратегію роботи на ринку. Наприклад, дистриб'ютори спеціалізованого напрямку прагнуть додати в асортимент нові товарні лінії, щоб зробити свій бізнес менш чутливим до змін ринкової кон'юнктури.

Оптові торговці діляться залежно від спектру виконуваних функцій на повнофункціональних і обмежено-функціональних.

Повнофункціональні оптові торговці надають повний цикл послуг для роздрібних і ділових споживачів. Вони здатні швидко поставити споживачам товар, мінімізуючи запаси. Оптовий торговець має свій штат продавців, які зв'язуються з роздрібними споживачами, організують доставку і надають кредит для клієнтів.

Повнофункціональні оптовики в основному займаються продажем лікарських препаратів, бакалійних виробів, на ринку ділових товарів реалізують невеликі машини й устаткування. До повнофункціональним оптовикам відносяться поличні оптовики (стелажні джоббери).

Джоббер - посередник на ринку, який подібно оптовикові збирає товар в загальну категорію від ряду виробників і продає роздрібним збутовикам. Джобери зберігають і поставляють роздрібним торговцям товари (засоби гігієни, іграшки), розміщують їх на полицях, оформляють місця продажів, призначають ціни, відвідують магазини для поповнення полиць своїми товарами.

Обмежено-функціональні оптові торговці діляться на чотири категорії: оптовики "плати і забирай", оптовик на вантажівці, пересильщиків за скороченим маршрутом, оптовики поштових замовлень.

Оптовик "плати і забирай" (Cash & Carry) виконує всі оптові функції, крім фінансування і доставки, продає товари поточного попиту дрібним магазинам та іншим роздрібним точкам за готівку. Клієнт приїжджає за товаром сам, оплачує його і відвозить в свою роздрібну мережу. Націнка німецької мережі Metro Cash & Carry на товари для ділових покупців (кафе,

дрібні торговці) становить близько 10%, в той час як у великих російських мережах для індивідуального покупця вона досягає 25-50%.

Оптовик на вантажівці займається реалізацією продукції з обмеженим терміном зберігання (хліб, молоко, фрукти, цукерки та ін.). Проводиться доставка товарів невеликими партіями за готівку в магазини, кафе, супермаркети.

Посередник за скороченим маршрутом приймає замовлення від споживачів і направляє їх виробникові, який у свою чергу пересилає замовлений продукт споживачеві. Посередник приймає на себе право власності на замовлені товари, несе ризики на період від моменту отримання замовлення до поставки продукції клієнту, але не зберігає запасів і не транспортує їх. Працює на ринку продажів громіздких товарів - лісоматеріалів, вугілля та ін.

Оптовик поштових замовлень розсилає поштою або за допомогою автомобільної доставки каталоги, які клієнти замовляють поштою або по телефону. Працює на ринку ювелірних виробів, косметики, спортивних товарів.

Зовсім оптові посередники можуть прийняти на себе зберігання і транспортування продукції, але ніколи не беруть право власності на товар. Головна функція - звести разом покупця і продавця.

1.2. Особливості ринків збуту і проблеми їх аналізу

Розуміння проблеми являє собою важливу частину аналізу, для виявлення проблеми було виокремлено наступні етапи:

а) Оцінка симптомів. Клієнти переважно сприймають проблеми як обставини, котрі з погляду певної групи осіб є небажаними. Вони, як правило, пояснюють проблеми експертам, радше покликаючись на ці небажані обставини, чи симптоми, ніж указуючи на глибинні причини. Завдання експерта – дати оцінку симптомам і пояснити (створити модель), як вони виникають. Оцінка

симптомів передбачає визначення їхньої емпіричної основи. У вузькому розумінні це означає спробу виявити дані, які допомагають розглянути симптоми в кількісній перспективі. Наприклад, якщо вашого клієнта турбують автомобільні аварії у вашому регіоні, спричинені нетверезими водіями, можна спробувати виявити дані, які допоможуть підрахувати кількість таких аварій, показник зміни цієї кількості з плином часу, процентне відношення до загального числа аварій та інші показники, що допомагають визначити масштаб і тенденцію розвитку цього симптому.

У широкому розумінні вам слід ознайомитися з поточним обговоренням цієї проблеми громадськістю та історією утвердження тих або інших варіантів політики, що загалом сприймаються як доречні в цьому випадку. Наприклад, можна виявити, що, незважаючи на стабільне зменшення числа аварій, викликаних водіями у нетверезому стані за останні роки, увагу громадськості до небезпеки водіння автомобіля в нетверезому стані привернула особливо трагічна автокатастрофа.

Ваша оцінка симптомів здебільшого виступає в аналізі проблеми як підоснова, вона передає відносну важливість і настійність проблеми та зміцнює ваш авторитет як особи, що розуміється на таких речах. Проте сама тільки оцінка симптомів – недостатня основа для вашого аналізу. Ви мусите ідентифікувати причинний зв'язок, що поєднує симптоми з чинниками, які можна змінити засобом державної політики. Інакше кажучи, ви мусите визначити і змоделювати проблему.

б) Визначення проблеми. Тут необхідно враховувати, що ефективність не є єдиною метою державної політики. Тому будь-яка реалістична структура "проблема-аналіз" має дозволяти експертові інтегрувати в аналітичний процес інші цілі та обмеження. Ця структура, крім того, має дати можливість експерту послідовно визначити ці цілі. Водночас вам слід починати з концентрації уваги на ефективності.

Перший крок означає з'ясування, чи є в даній ситуації неспроможності ринку. Вам треба визначити, чи ринок функціонує так, щоб задовольнити

індивідуальні вподобання, уникнути впливу на ціни окремими суб'єктами, використання ними асиметричності інформації тощо. Для цього знадобиться вивчення тем 4-10 нашого курсу. Якщо на неспроможність ринку не вказують ні теоретичні аргументи, ні емпіричні свідчення, доречно припустити, що ринок розподіляє рідкісні ресурси найефективнішим способом. Однак навіть за наявності ефективного ринку може зберігатися втручання держави, яке призводить до неефективності. Інакше кажучи, може виникнути ситуація, коли скасування державного втручання покращить ринкові розміщення.[3]

Конфлікти всередині каналів розподілу: зазвичай канал розподілу товарів і послуг складається з одного або декількох незалежних виробників, оптових і роздрібних торговців, кожен з яких представляє окремий бізнес і намагається максимізувати власний прибуток. У той же час ніхто з них не володіє достатнім контролем над іншими учасниками ланцюжка товароруку. Учасники каналів розподілу повинні тісно співпрацювати між собою з метою підтримки нормального прибутку і збільшення продажів, оскільки успіх кожного з них залежить від ефективної роботи всього каналу. Було б ідеально, якби всі ланки каналу товароруку працювали разом і злагоджено. У реальності ж учасники каналу розподілу, незважаючи на взаємну залежність, часто прагнуть діяти у власних інтересах, ігноруючи інтереси партнерів. Розбіжності з приводу цілей і функцій призводять до конфліктів всередині каналу розподілу.

Деякі конфлікти в каналі можуть приймати форму здорової конкуренції. Співпраця, раціональний розподіл ролей і управління конфліктами може бути забезпечений тільки сильним керівництвом.

Лідери каналів нерідко займаються вирішенням конфліктів між учасниками каналу і мережі розповсюдження. Канал може працювати успішно тільки за умови співпраці його учасників. Як тільки члени каналу починають конкурувати між собою, виникають конфлікти. Конфлікти між виробниками та посередниками пов'язані з розходженнями, що існують в основі їхньої діяльності. Виробник належить до торгівлі як до консультативного процесу, що включає в себе ряд стадій: встановлення контакту з клієнтом, презентація товару, зняття сумнівів

покупця, досягнення бажаного результату - придбання товару. Виробник упевнений, що конкурентні переваги кроються в якості товару, його характеристиках, наданні знижок і гарантій. Для дистриб'ютора конкурентні переваги пов'язані з встановленням взаємовигідних відносин з клієнтами, доступності продукту. Виробник концентрує увагу на ринку, цільових сегментах, дистриб'ютор - на клієнта, його потребах і перевагах. Відмінності в діяльності виробника і дистриб'ютора призводять до втрати довіри один до одного. Їхні стосунки повинні бути переглянуті таким чином, щоб відповідати мінливої ситуації. Конфлікти виникають на двох рівнях.[3]

Горизонтальний конфлікт маркетингових каналів виникає між членами каналів одного рівня (двома або більше оптовиками, двома дилерами). Великі оптовики можуть зажадати від постачальника великі знижки, погрожуючи відмовитися від збуту його виробів. Дрібні оптовики можуть об'єднуватися, вимагаючи таких же знижок для себе. Конкурувати можуть дилери, намагаючись витіснити зі своєї території конкурента. Наприклад, компанія "Вімм-Білл-Данн" протягом декількох років скоротила кількість дистриб'юторів з декількох десятків до восьми з через цінові війни, що мали місце серед посередників. За кожним з охоплених регіонів закріплений один з дистриб'юторів.

Досвід показує, що один дилер може забезпечити ті ж обсяги продажів, що були до того моменту, як він отримав виключне право збуту продукції на даному ринку. Наприклад, фірма, що реалізує взуття в Єкатеринбурзі через декількох дилерів, прийняла рішення надати ексклюзивне право па продаж взуття одному дилеру - місцевої компанії. Була надана короткострокова відстрочка платежу за товар. Місця продажу оформлялися відповідно до вимог підприємства постачальника. Компанія дилера не тільки утримала об'єми закупівель для потреб регіону, а й зуміла збільшити їх удвічі.

Часто конфлікти виникають між посередниками різних типів, що реалізують аналогічні товари. Наприклад, оргтехніку фірма може придбати через роздрібну мережу, дилера або Інтернет. Якщо інші посередники створюють жорстку

конкуренцію, дилер може відмовитися від збуту оргтехніки або зажадати особливих умов у виробника.

Продаж через Інтернет може призвести до скрутному становищу, оскільки прямі продажі посередники можуть розглядати як загрозу своєму бізнесу. Тому багато компаній використовують інтернет-сайти для надання інформації клієнтам, а для здійснення покупки направляють їх до дистриб'юторів.

Вертикальний конфлікт виникає між членами одного капала (виробником і ритейлером, дилером або дистриб'ютором). Виробник може відкрити фірмовий магазин, відбираючи у незалежних посередників частину обсягу продажів. Роздрібна мережа може розробити приватну марку товару, що конкурує з маркою виробника.[7]

Іноді конфлікти відбуваються в ситуаціях, коли фірма має декілька різних каналів для різних груп клієнтів, різних цінових ніш, типів використання. Для зниження конфліктів між каналами виробники можуть використовувати різні цінові орієнтири і надавати різну продукцію або бренди різних каналах.

Кожен учасник товаропровідної мережі повинен усвідомлювати себе ланкою єдиного ланцюжка реалізації, оскільки конфлікти в каналі неодмінно призведуть до краху всіх його учасників. Для зниження ризику виникнення конфліктів та підвищення ефективності дистрибуції створюються вертикальні маркетингові системи (ВМС). [10]

Підводячи підсумки, скажемо, що розуміння проблеми означає дослідження симптомів, що цікавлять певного клієнта, окреслення небажаних обставин через з'ясування неспроможності ринку та влади, і розробку моделей поведінки, що пов'язують обставини проблеми із змінними, якими можна маніпулювати через державну політику. Кожне з цих трьох завдань, і особливо моделювання, можна виконати тільки попередньо на старті аналізу. Як ми обговоримо нижче, кожний з інших етапів аналізу державної політики змушує переглядати початкове визначення проблеми.

1.3 Огляд методів аналізу ринків збуту товарів та послуг

Світ інформаційних технологій відповів на такий виклик створенням досить потужного набору інструментів, об'єднаних загальною назвою «Інтелектуальний аналіз даних» (Data Mining). В рамках цього підходу розроблені досить ефективні алгоритми, орієнтовані на пошук важливих, корисних і часто зовсім неочевидних закономірностей, прихованих у великих обсягах даних. Ці алгоритми спираються не тільки на статистичні ідеї, але також широко використовують останні досягнення в області машинного навчання і нейронних мереж. І якщо статистика дозволяє знаходити правильні відповіді на коректно поставлені питання, то інтелектуальний аналіз даних якраз і допомагає їх задавати.

Природно, виникає питання про причини настільки малого використання цих підходів на практиці. Основних причин дві:

1. Мала інформованість про можливості інтелектуального аналізу даних і про те, що багато прогнозів, що виконуються «на око», можна зробити більш повними і обґрунтованими;

2. Перебільшене уявлення про складність і вартості відповідного програмного забезпечення. На сьогоднішній момент можна знайти як безкоштовні пакети (Weka, RapidMiner і ін.), Апробовані засоби відомих виробників (Microsoft, Oracle, IBM і ін.), так і зручні і створені на пострадянському просторі рішення (Deductor і ін.). У будь-якому випадку, використання коштів інтелектуального аналізу даних на практиці показує високе відношення отриманого ефекту до витрат на впровадження.

І якщо друга причина носить чисто «технічний» характер, то перша - явно системний. Без подолання першого бар'єра технології інтелектуального аналізу даних не будуть затребувані білоруськими споживачами. Які ж можливості інтелектуального аналізу даних в торгівлі?

У застосуванні до ринків збуту товарів та послуг про методи інтелектуального аналізу даних згадують найчастіше в зв'язку з аналізом купівельної корзини, коли визначаються самі «ходові» поєднання товарів. Лише

тільки на перший погляд це завдання здається простою. Наприклад, якщо в номенклатурі присутні 10 000 товарів і мова йде тільки про виявлення тільки наборів з двох, то вже виходить 100 000 000 поєднань, які необхідно порівняти з наявною базою продажів. Але ж при цьому ще потрібно враховувати і те, що певні товари купуються часто і в будь-якому поєднанні. Рішення задач такого виду - звичайна сфера застосування аналізу асоціацій, як раз і орієнтованого на пошук корисних зв'язків між об'єктами. У число цих об'єктів можуть включатися не тільки товари, але, наприклад, час доби і день тижня, адже поєднання на кшталт «кефір + соус + вечір» теж виявляються корисними. Одні з найяскравіших прикладів вдалого використання аналізу купівельної корзини був представлений мережею супермаркетів Wall Mart в 1998-му році. Виявилось, що в п'ятницю ввечері добре продається поєднання пиво + підгузки. Причиною стала американська звичка купувати продукти в кінці тижня на кілька днів вперед, а основними «любителями» такої комбінації виявилися молоді батьки сімейств, що відправляються в магазин дружинами, зайнятими метушною з дітьми. До речі, реакція торгової мережі теж була адекватна: дорогі сорти пива перекочували на полиці поруч з памперсами. Приклад досить простий і виразний, але він наочно демонструє, як застосування аналізу асоціацій дозволяє торгової мережі збільшити виручку за певними групами товарів.[6]

Великі обсяги даних ускладнюють не тільки обробку інформації, а й усвідомлення отриманого результату. З цієї причини дані постійно розбиваються на невелику кількість щодо однорідних груп. На жаль, і ця проблема не завжди вирішується оптимальним чином. Так, клієнтська база часто розбивається на кілька груп за віковою ознакою (4-5 груп), потім - за рівнем доходів (3-4 групи). В результаті кількість груп йде на десятки, і немає впевненості, що не доведеться ще доробити групи, наприклад, за родом діяльності, або, навпаки, деякі з груп виявляються занадто схожими і їх варто об'єднати. Аналогічні операції проводяться і з товарною базою. Товари поділяють в залежності від рівня продажів (зазвичай на групи, що дають 50%, 30% і 20% від загального обсягу продажів), а також по стабільності попиту (до 10%, до 25% і понад 25% коливань від середнього рівня -

ABC -XYZ-аналіз). Головний недолік цих традиційних методів - штучні правила поділу, тобто спроба «розіпхати» дані по групах, певним заздалегідь. Що в такій ситуації нам можуть запропонувати технології аналізу даних? Відповідь така: більше коректна сегментація як клієнтської бази, так і товарної, проводиться на основі кластерного аналізу. При цьому до уваги приймаються всі значущі чинники і знаходяться природні групи. Заодно стає зрозумілим, на скільки груп слід розділити дані і переконатися в їх взаємному відмінності і внутрішньої однорідності. Як результат, на виході ми маємо, наприклад, адекватну сегментацію клієнтів конкретного підприємства і отримана інформація дасть основу для подальшого вдосконалення роботи профільних служб. Та й маркетингові дослідження стають простіше: достатньо взяти для дослідження невелике число представників кожного кластера і з меншими витратами сил, часу і грошей оцінити реакцію ринку на ту чи іншу маркетингову акцію, на новий продукт (послугу) або зміни в роботі.

З кластеризацією даних досить сильно пов'язана задача класифікації. Які покупці характеризуються лояльним ставленням до торгової мережі, а які збираються найближчим часом перейти на використання послуг конкурента? Хто поставиться позитивно, нейтрально або негативно до чергову новинку? Чим відрізняються комерційно успішні товари від інших? Цілий набір алгоритмів (дерева рішень, логістична регресія, висновок правил) намагаються відповісти на подібні питання на підставі вже наявних у підприємства даних, а не умоглядних висновків. Зазвичай отримані правила обчислюються десятками, але це вже доступне для огляду кількість чітко сформульованих тверджень, цілком готових до використання. Наприклад, набір правил виду «ЯКЩО Дохід з клієнта, тис. Руб. <5565,5 І Загальна кількість покупок> = 19 І Вік <54,5 І Звернень в службу підтримки <2 І Термін обслуговування <3 років, ТО Імовірність відгуку на рекламну розсилку <5% несе в собі чимало корисної інформації. Освоєння нового ринку, розширення асортименту або перенесення діяльності в інший регіон може вже плануватися більш точно і усвідомлено. [16]

Методи дослідження ринку:

Системний аналіз знаходить широке застосування в маркетингових дослідженнях, оскільки дозволяє розглядати будь-яку ринкову ситуацію як певний об'єкт для вивчення з великим діапазоном внутрішніх і зовнішніх причинно-наслідкових зв'язків. Так, зміни на ринку споживчих товарів можуть бути причиною, з одного боку, змін зовнішніх процесів, що відбуваються в сфері ринку засобів виробництва, фінансового та міжнародного ринків, а з іншого - внутрішніх процесів, а саме, змін у розвитку ринків окремих взаємопов'язаних товарів.

Комплексний підхід дозволяє досліджувати ринкову ситуацію, розглядаючи її як об'єкт, що має різні прояви. Наприклад, проблематика ринку окремого товару може бути пов'язана зі зміною попиту, товарної пропозиції або ціни, тобто з такими аспектами досліджуваного об'єкта (або ситуації), за допомогою яких можна визначити і прийняти стратегічні і тактичні рішення виходу з ситуації, що створилася. [18]

Системний аналіз і комплексний підхід тісно взаємопов'язані і, по суті, не можуть бути реалізовані один без іншого. Однак в конкретній ситуації необхідно досліджувати, по-перше, все її взаємозв'язку (внутрішні та зовнішні), а по-друге, всі сторони і аспекти її прояву (структуру, обсяг).

Програмно-цільове планування широко використовується при виробленні та реалізації стратегії і тактики маркетингу. Більш того, можна сказати, що маркетинг - це і є програмно-цільовий підхід до сфери ринку, на підставі якого будується вся планова маркетингова діяльність на підприємстві (програми або плани маркетингу).

Аналітично-прогностичні методи являють собою сукупність методів лінійного програмування, теорії масового обслуговування, теорії зв'язку, теорії ймовірностей, мережевого планування, методів ділових ігор, методів функціонально-вартісного аналізу, економіко-математичного моделювання, методів експертних оцінок. [18]

Лінійне програмування як математичний метод для вибору найбільш сприятливого рішення (з мінімальними витратами, максимальним прибутком,

найменшими витратами часу або зусиль) застосовується в маркетингу, наприклад, при розробці більш вигідного асортименту в умовах обмежених ресурсів, розрахунку оптимальної величини товарних запасів, планування маршрутів руху збутових агентів.

При вирішенні проблем вибору черговості обслуговування замовників, складання графіків поставок товарів та інших аналогічних завдань застосовуються методи теорії масового обслуговування, які дають можливість, по-перше, вивчити складаються закономірності, пов'язані з наявністю потоку заявок на обслуговування, і по-друге, дотримати необхідну черговість їх виконання.

Теорія зв'язку, яка розглядає механізм зворотних зв'язків, дозволяє отримати сигнальну інформацію про процеси, що виходять за межі встановлених параметрів. У маркетинговій діяльності використання підходу дає можливість управляти товарними запасами (регулювання надходженнями і відвантаженнями), процесами виробництва і збуту (ув'язка виробничих потужностей з можливостями збуту). Застосування теорії зв'язку в організаційних структурах маркетингової діяльності допомагає вдосконалювати зв'язок підприємств з ринком, підвищувати ефективність використання одержуваних даних. [18]

Методи теорії ймовірностей допомагають приймати рішення, які зводяться до визначення значення ймовірностей настання певних подій і вибору з можливих дій найбільш пріоритетним, наприклад, виробляти або не виробляти продукт А чи В, реорганізувати або розширювати виробництво, проникати або НЕ проникати на ринок і ін.

Метод мережевого планування дає можливість регулювати послідовність і взаємозалежність окремих видів робіт або операцій в рамках будь-якої програми, дозволяє чітко фіксувати основні етапи роботи, визначати терміни їх виконання, розмежовувати відповідальність, економити витрати, передбачати можливі відхилення. Цей метод ефективний при розробці програми виробництва нового товару і організації пробних продажів, підготовці і проведенні збутових і рекламних кампаній.

Дозволу реальних маркетингових ситуації в значній мірі допомагає метод ділових ігор, що дозволяє "програвати" при пошуку оптимальних варіантів спрощені моделі поведінки конкурентів і стратегії виходу на нові ринки. [16]

Для комплексного вирішення завдань, пов'язаних з підвищенням якості продукції, і одночасної економії матеріальних і трудових ресурсів використовується метод функціонально-вартісного аналізу. Наприклад, навіщо створювати зайву міцність товару, якщо він швидко застаріває морально і виходить з ужитку? Продукт виробництва як потенційна споживча вартість виявляється лише тоді, коли перетворюється в об'єкт споживання. [16]

Отже, предметом дослідження функціонально-вартісного аналізу повинні бути не тільки виробничий процес, а й вимоги ринку, причому процеси виробництва і вивчення ринку в рівній мірі можуть впливати один на одного. Більш значну економічну вигоду отримає те підприємство, яке, опанувавши мистецтвом маркетингу, застосує знання про ринок для формування такої структури виробничої програми, в якій буде віддана перевага продукції, що користується підвищеним попитом. У міру розвитку конкуренції взаємодія функцій виробництва і маркетингу ускладнюється. Застосування функціонально-вартісного аналізу дозволить підприємствам вести пошук таких раціональних рішень в області конструювання, технології, організації виробництва і обслуговування продукції, які забезпечать їй підвищену конкурентоспроможність.

Систему відомих або передбачуваних зв'язків між подіями, діями і процесами можна описати за допомогою методів моделювання. Більш ефективними є економетричні (економіко-математичні) моделі, які дають можливість з урахуванням діючих факторів ємності ринку визначити найбільш раціональні стратегії маркетингу та можливі кроки конкурентів, оцінити оптимальні витрати для отримання необхідного розміру прибутку.

Особливе місце в методах маркетингу займають методи експертних оцінок ("Дельфін", "Мозкова атака", "Адвокат диявола" і ін.), Які дозволяють досить швидко отримати відповідь про можливі процесах розвитку тієї чи іншої події на ринку, виявити сильні і слабкі сторони підприємства, оцінити ефективність тих чи

інших маркетингових заходів. Чистота проведення експертизи залежить від вирішення низки питань, пов'язаних з формуванням експертної групи, процедурою проведення експертизи, вибором методів обробки результатів експертних оцінок. Основні вимоги, що пред'являються до експертів, — це компетентність, професіоналізм, авторитетність, неупередженість. Процедура проведення експертизи передбачає генерування ідей на основі дискусій або опитувань (звичайних або багатоступеневих) із застосуванням анкет, а використання засобів обчислювальної техніки допомагає не тільки обробити отримані дані, а й побудувати Аналітичні і імітаційні моделі. [5]

Висновки до розділу 1

Поняття «Ринок збуту» є набагато ширшим, ніж просто економічна категорія. Адже збут товару є ланцюжком, що пов'язує підприємство-виробника із споживачем через проміжні ланки: збувальників, торгових посередників, ініціаторів покупки.

Для ефективної дії на споживача потрібне вживання особливих методів. Звичайно саме на стимулювання покладається задача пошуку прийомів, відповідних кожній стадії в процесі збуту.

Оскільки, головною метою будь-якого підприємства є отримання максимального прибутку, то збутовій діяльності повинно надаватися якомога більше уваги. На кожному підприємстві повинна бути добре організований служба маркетингу що займається шляхами просування своєї продукції на ринку вільної конкуренції. Основною метою роботи є аналіз ринку збуту товарів та послуг і розробка інструментарію по вдосконаленню діяльності, збільшення прибутків за рахунок пошуку та аналізу інформації, яку можна використати на практиці маркетологам для збільшення прибутків підприємства. Через це треба досягти мети роботи, адже технології Data Mining здатні допомогти підприємствам збільшити свою економічну силу на ринку у будь якій сфері діяльності.

РОЗДІЛ 2.

МЕТОДИ DATA MINING ДЛЯ ВИВЧЕННЯ РИНКІВ ЗБУТУ

2.1. Сутність і особливості методів Data Mining

Технологія інтелектуального аналізу даних (Data mining) може розглядатися як результат природної еволюції інформаційних технологій. По дорозі такої еволюції індустрія баз даних займалася розробкою наступних функціональностей: накопичення даних, керування даними (включаючи зберігання та вилучення, а також виконання транзакцій), а також аналіз даних (включаючи розробку сховищ даних і технологію інтелектуального аналізу даних).

З 1960-х років інформаційні технології послідовно еволюціонували від примітивних систем обробки файлів до складних, потужних систем управління базами даних. Дослідження в області баз даних з 1970-х років зміщувалися від ранніх ієрархічних і мережевих баз даних до реляційних СУБД, інструментів моделювання даних, а також до питань індексування і організації даних. На додаток люди отримали гнучкий і зручний доступ до даних за допомогою мов запитів (наприклад, SQL), призначені для користувача інтерфейси, оптимізовану обробку запитів і управління транзакціями. Ефективні методи онлайн-обробки транзакцій (on-line transaction processing - OLTP) внесли великий вклад в еволюцію і широке впровадження реляційної технології в якості головного інструмента ефективного зберігання, вилучення та управління великими обсягами даних. [11]

Технологія баз даних починаючи з середини 1980-х характеризувалася популяризацією, широким впровадженням, і концентрацією дослідницьких зусиль на нові, все більш потужні СУБД. З'явилися нові моделі даних, такі як об'єктно-орієнтовані, об'єктно-реляційні, дедуктивні моделі. Виникли предметно орієнтовані СУБД, включаючи просторові, временні, мультимедійні, наукові системи баз даних, бази знань, бази офісної інформації. Розглядалися питання, пов'язані з розподілом, диверсифікацією та поділом даних. З'явилися гетерогенні

системи баз даних, а також глобальні інформаційні системи, такі як Всесвітня Павутина - World Wide Web (WWW), які грають ключову роль в індустрії інформаційних технологій.

Дивно швидкий прогрес комп'ютерних апаратних засобів за останні сорок років привів до масового пропозицією потужних і доступних комп'ютерів і накопичувачів даних. Це сприяло сплеску індустрії інформаційних технологій і зробило величезну кількість баз даних і сховищ інформації доступними для управління транзакціями, вилучення даних та аналізу даних.[11]

Дані тепер можуть зберігатися в різних типах баз даних. Одна з недавно з'явилися архітектур - це сховища даних, репозиторій безлічі різнорідних джерел даних, організованих в рамках єдиної схеми в одному місці, призначений для прийняття управлінських рішень. Технологія сховищ даних включає очищення даних, інтеграцію даних, а також онлайнкову аналітичну обробку (On-Line Analytical Processing - OLAP), тобто технологію аналізу з такими функціональностями, як консолідація, агрегація, підведення підсумків, перегляд інформації "під різними кутами". Хоча технологія OLAP дозволяє проводити багатовимірний аналіз для прийняття рішень, для більш глибокого аналізу потрібні додаткові методи, такі як методи класифікації даних, кластерного аналізу, характеристизації змін даних в часі і т.д.

Надлишок даних і недолік в хороших методах їх аналізу приводив до ситуації багатства даними, але бідності інформацією. Швидко зростаючі обсяги накопичених даних швидко перевищили здатності людини в їх обробці. В результаті великі бази даних стали «могилами» даних - архівами, які рідко відвідуються. Як наслідок, важливі рішення приймаються не на основі інформаційно-насичених баз даних, а на основі інтуїції людини, яка приймає рішення, так як він не має відповідних інструментів для вилучення корисних знань з величезних обсягів даних. Технологія Інтелектуального Аналізу Даних дозволяє витягти корисні знання, важливі патерни, сприяючи вдосконаленню бізнес-стратегій, баз знань, наукових і медичних досліджень. [12]

Інтелектуальним аналізом даних ми будемо називати процес визначення

нових, коректних і потенційно корисних знань на основі великих масивів даних. В англійській літературі замість терміна «інтелектуальний аналіз даних» зазвичай використовується термін Data Mining (дослівний переклад - «видобуток даних»), а також близький термін Knowledge Discovery in Databases (KDD) - «Виявлення знань у великих базах даних».

Витягнуте знання в результаті інтелектуального аналізу даних ми будемо називати терміном патерн. Патерном може бути, наприклад, деякий нетривіальне твердження про структуру даних, про наявні закономірності, про залежності між атрибутами і т.д. Таким чином, завданням інтелектуального аналізу даних є ефективно витяг осмислених патернів з наявного масиву даних великого розміру. Для відсіву великої кількості можливих малокорисних патернів може вводитися функція корисності. В реальності оцінка корисності знання має суб'єктивний характер, тобто залежить від конкретного користувача. Можна виділити дві головні характеристики «цікавого» знання:

-Несподіванка. Знання «дивно» для користувача і потенційно несе нову інформацію.

-Застосовність. Користувач може використовувати нове знання для досягнення своїх цілей.

Цікаві знання, закономірності, Високорівнева інформація, отримані в результаті аналізу даних, можуть бути використані для прийняття рішень, контролю за процесами, управління інформацією та обробки запитів. Тому технологія інтелектуального аналізу даних розглядається як одна з найважливіших і перспективних тем для досліджень і застосування в галузі інформаційних технологій. [24]

Області застосування: наведемо деякі приклади областей, де велика кількість даних зберігається в централізованих або розподілених базах даних і вимагає аналізу:

Електронні бібліотеки, в яких систематизовано зберігаються тексти в різних форматах.

Архіви зображень, що складаються з великої кількості зображень в сирій

або стислій формі. До зображень може додаватися текст.

Бази даних геномних досліджень. Як відомо, організм людини складається з більш ніж 50000 видів генів і білків в різних поєднаннях. Дослідженням і інтерпретацією величезних баз даних, що виникли в результаті розшифровки генома людини, займається біоінформатика.

Медичні зображення. Велика кількість медичних відомостей мають вигляд зображень: ЕКГ, знімки внутрішніх органів і т.д. Аналіз цих зображень має велике значення для медицини.

Фінансові дані також є важливою сферою застосування методів інтелектуального аналізу даних. Ці дані являють собою котирування акцій, золота, ринкові індекси, процентні ставки, кредитні операції банків, транзакції по кредитних картах, виявлені шахрайські операції, і т.д.

Бази даних підприємств зазвичай зберігають докладні відомості про основні бізнес-операціях організації. Наприклад, відомості про клієнтів можуть представляти інтерес для вироблення маркетингової політики організації, політики утримання клієнтів, визначення індивідуальних переваг клієнтів.

Телекомунікаційні системи є джерелом таких даних, як історія викликів, збоїв, перевантажень, вмісту трафіку, і т.д.

Всесвітня Павутина містить величезний обсяг різномірної мультимедійної інформації різного типу. Її можна вважати найбільшою розподіленою базою даних, яка коли-небудь існувала в світі.

Біометричні дані людини (відбитки пальців, знімки осіб, і т.д.) знаходять все більше застосування в системах однозначної ідентифікації людини. Це породжує необхідність розвитку методів пошуку та аналізу в подібних базах даних.[28]

Кластерний аналіз: на відміну від класифікації та прогнозування, які аналізують безліч об'єктів навчальної вибірки, що має відомі мітки класу, кластеризація або кластерний аналіз аналізує об'єкти, у яких мітки класів невідомі. Кластеризація як раз покликана згенерувати ці мітки. Об'єкти кластеризуються або групуються на основі принципу максимізації внутрикласової близькості і мінімізації міжкласової близькості. Таким чином,

кластери об'єктів формуються так, що об'єкти одного кластера схожі один з одним, а об'єкти різних кластерів несхожі.

Кожен отриманий кластер може розглядатися як клас об'єктів, який в свою чергу може використовуватися в інших видах аналізу для отримання різних правил і закономірностей. [29]

Існують також методи ієрархічної кластеризації, які дозволяють організувати безліч спостережень в ієрархію класів.

Еволюційний аналіз: еволюційний аналіз даних описує і моделює регулярності і тренди для об'єктів, чия поведінка змінюється в часі. Незважаючи на те, що тут можуть застосовуватися розглянуті до цього характеристика і дискримінація, аналіз асоціацій, класифікація, кластеризація, у даного виду аналізу є відмінні риси і свої власні методи, які включають аналіз тимчасових рядів, аналіз послідовності та періодичності, пошук близькості.[30]

2.2. Вибір моделей Data Mining для сегментації ринку збуту товарів та послуг

Для більш ефективного аналізу ринку збуту товарів та послуг розглянемо існуючі моделі, розглянемо їх переваги та недоліки та оберемо модель, яка за своїми параметрами краще підходить для реалізації мети поставленої перед нами в даній роботі.

Класифікація Баєса: метод Баєса класифікації є статистичним методом. Він дозволяє передбачити ймовірність приналежності об'єкта до заданого класу. Метод Баєса класифікації заснований на теоремі Баєса, наведеної нижче. Перевагами методу є як точність, так і швидкість при роботі з великими масивами даних.

Нехай X - деякий об'єкт, клас якого невідомий. Нехай H - гіпотеза, яка полягає в тому, що X належить до класу C . Для проблеми класифікації ми хочемо визначити $P(H | X)$, ймовірність виконання гіпотези H при спостережуваних даних X .

Мовою теорії ймовірностей $P(H | X)$ - це ймовірність а posteriori настання

Н за умови Х. Наприклад, розглянемо в якості безлічі об'єктів фрукти, описуваних в базі даних кольором і формою. Припустимо, що Х - червоного кольору і круглої форми, а Н.

Гіпотеза, що Х - це яблуко. Тоді $P(H|X)$ - ступінь достовірності того, що Х — це яблуко при тому, що ми бачимо, що Х - червоне і кругле. У той же час $P(H)$ — це ймовірність а priori настання Н. Для нашого прикладу $P(H)$ - це ймовірність, що довільно взятий об'єкт з нашої бази даних буде яблуком. Ймовірність а posteriori $P(H|X)$ базується на більшій кількості інформації, ніж ймовірність а priori $P(H)$, яка не залежить від Х.

Аналогічно, $P(X|H)$ - це ймовірність а posteriori настання Х за умови Н. Тобто це ймовірність, що Х - круглої форми і червоного кольору при тому, що ми знаємо, що Х - яблуко.

Теорема Баєса говорить, що

$$P(H|X) = \frac{P(H)P(X|H)}{P(X)} \quad (2.1)$$

Розглянемо так званий наївний метод Баєса класифікації, як дуже простий і ефективний при великих розмірах бази даних. У ньому передбачається, що всі атрибути незалежні один від одного.

Нехай будь-який об'єкт заданий за допомогою n атрибутів, тобто об'єкт Х може бути представлений у вигляді вектора $X = (X_1, \dots, X_n)$. Припускаємо для простоти, що всі атрибути категоріальні, тобто можуть приймати лише кінцеве число значень. Нехай m — це кількість класів.

Теоретично, метод Баєса має мінімальну ступінь помилок в порівнянні з іншими класифікаторами. Однак на практиці це не завжди вірно, так як умова незалежності атрибутів - занадто сильне умова. Крім того, часто необхідних статистичних даних не вистачає для виконання класифікації. Проте, різні емпіричні дослідження і порівняння даного методу з деревами рішень і з нейронними мережами показують, що в ряді областей наївний метод Байєсова

класифікації цілком можна застосувати. [17]

Класифікація з навчанням: що таке класифікація з навчанням? Нехай є набір об'єктів, кожен з яких належить одному з m класів. Як приклад можна привести клієнтів банку, які можуть бути віднесені до класу сумлінних або недобросовісних позичальників, а також безліч танків на фотознімку, які можна розділити на «своїх» і «чужих». Завданням класифікації з навчанням є складання правила, за яким для будь-якого об'єкта можна з великим ступенем вірогідності визначити клас, якому даний об'єкт належить.

Нехай x_1, \dots, x_k - атрибути об'єкта, m - кількість класів. Внаслідок класифікації повинна бути отримана деяка функція $f(x_1, \dots, x_k)$, значення якої належить $1, \dots, m$, і задає номер (мітку) класу, якому належить об'єкт з атрибутами x_1, \dots, x_k . У розпорядженні у дослідника зазвичай є деякий набір об'єктів, у яких мітка класу вже відома. Ці об'єкти можуть бути використані для навчання моделі, тобто підбору параметрів моделі класифікації, і для тестування побудованої моделі класифікації.

Класифікація з навчанням має на увазі наступні дії:

1. Підготовка даних. Наявний набір об'єктів з відомими мітками класів розбивається на 2 частини: навчальну вибірку і тестову вибірку. Бажано, щоб це розбиття було вироблено випадковим чином. Найчастіше навчальна вибірка має розмір більше, ніж тестова.

2. Навчання моделі. Параметри моделі класифікації підбираються на основі навчальної вибірки таким чином, щоб домогтися максимальної відповідності умовам між передбаченими і фактичними мітками класів.

3. Тестування моделі. Отримана в результаті навчання модель перевіряється на достовірність. Для цього обчислюється відсоток невірних результатів класифікації об'єктів з тестової вибірки.

Класифікація з навчанням має безліч застосувань, наприклад, в таких областях, як кредитування, медична діагностика, прогноз доходів, маркетинг. Ми розглянемо три відомих методи класифікації з навчанням: дерева рішень,

нейронні мережі і метод Naive Bayes. [2]

Дерево рішень — це дерево, в якому кожної внутрішньої вершині поставлений у відповідність деякий атрибут, кожна гілка, що виходить з даної вершини, відповідає одному з можливих значень атрибута, а кожному листу дерева зіставлений конкретний клас або набір ймовірностей класів. Приклад дерева рішень, що дозволяє передбачити, чи є потенційний клієнт добросовісним позичальником, представлений на малюнку

Для того, щоб класифікувати новий об'єкт, необхідно рухатися по дереву зверху вниз, починаючи з кореня. При цьому на кожному внутрішньому вузлі дерева вибирається та гілка, яка відповідає фактичним значенням відповідного атрибута. Добравшись до листа дерева, отримуємо той клас, якому належить об'єкт згідно класифікуються правилом.[9]

Основна проблема полягає в тому, щоб побудувати досить гарне дерево рішень. Один з алгоритмів вирішення цього завдання, відомий як алгоритм ID3.

На кроці даного алгоритму використовується поняття інформаційного виграшу атрибута. Нехай навчальна вибірка S складається з s об'єктів, m - це кількість розглянутих класів, s_i - це число об'єктів з S , що належать класу з номером i . Кількість інформації, необхідне для того, щоб повідомити клас.[19]

Нейронні мережі: у мозку людини є близько 10^{11} нейронів - клітин, що відповідають за обробку даних. Нейрони пов'язані між собою численними сполуками - аксонами і дендритами. Біла речовина мозку складається з нейронів, а сіре - з аксонів і дендритів. Кожен нейрон отримує сигнал через безліч своїх дендритів, а передає результат його обробки через єдиний аксон, що розгалужується на безліч (тисячі) синапсів. Таким чином, мозок містить приблизно 1 015 взаємозв'язків. [13]

Такому пристрою мозку, як вважають нейрофізіологи, людина зобов'язана своїм розумом. Штучні нейронні мережі (Artificial Neural Networks) — це штучна обчислювальна система, що імітує поведінку біологічних нервових систем.

Завдання, які можуть вирішуватися за допомогою штучних нейронних мереж, включають завдання класифікації, кластерний аналіз, апроксимацію

функцій, завдання прогнозу, оптимізації, пошуку по вмісту і розпізнавання образів. Штучні нейронні мережі (ШНМ) можуть бути представлені, як зважені орієнтовані графи, в яких вершини відповідають нейронам, а орієнтовані ребра з вагами відповідають зв'язкам між виходами нейронів і входами нейронів.

За структурою зв'язків нейронні мережі можуть розділені на два класи:

1. Мережі прямого поширення: відповідний мережі граф не має петель, тобто зворотний зв'язок неможливі. Прикладами таких мереж є одношаровий перцептрон, багатошаровий перцептрон, мережі Кохонена.

2. Рекурентні мережі (мережі зворотного поширення): можливі цикли, а значить зворотні зв'язки. Прикладом є мережа Хопфілда.

Ми розглянемо застосування багатошарового перцептрона до задачі класифікації. Багатошаровий перцептрон складається з декількох шарів нейронів: вхідного шару, вихідного шару і декількох прихованих шарів.[22]

Нейронна мережа може розглядатися, як обчислювальна система, якій на вхід подається вектор введення, а результатом обчислень є вектор виведення. При цьому кожна компонента вектора введення подається через відповідний нейрон вхідного шару, а вектор виведення відповідає нейронам вихідного шару. [25]

Пошук асоціативних правил: асоціативні правила - це зв'язки між логічними атрибутами об'єктів. Дана глава присвячена методам пошуку цікавих з точки зору дослідника асоціативних правил в великих наборах даних. Розглядаючи методи стосовно відповідних довільної природи. Типовим прикладом області, в якій пошук асоціативних правил має важливе значення, є аналіз ринкового кошика. Розглянемо, наприклад, деякий супермаркет, в якому продається безліч товарів. Покупці вибирають необхідні їм товари, складають їх в кошик і потім оплачують.

Продавця цікавлять асоціацію між різними товарами, які покупець складає в кошик. Наприклад, інтерес може представляти питання, наскільки ймовірним є те, що покупець, який купив хліб, приобетет з ним і молоко? Які товари зазвичай купуються разом?

Така інформація може допомогти продавцям виробити маркетингову або рекламну стратегію. Наприклад, це допоможе:

- ефективно розташувати товари на території супермаркету;
- розробити систему знижок на одні товари для того, щоб стимулювати продажі інших товарів;
- вибрати ідею реклами на товар, використовуючи те, що цей товар зазвичай використовується спільно з іншим товаром. [14]

Іншими областями аналізу даних, де використовуються методи пошуку асоціативних правил, є: виявлення шахрайських операцій по кредитних картах, страховими випадками; визначення причин збоїв в телекомунікаційних системах; аналіз ДНК живих організмів; обробка даних соціологічних досліджень. і т.д.

Розглянемо базу даних, що складається з клієнтських транзакцій, де кожна транзакція характеризується великою кількістю елементів. Асоціативне правило формулюється зазвичай у вигляді:

$X \Rightarrow Y$, частота = $s\%$, достовірність = $c\%$, де X і Y - деякі безлічі елементів, s , c - числа від 0 до 100. Дане правило означає, що $c\%$ транзакцій, що містять елементи X , містять і елементи Y ; при цьому $s\%$ всіх транзакцій містять одночасно X і Y .

Сегментування споживчого ринку з урахуванням екологічних факторів може здійснюватися на підставі традиційних принципів:

1. Сегментування за географічним принципом передбачає розбивку ринку за різними географічними одиницями: держави, регіони, області, міста. Підприємство може вирішити діяти або в одному або декількох географічних районах або у всіх районах, але з урахуванням розбіжностей у потребах і перевагах, зумовлених географією.

2. Сегментування за демографічним принципом полягає в розбивці ринку на групи на основі таких демографічних даних як стать, вік, розмір родини, етап життєвого циклу родини, рівень доходів, рід занять, освіта, релігійні переконання, раса й національність.

3. Сегментування за психографічним принципом. При даному способі сегментування покупців поділяють на групи за ознаками приналежності до суспільного класу, способу життя чи певних характеристик особи.

4. Сегментування за поведінковим принципом - покупців ділять на групи залежно від їхніх знань, відносин, характеру використання товару й реакції на цей товар. Однак покупці екологічно чистих товарів являють собою досить складний об'єкт сегментації. Тут при сегментації ринку необхідно враховувати безліч факторів у їхній комбінації: регіональну демографію, вік, стать, родинний стан, професію, освіту, національність, релігійні переконання, житлові умови, добробут родини, сімейні переваги, екологічну ситуацію в країні проживання [24]

У таблиці 2.1 подані приклади суб'єктивних факторів і типу товару, що обирається, від яких залежить тип споживчого вибору.

Таблиця 2.1

Вплив суб'єктивних факторів на споживчий вибір

Фактор впливу	Емоційний вибір	Раціональний вибір
Тип товару	Публічне споживання (костюм)	Непублічне споживання (засоби миття для посуду)
Ціна	Низька ціна необхідного товару (йогурт)	Висока ціна необхідного товару (металеві вхідні двері)
Симбіоз суб'єктивного фактора й типу товару	Висока ціна товару розкоші/підтвердження статусу (прикраси)	Порівняльно низька ціна товару розкоші / підтвердження статусу (дорогий автомобіль)
Суб'єктивні фактори	Високий дохід споживача	Низький дохід споживача
Забезпечення потреб	Забезпечення психологічно обумовлених потреб	Забезпечення біологічно обумовлених потреб

При сегментації ринку також необхідно враховувати реакцію на товар:

швидко або повільна адаптація до нового екологічно чистого товару. При формуванні сегмента ринку з урахуванням екологічних переваг необхідно враховувати наступні ознаки:

- 1) подібність між споживачами, що формує стабільність переваг даної групи споживачів до даного товару;
- 2) відмінності між споживачами, що дозволяють об'єднати їх у сегмент;
- 3) наявність показників, що дозволяють оцінити економічні й екологічні вимоги й переваги споживачів. [24]

Таблиця 2.2

Типи мотивів споживання екологічних товарів

Тип мотивації	Приклад мотивів споживання
Раціональні мотиви	<ul style="list-style-type: none"> • якість • економічність • експлуатаційні параметри
Емоційні мотиви	<ul style="list-style-type: none"> • унікальні властивості (наприклад, найвища точність обробки матеріального верстата) • стиль життя (уживати тільки натуральні продукти) • відчуття страху (наприклад, якщо не придбати фільтр для питної води, то можуть утворитися камені в нирках) • відчуття провини (наприклад, придбання товарів, зроблених інвалідами)
Моральні мотиви	<ul style="list-style-type: none"> • збереження природного середовища (екологічність товару) • відчуття причетності (купуй вітчизняне)

Прийнявши рішення про сегментацію й позиціонування товару, підприємство може переходити до розробки комплексу маркетингу й розробляти новий більш конкурентний товар, ніж раніше пропонований на ринку.[24]

Кластерний аналіз (класифікація без навчання): нехай є безліч об'єктів для аналізу. Нехай при цьому, на відміну від випадку класифікації з навчанням, мітка класу для об'єктів не задана. Кластерний аналіз - це процес групування даних в класи або кластери таким чином, що об'єкти одного кластера мають високу схожість один з одним і високу ступінь відмінності від об'єктів інших класів. Ступінь відмінності між об'єктами визначається на основі значень атрибутів, що описують об'єкт.

Кластерний аналіз, або кластеризація - важлива діяльність людини. Ще в ранньому дитинстві людина вчиться розрізняти кішок і собак або тварин і рослин. Кластеризація широко застосовується в багатьох областях, включаючи статистику, біологію, машинне навчання, розпізнавання образів, маркетинг, політологію, соціологію і т.д. У бізнесі кластеризація може допомогти маркетологам визначити групи споживачів на основі наявної бази даних покупок. Кластеризація може допомогти ідентифікувати земельні площі однакового використання в базі даних спостережень земної поверхні. Страхові компанії можуть визначити групи власників полісів автострахування. Політичні партії на основі опитувань можуть виявити групи виборців. [23]

На відміну від класифікації з навчанням, кластерний аналіз не має справу з зумовленими класами. Немає також і навчальної вибірки, тобто прикладів з відомими мітками класів. Тому кластерний аналіз часто називають класифікацією без навчання.

Етапи кластерного аналізу наступні:

1. виявлення відповідних класів;
2. опис кожного отриманого класу.

На першому кроці застосовуються формальні алгоритми кластеризації, що розглядаються далі в цій главі.

Другий крок призначений для отримання характерних властивостей кожного класу. При цьому часто застосовуються розглянуті раніше алгоритми класифікації з навчанням, використовуючи мітки класів, отримані на першому кроці.

Типи даних в кластерному аналізі: в даному розділі ми вивчимо типи даних,

також комбінацією цих видів змінних. [26]

Цілочислові змінні (interval-scaled variables) — це кількісні вимірювання будь-яких властивостей. Наприклад, вага, зріст, тривалість, координати по вертикалі і по горизонталі, температура і т.д.

Використовуваний масштаб вимірювань при цьому може впливати на результати кластерного вимірювання. Наприклад, перехід в одиницях виміру від метрів на сантиметри, або від кілограм на фунти може привести до отримання зовсім іншої кластерної структури. Для уникнення такої залежності від вибору одиниць вимірювання дані повинні бути якимось чином стандартизовані. [15]

2.3. Побудова та оптимізація моделей аналізу ринку збуту

Для подальшої роботи над досягненням мети треба обрати модель аналізу ринку збуту та вивчити етапи її побудови та оптимізувати її під обраний маркетинг

Традиційно виділяються наступні етапи в процесі інтелектуального аналізу даних:

1. Вивчення предметної області, в результаті якого формулюються основні цілі аналізу.
2. Збір даних.
3. Попередня обробка даних:
 - Очищення даних — виняток протиріч і випадкових "шумів" з вихідних даних
 - Інтеграція даних — об'єднання даних з декількох можливих джерел в одному сховищі
 - Перетворення даних. На даному етапі дані перетворюються до форми, придатної для аналізу. Часто застосовується агрегація даних, дискретизація атрибутів, стиснення даних і скорочення розмірності.
4. Аналіз даних. В рамках даного етапу застосовуються алгоритми інтелектуального аналізу з метою вилучення патернів.

5. Інтерпретація знайдених патернів. Даний етап може включати візуалізацію витягнутих патернів, визначення дійсно корисних патернів на основі деякої функції корисності.

6. Використання нових знань.

Компоненти систем інтелектуального аналізу:

Зазвичай в системах інтелектуального аналізу даних виділяються наступні головні компоненти:

1. База даних, сховище даних або іншої репозиторій інформації. Це може бути одна або кілька баз даних, сховище даних, електронні таблиці, інші види репозиторіїв, над якими можуть бути виконані очищення і інтеграція. Види баз даних:

- Реляційні бази даних;
- Сховища даних;
- Транзакційні бази даних;
- Об'єктно-орієнтовані бази даних;
- Об'єктно-реляційні бази даних;
- Просторові бази даних (Spatial databases);
- Тимчасових бази даних (Temporal databases);
- Текстові бази даних;
- Мультимедійні бази даних;
- Різноманітні бази даних;
- Всесвітня павутина.

2. Сервер бази даних або сховища даних. Зазначений сервер відповідає за вилучення істотних даних на підставі призначеного для користувача запиту.

3. База знань. Це знання про предметну область, які вказують, як проводити пошук і оцінювати корисність результуючих патернів.

4. Служба видобутку знань. Вона є невід'ємною частиною системи інтелектуального аналізу даних і містить набір функціональних модулів для таких завдань, як характеристика, пошук асоціацій, класифікація, кластерний аналіз та аналіз відхилень.

5. Модуль оцінки патернів. Даний компонент обчислює заходи інтересу або корисності патернів.

6. Графічний користувальницький інтерфейс. Цей модуль відповідає за комунікації між користувачем і системою інтелектуального аналізу даних, візуалізацію патернів в різних формах. [10]

Парсинг(від англ. Parse) – процес аналізу або розбору певного контенту на складові за допомогою роботів-парсерів (спеціальних програм або скриптів). У SEO цим контентом є html-код сторінок сайтів.

Парсинг — це процес збору деяких даних і складання з них бази. Наприклад, можна зібрати базу гостьових книг. Або базу каталогів сайтів. Ось навіщо це робити і робити взагалі — кожен вирішує сам. Найчастіше такі бази використовуються спамерами. Але не обов'язково. [21]

Найвідоміші парсери в мережі – це пошукові роботи, які аналізують сторінки, зберігають дані аналізу у себе в базі і потім при пошуку видають релевантні та актуальні документи.

Часто парсинг плутають з граббінгом. Це близькі поняття, але все ж мають різні значення. Граббер дозволяє скачувати інформацію з мережі (html-сторінки, rss-стрічки, xml-документи) в свою базу, а парсер дозволяє виявити з цієї купи корисну інформацію і обробити її, залежно від поставлених завдань.

Синтаксичний аналізатор — це програмний компонент, який приймає вхідні дані (часто текст) і створює структуру даних — часто дерево розбору, абстрактне дерево синтаксису або іншу ієрархічну структуру — забезпечує структурне представлення вводу, перевіряє правильність синтаксису в процесі. Для аналізу можуть передувати або слідувати інші кроки, або їх можна об'єднати в один крок. Аналізатору часто передує окремий лексичний аналізатор, який створює токени з послідовності введених символів; Крім того, їх можна об'єднати у парсинг без сканування. Аналізатори можуть бути запрограмовані вручну або автоматично або напів автоматично генератором парсерів. Розбір допомагає шаблону, який виробляє відформатований вихід. Вони можуть використовуватись у різних

ділянках, але часто з'являються разом, наприклад, пара `scanf/printf`, або як вхідний (аналіз вхідних даних) та вихідний етапи (створення кінцевого коду) компілятора.

Вхідними даними для синтаксичного аналізатора часто є текст на деякій комп'ютерній мові, але також може бути текстом природною мовою або менш структурованими текстовими даними, в цьому випадку, як правило, витягуються лише окремі частини тексту, а не дерево розбору. Параметри відрізняються від дуже простих функцій, таких як `scanf`, до складних програм, таких як інтерфейс компілятора C++ або HTML-аналізатор веб-браузера. Важливий клас простий синтаксичний аналіз виконується за допомогою регулярних виразів, в яких група регулярних виразів визначає регулярну мову та двигун регулярного виразу, автоматично генеруючи аналізатор для цієї мови, що дозволяє узгодити шаблон та вилучення тексту. В інших контекстах регулярні вирази замість цього використовуються перед розбором, як етап лексизації, вихід якого потім використовується аналізатором.[14]

Використання аналізаторів залежить від вхідних даних. У випадку з мовами даних часто використовується синтаксичний аналізатор як функція читання файлів у програмі, наприклад, читання в HTML або XML-тексті; ці приклади є мовами розмітки даних. У випадку мов програмування є компонентом компілятора або інтерпретатора, який аналізує початковий код мови комп'ютерного програмування для створення певної форми внутрішнього представлення; аналізатор є ключовим кроком в інтерфейсі компілятора. Мови програмування, як правило, вказуються в термінах детерміністичної контекстно-вільної граматики, оскільки для них можуть бути написані швидкі та ефективні аналізатори. Для компіляторів сам аналіз може бути виконаний за один прохід або кілька проходів — див. одно-прохідний компілятор і багатопрохідний компілятор.

Майбутні недоліки компілятора з одним прохідним процесом у значній мірі можуть бути вирішені шляхом додавання виправлень, коли передбачається виправлення впродовж прямого переходу, а виправлення застосовуються в зворотньому напрямку, коли поточний сегмент програми є таким, що має бути завершений. Приклад, коли такий механізм виправлення може бути корисним,

буде формальним твердженням GOTO, де ціль GOTO невідома, доки не буде пройдено сегмент програми. У такому випадку застосування виправлення буде відкладено, доки не буде визначено куди вказує GOTO. Очевидно, що відсталий GOTO не вимагає виправлення. [15]

Контекстні граматики обмежені в тій мірі, в якій вони можуть виразити всі вимоги до мови. Неформально, причиною є те, що пам'ять в такій мові обмежена. Граматика не запам'ятовує наявності конструкції над довільним введенням; це необхідно для мови, в якій, наприклад, ім'я повинно бути оголошено, перш ніж може бути посилання на нього. Однак більш потужні граматики, які можуть обійти це обмеження, не можуть бути ефективно розібрані. Таким чином, загальною стратегією є створення аналізатора для контекстно-вільної граматики, який приймає потрібні конструкції мови (тобто він приймає деякі недійсні конструкції); пізніше, небажані конструкції можуть бути відфільтровані на етапі семантичного аналізу (контекстного аналізу).

Пошуковий робот або Web Crawler - програма, що є складовою частиною пошукової системи і призначена для перебору сторінок Інтернету з метою занесення інформації про них в базу даних пошукової системи.

За принципом дії, павук нагадує звичайний браузер. Він аналізує вміст сторінки, зберігає його в деякому спеціальному вигляді на сервері пошукової машини, якій належить, і відправляється по посиланнях на наступні сторінки.

Власники пошукових машин нерідко обмежують глибину проникнення павука всередину сайту і максимальний розмір сканованого тексту, тому надто великі сайти можуть виявитися не повністю проіндексованими пошуковою машиною. Крім звичайних павуків, існують так звані «дятли» - роботи, які «простукують» проіндексований сайт, щоб визначити, що він доступний.

Порядок обходу сторінок, частота візитів, захист від зациклення, а також критерії виділення значущої інформації визначаються алгоритмами інформаційного пошуку.

У більшості випадків перехід від однієї сторінки до іншої здійснюється за такими адресами, що містяться на першій і наступних сторінках.

Також багато пошукові системи надають користувачеві можливість самостійно додати сайт в чергу для індексування. Зазвичай це істотно прискорює індексування сайту, а у випадках, коли ніякі зовнішні посилання не ведуть на сайт, взагалі виявляється практично єдиною можливістю вказати на його існування. Ще одним способом швидкої індексації сайту є додавання на сайт систем веб-аналітики, що належать пошукових сервісів. Наприклад, таких як Google Analytics. Обмежити індексацію сайту можна за допомогою файлу robots.txt. Повний захист від індексації може бути забезпечена іншими механізмами, наприклад установкою пароля на сторінці або вимогою заповнити реєстраційну форму перед тим, як отримати доступ до вмісту. [15]

Робот не аналізує знайдений контент, він тільки передає його на сервера пошукової системи, де вже відбувається його обробка.

Висновки до розділу 2

В аналітичному огляді розглянуті основні методи Data Mining, такі як асоціативні правила, нейромережеві алгоритми, дерева рішень, алгоритми кластеризації, теорема Баєса, класифікація за навчанням, пошук асоціативних правил, розглянуті такі методи, як парсинг та технологія пошукового робота (Web Crawler). Були описані ключові етапи процесу інтелектуального аналізу даних, компоненти систем інтелектуального аналізу.

Data Mining має вагомим практичне значення як технологія, призначена для вирішення актуальних повсякденних проблем, пов'язаних з пошуком і аналізом великих масивів даних. Великі дані здатні змінити наш спосіб життя, праці й мислення.

Однією з умов успішного розвитку ринку збуту товарів та послуг на сучасному етапі стає можливість фіксувати й аналізувати величезні масиви і потоки інформації за допомогою технології Data Mining.

У результаті проведених досліджень можна дійти висновку що створених технологій і методів, для реалізації поставлених цілей перед аналізом ринків

збуту, вистачає. Для досягнення поставленої мети було вирішено створити пошуковий робот який буде збирати інформацію з web-сторінок та сортувати її для легшого аналізу ринку збуту товарів та послуг.

РОЗДІЛ 3. ПРОГРАМНА РЕАЛІЗАЦІЯ МЕТОДІВ DATA MINING НА РИНКУ ЗБУТУ

3.1. Вибір інструментарію Data Mining для реалізації моделі аналізу

Для створення інструменту, який може допомогти винести більше корисної інформації з обраного ринку збуту було вирішено створити власний Web Crawler. Для створення було обрано мову програмування Python.

Python - інтерпретована об'єктно-орієнтована мова програмування високого рівня зі строгою динамічною типізацією. Структури даних високого рівня разом із динамічною семантикою та динамічним зв'язуванням роблять її привабливою для швидкої розробки програм, а також як засіб поєднування наявних компонентів. Python підтримує модулі та пакети модулів, що сприяє модульності та повторному використанню коду. Інтерпретатор Python та стандартні бібліотеки доступні як у скомпільованій, так і у вихідній формі на всіх основних платформах. В мові програмування Python підтримується кілька парадигм програмування, зокрема: об'єктно-орієнтована, процедурна, функціональна та аспектно-орієнтована. Серед основних її переваг можна назвати такі:

- чистий синтаксис (для виділення блоків слід використовувати відступи);
- переносність програм (що властиве більшості інтерпретованих мов);
- стандартний дистрибутив має велику кількість корисних модулів (включно з модулем для розробки графічного інтерфейсу);
- можливість використання Python в діалоговому режимі (дуже корисне для експериментування та розв'язання простих задач);
- стандартний дистрибутив має просте, але разом із тим досить потужне середовище розробки, яке зветься IDLE і яке написано на мові Python;
- зручний для розв'язання математичних проблем (має засоби роботи з комплексними числами, може оперувати з цілими числами довільної величини, у діалоговому режимі може використовуватися як потужний калькулятор);

Python має ефективні структури даних високого рівня та простий, але ефективний підхід до об'єктно-орієнтованого програмування. Елегантний синтаксис Python, динамічна обробка типів, а також те, що це інтерпретована мова, роблять її ідеальною для написання скриптів та швидкої розробки прикладних програм у багатьох галузях на більшості платформ. [20]

Багата стандартна бібліотека є однією з привабливих сторін Python. Тут є засоби для роботи з багатьма мережевими протоколами та форматами Інтернету, наприклад, модулі для написання HTTP-серверів та клієнтів, для розбору та створення поштових повідомлень, для роботи з XML, тощо. Набір модулів для роботи з операційною системою дозволяє писати крос-платформні застосунки. Існують модулі для роботи з регулярними виразами, текстовими кодуваннями, мультимедійними форматами, криптографічними протоколами, архівами, серіалізацією даних, юніт-тестуванням та ін. [4]

Web Crawler або пошуковий робот сайтів (пошукових систем) – це автоматична програма, яка здатна відвідувати мільйони веб-сторінок, швидко переміщуючись по інтернету без втручання оператора. Боти постійно сканують простір Всесвітньої павутини, знаходять нові інтернет-сторінки і регулярно відвідують вже проіндексовані. Інші назви пошукових роботів: павуки, краулери, боти.

Основна функція, яку виконують пошукові роботи, – індексація веб-сторінок, а також текстів, зображень, аудіо - та відеофайлів, що знаходяться на них. Боти перевіряють посилання, дзеркала сайтів (копії) і оновлення. Роботи також здійснюють контроль HTML-коду на предмет відповідності нормам Світової організації, яка розробляє та впроваджує технологічні стандарти для Всесвітньої павутини. [21]

У загальних рисах принцип роботи робота полягає в наступному: програма «приходить» на сайт по зовнішніх посиланнях і, починаючи з головної сторінки, «читає» веб-ресурс (в тому числі переглядаючи ті службові дані, які не бачить користувач). Бот може переміщатися між сторінками одного сайту, так і переходити на інші. Як програма вибирає, який індексувати сайт? Найчастіше

«подорож» павука починається з новинних сайтів або великих ресурсів, каталогів і агрегаторів з великою посилальною масою. Пошуковий робот безперервно сканує сторінки одну за одною, на швидкість і послідовність індексації впливають наступні фактори:

- внутрішні: перелиновка (внутрішні посилання між сторінками одного і того ж ресурсу), розмір сайту, правильність коду, зручність для користувачів і так далі;
- зовнішні: загальний обсяг посилальної маси, яка веде на сайт.

Всупереч поширеній думці, "павуки" не обробляють отриману інформацію. Програма сканує тільки і зберігає веб-сторінки, а подальшою обробкою займаються зовсім інші роботи. Також багато користувачів вважають, що пошукові роботи чинять негативний вплив і «шкідливі» інтернету. Дійсно, окремі версії "павуків" можуть значно перевантажувати сервера. Має місце і людський фактор – веб-майстер, який створював програму, може допускати помилки в налаштуваннях робота. Все ж більшість діючих програм добре спроектовані і професійно управляються, а будь-які виникаючі проблеми оперативно усуваються. [9]

Основним ринком збуту була обрана індустрія комп'ютерних ігор, а саме, найбільша платформа для гравців Steam. Steam — сервіс компанії Valve, відомого розробника відеоігор, який надає послуги цифрової дистрибуції, багатокористувацьких ігор і спілкування гравців. Steam використовується також для отримання автоматичних оновлень і новин, як для продуктів самої Valve, так і сторонніх розробників, серед яких є Activision, Codemasters, Eidos Interactive, Epic Games, GSC Game World, id Software, Sega, THQ, Atari, Rockstar Games, Ubisoft, NCSoft та багато інших. Крім відеоігор, сервіс пропонує інструменти для розробників ігор, музику, фільми та геймерське обладнання.

Станом на серпень 2017, через Steam поширюється більше 23000 товарів, а кількість активних користувачів досягла 125 мільйонів, з них 33 мільйонів активних. Щодня сервісом у середньому користуються 14 мільйонів чоловік. Хоча

Valve ніколи не висвітлювала особисто показники продажів, підраховано, що їхній сервіс охоплює 70 % ринку цифрової дистрибуції відеоігор.

Всі користувачі сервісу мають власну сторінку, на якій розміщують дані про себе, переглядають новини друзів і груп. Сторінки мають елементи відеогри, такі як рівні користувача. Кожна куплена відеогра впродовж грання дає картки, збір яких підвищує рівень. Крім того його підвищує виконання базових дій, таких як перша покупка, перше опублікування скриншоту, коментар і т. д. Виконавши низку «завдань» з освоєння у сервісі та збираючи комплекти карток, користувач додатково здобуває значки, які є предметом колекціонування.

Збираючи картки, гравець добуває додаткові матеріали для свого профілю, такі як оформлення, смайли для чату, а також купони на знижки. Все це складається до так званого інвентаря. Самі ігри мають систему досягнень, які спонукають повертатися до гри аби здобути їх. Вони дають стимул до змагання між гравцями у їх зборі, але не приносять фінансової вигоди або внутрішньоігрової винагороди.

Статус кожного учасника відображається на його сторіночці: В мережі, Немає на місці, Не турбувати, Бажає грати, Бажає обмінятись, Поза мережею, В грі (з вказівкою назви гри). Також на особистій сторінці відображається статистика часу гри за останні два тижні, окремо для кожної гри. Час, проведений за не пов'язаною зі Steam грою, в статистиці не показується.

Виходячи з цієї інформації було вирішено створити краулер, який буде обробляти відкриту інформацію з особистих сторінок користувачів сервісу, адже всі користувачі є цільовою аудиторією для розробників майбутніх продуктів, які будуть продавати на даній платформі. Компанія Valve має власний сервіс під назвою Steam Database, на якому висвітлює величезні об'єми даних що стосуються абсолютно різної статистики яку збирає Steam(кількість активних користувачів, графіки зі статистикою будь-якої гри на сервісі, вся інформація про платформи, відео-карти, процесори, кількість оперативної пам'яті, яку використовують покупці), але немає статистичної інформації стосовно особистих сторінок користувачів.

3.2. Проектування та програмна реалізація

На основі розглянутих теоретичних засадах було розроблено програмне забезпечення, яке має на меті зібрати дані про користувачів скануючи інформацію яку висвітлює сам користувач на своїй особистій сторінці.

Для пошуку імені на сторінці було розроблено наступний алгоритм зображений на рисунку 3.1.

```
name = find_item(self.re_name, html1, "user name", self.current_user)
if check_existence:
    if len(self.existlist) >= EXIST_LIST_SIZE: self.existlist = self.existlist[1:]
    self.existlist+= "0" if name == None else "1"
if name == None: return 0
```

Рис. 3.1 Пошук імені користувача [Розробка автора]

Для збору статистики про SteaID користувачів було створено наступний алгоритм зображений на рисунку 3.2. Дані про SteaID та особливу (зміну) URL адресу користувача доступні всім, не залежно від налаштувань профілю користувача, адже свої персональні сторінки іноді користувачі приховують і дістати інформацію завдяки траулеру неможливо.

```
steamid = find_item(self.re_steamid, html1, "Steam id", self.current_user)
c_user_id = user_url_to_user(self.current_user) if self.current_user[0] == 'p' else None #profiles/...
if steamid:
    if c_user_id and steamid != c_user_id:
        print "Mismatch: " + self.current_user + " != " + steamid
    else: steamid = c_user_id if c_user_id else None
if not steamid: return 0
```

Рис. 3.2 Пошук SteamID [Розробка автора]

Кожен користувач може редагувати свій власній SteamID, таким чином велика кількість власників особистих сторінок створюють собі унікальні URL адреси, щоб їх сторінки було легше знайти, та URL адреса сторінки легше сприймалась наочно. На рисунку 3.3 показано код який відрізняє та записує

інформацію у тих випадках коли користувач змінив свою стандартну URL адресу на унікальну.

```
if self.current_user[0] == 'p': #profiles/...
    customurl = find_item(self.re_customurl, html1, "Custom URL", self.current_user)
    if customurl and r"\/id\/" in customurl:
        self.current_user = "id/" + customurl.split("/")[-2]
```

Рис. 3.3 Пошук унікальної URL адреси профілю [Розробка автора]

Через особливі налаштування приватності особистої сторінки користувача багато даних можуть бути приховані від інших. Проте, з прихованих сторінок можна винести корисну інформацію яка наявна всім, це стосується зображення профілю (його наявність або відсутність), унікальна URL адреса (стандартне URL або унікальне), псевдонім гравця (користувач має можливість змінювати свій псевдонім безліч разів). Вся інша інформація доступна тільки на відкритих сторінках, така як – рівень профілю, кількість друзів, унікальний фон сторінки чи наявність VAC блокування.

Valve Anti-Cheat (скорочено VAC) — розробка компанії Valve для запобігання використанню гравцями нечесних прийомів гри розроблена компанією Valve, як компонент платформи Steam.

Гравці з таким блокуванням не мають можливості грати в деякі ігри, які існують на платформі Steam.

Наступна інформація, яка буде збиратися та аналізуватися за допомогою власної розробки доступна у випадку відкритого профілю користувача. Щоб запобігти скануванню прихованих сторінок було створено алгоритм який буде відрізняти приховані сторінки від публічних. Код розробки алгоритму зображено на рисунку 3.4.


```

if not private_profile:
    #parts of the page
    #left and right cols
    leftcol_index = html1.find(self.se_leftcol)
    rightcol_index = html1.find(self.se_rightcol)
    #right col
    html1_right = ""
    if rightcol_index == -1:
        print "Couldn't find right collumn for " + name
    elif leftcol_index < rightcol_index:
        html1_right = html1[rightcol_index:]
    else: html1_right = html1[rightcol_index:leftcol_index]
    #left col
    html1_left = ""
    if leftcol_index == -1:
        print "Couldn't find left collumn for " + name
    elif rightcol_index < leftcol_index:
        html1_left = html1[leftcol_index:]
    else: html1_left = html1[leftcol_index:rightcol_index]
    #comments
    html1_comments = ""
    comments_index = html1_left.find(self.se_comments)
    if comments_index == -1:
        print "Couldn't find comments for " + name
    else: html1_comments = html1_left[comments_index:]
    #top friends
    html1_topfriends = ""
    topfriend_index = html1_right.find(self.se_topfriends)
    if topfriend_index == -1:
        print "Couldn't find top friends for " + name
    else: html1_topfriends = html1_right[topfriend_index:]

```

Рис. 3.4 Робота з не приватними сторінками [Розробка автора]

У кожного власника особистої сторінки може бути особливий фон профілю або його відсутність. Користувачі не можуть обирати власні зображення, тільки серед запропонованих в середині сервісу, або придбавши на внутрішній торговельному маркеті. Інформація про наявність фону та інформацію саме про зображення записує алгоритм зображений на рисунку 3.5.

```

has_background = html1.find(self.se_background) != -1
if has_background:
    bg_image = find_item(self.re_bgimage, html1, "background", name, 2)
    if bg_image:
        bg_image = trim_bgurl(bg_image)
        if bg_image in self.bg_images:
            bg_image = self.bg_images.index(bg_image) + 1
        else:
            self.bg_images.append(bg_image)
            game = bgurl_to_game(bg_image)
            if game not in self.games: self.games_queue.append(game)
            bg_image = len(self.bg_images)
    else: bg_image = 0
else: bg_image = 0

```

Рис. 3.5 Наявність фону особистої сторінки [Розробка автора]

Кожен користувач має особистий рівень своєї сторінки, він створений для боротьби проти шахрайства, у сторінок з низьким рівнем обмежені можливості на ресурсі. Інформацію про рівні користувачів збирає алгоритм зображений на рисунку 3.5.

```

level = int(find_item(self.re_level, html1, "level", name))
if level: items = [level]
else: items = [0]

```

Рисунок 3.6 Рівень особистої сторінки користувача [Розробка автора]

Для кожного показника що збирає розроблений пошуковий робот створено окремий алгоритм, що визначає вищі, різного роду, середні показники, та їх відхилення для параметрів від всіх попередніх алгоритмів. Алгоритми визначення зображені на рисунках 3.7, 3.8, 3.9.

```

def get_median(self):
    s = sorted(self.n.items(), key = itemgetter(0))
    start = 0
    startamount = s[start][1]
    end = len(s) - 1
    endamount = s[end][1]
    while start < end:
        amount = min(startamount, endamount)
        startamount -= amount
        endamount -= amount
        if(startamount == 0):
            start += 1
            startamount = s[start][1]
        if(endamount == 0):
            end -= 1
            endamount = s[end][1]
    self.median = s[start][0]
    return self.median

```

Рис. 3.7 Визначення медіани [Розробка автора]

```

def get_mean_deviation(self):
    if self.median == None: get_median()
    return float(sum(abs(i - self.median) * self.n[i] for i in self.n)) / float(self.count)

```

Рис. 3.8 Визначення абсолютного відхилення[Розробка автора]

```

def get_relative_deviation(self):
    if self.mean == None: get_mean()
    if self.mean == 0: return 0
    return 100.0 * math.sqrt(float(sum((i - self.mean) ** 2 * self.n[i] for i in self.n)) / float(self.count)) / self.mean

```

Рис. 3.9 Визначення відносного стандартного відхилення (коефіцієнт варіації)

[Розробка автора]

3.3. Результати використання програмної розробки

В результаті створення програмного продукту було зібрано та опрацьовано інформацію з 7 941 457 сторінок користувачів платформи Steam. Отримані результати несуть в собі користь для розробників ігор, які планують продавати

свій продукт на даній платформі. Зібрана інформація, яка буде проілюстрована у розділі, допоможе дізнатися про своїх потенційних клієнтів значно більше інформації, яку можна використовувати для збільшення продажів свого продукту за рахунок кращої взаємодії з аудиторією.

Інформація, яка була зібрана за рахунок виконання коду зображеного на рисунках 3.1, 3.2, 3.3, 3.4 та 3.5 зображена на рисунку 3.10. Варто зазначити, що результати поля «Наявність не стандартного фону» стосується відкритих сторінок користувачів, яких, в свою чергу, 85,4%.



Рис. 3.10 Головна інформація про сторінку [Розробка автора]

Наступним кроком було пророблено роботу про збору даних що стосуються псевдонімів користувачів, відсортовано за популярністю. На рисунках 3.11, 3.12, 3.13 зображено пої популярніші псевдоніми, використані слова та «чисті» псевдоніми.

Окремим класом було виокремлено алгоритм для визначення відносного стандартного відхилення. Отримані результати свідчать про те що мінливість даних які отримує алгоритм дуже висока на всіх показниках:

- Рівень профілю Steam – 112,5%;
- Кількість ігор – 245%;
- Кількість спільнот – 244,9%;
- Кількість друзів – 104%;
- Кількість змін псевдоніму протягом дня – 307,5%.

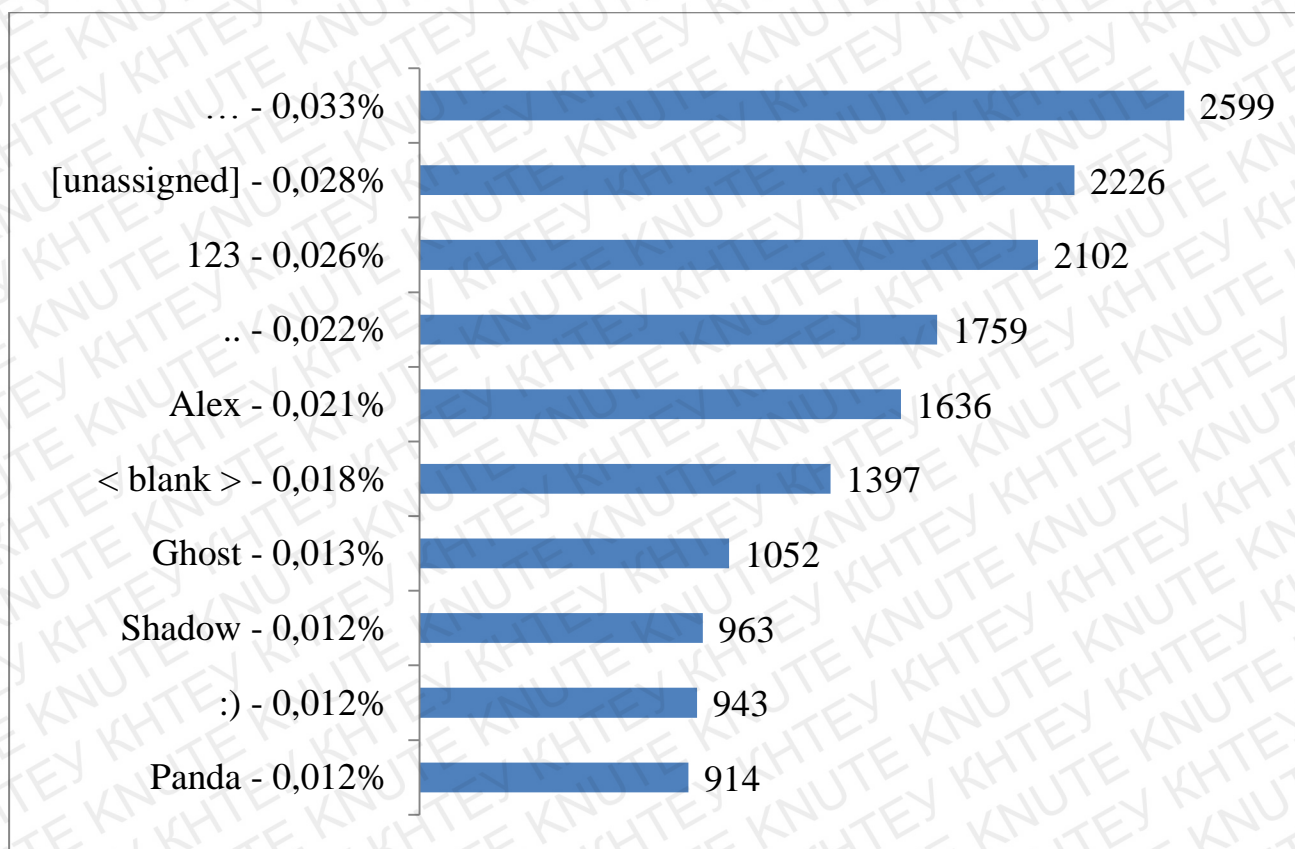


Рис. 3.11 Найпопулярніші псевдоніми [Розробка автора]

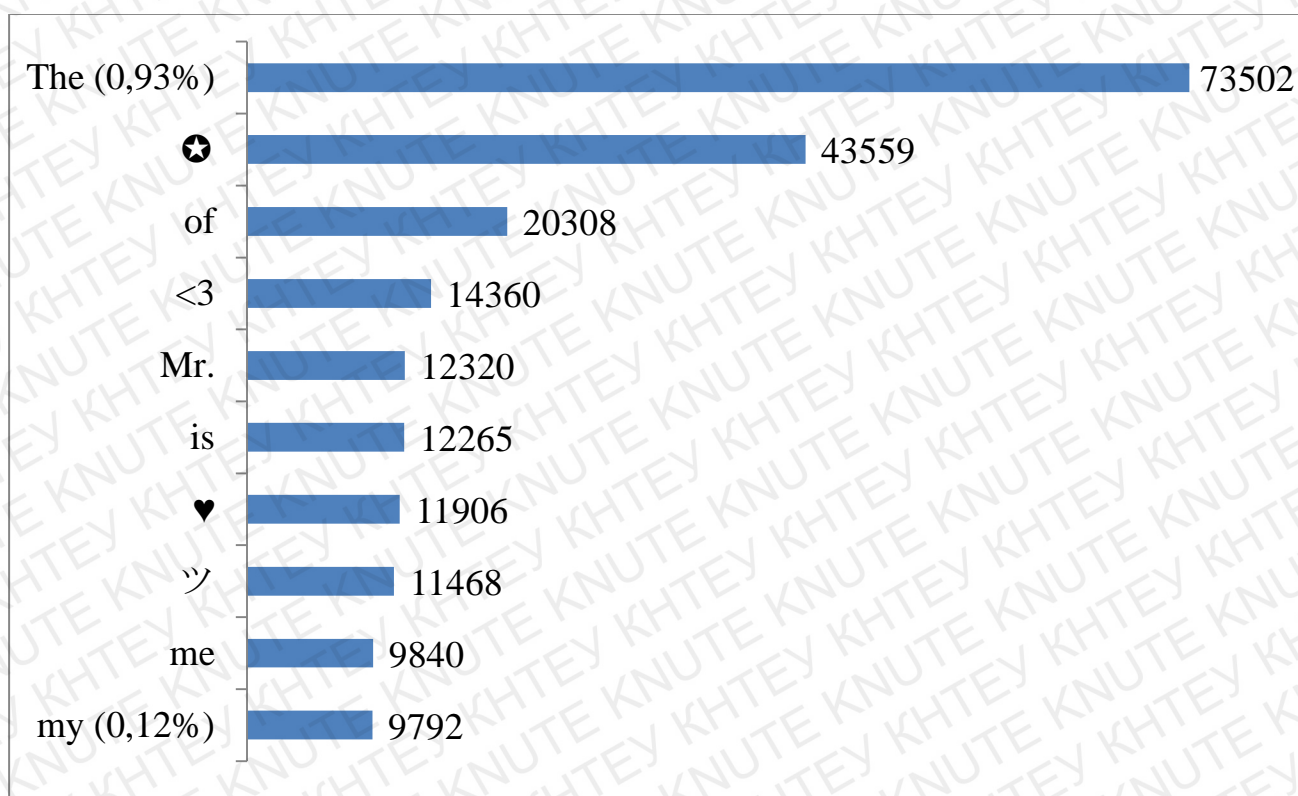


Рис. 3.12 Найпопулярніші слова в імені [Розробка автора]

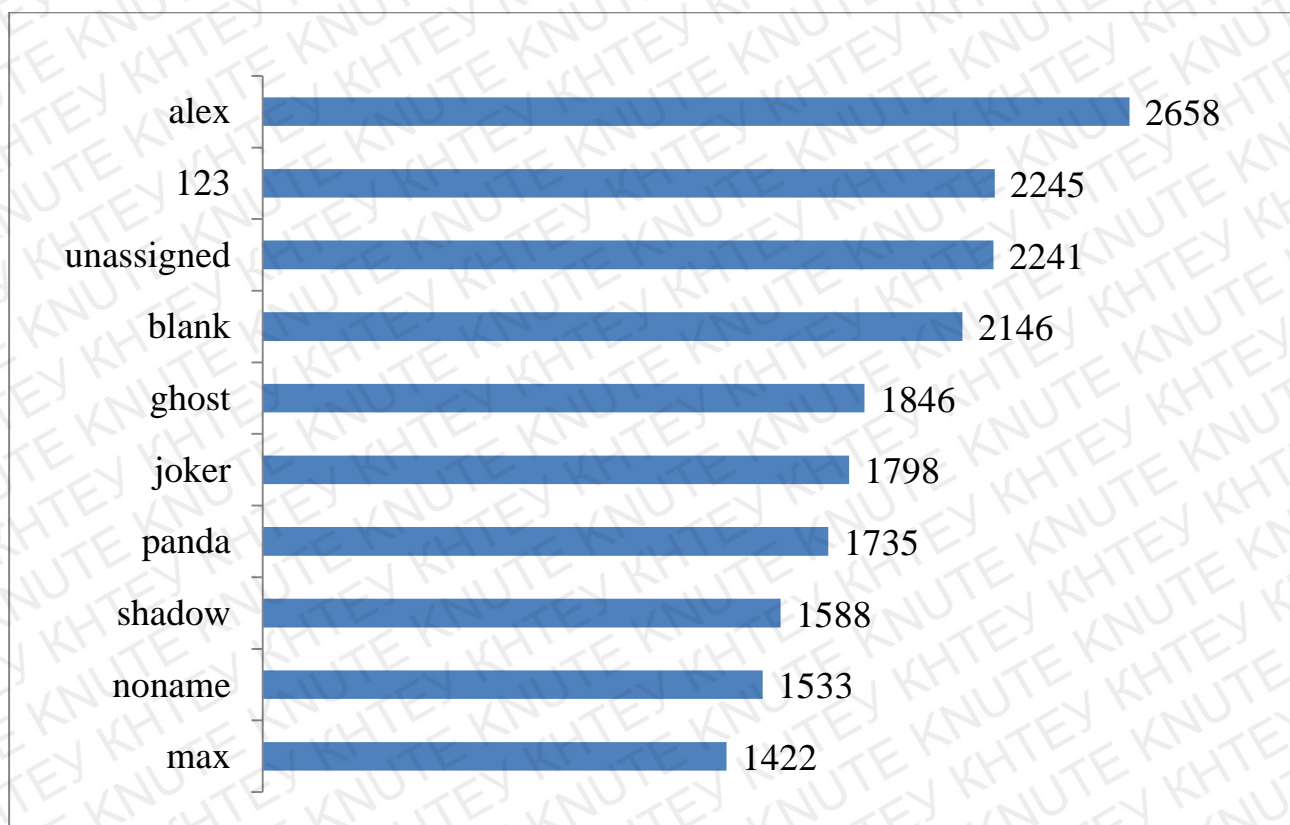


Рис. 3.13 Псевдоніми без символів [Розробка автора]

Наступні результати збору інформації, такі як рівень профілю користувача, кількість придбаних ігор, число спільнот в яких перебуває користувач, кількість друзів та частота зміни псевдоніму вираховувалася окремим алгоритмом винесеним в один клас (рисунок 3.7 – 3.9). Цей алгоритм отримував параметри від інших класів та підраховував середнє значення, медіану, абсолютне відхилення та відносне стандартне відхилення (коефіцієнт варіації). Результати роботи алгоритму зображені на рисунках 3.14 – 3.18.



Рис.к 3.14 Рівень профілю Steam [Розробка автора]



Рис. 3.15 Кількість ігор користувача [Розробка автора]



Рис. 3.16 Кількість спільнот до яких належить користувач [Розробка автора]



Рис. 3.17 Кількість друзів користувача [Розробка автора]



Рис. 3.18 Кількість змін псевдоніму протягом дня [Розробка автора]

Висновки до розділу 3

На основі проведених теоретичних досліджень у перших двох розділах було обрано ринок збуту для аналізу великих масивів даних та розроблено інструмент для пошуку та аналізу інформації в корисних цілях для підприємств пов'язаних з ігровою індустрією.

Ринком збуту було обрано платформу створену для продажу ігор — Steam, яка займає найбільшу частку ринку ігрової індустрії на домашніх комп'ютерах, зважаючи на те, що існують ресурси які надають багато даних пов'язаних з цією платформою, не знайшлося інформації про самих користувачів Steam, тому було вирішено створити Web Crawler, для сканування особистих сторінок користувачів та аналізу тих даних які надають особисто кожен учасник обраного ресурсу.

Було зібрано інформацію з 7 941 457 особистих сторінок, висвітленні дані можуть нести користь для компаній, які займаються продажем ігор. За рахунок отриманих даних маркетингові підрозділи зможуть краще працювати з власною аудиторією споживачів і за рахунок цього збільшувати свої продажі на ринку комп'ютерних ігор.

ВИСНОВКИ

У сучасному світі інтелектуальний аналіз даних набув широкого визнання як потужний і універсальний інструмент аналізу даних не лише в інформаційних технологіях. Обчислювальний процес аналізу великих обсягів даних має на меті вилучення корисної інформації.

Щодня мережу Інтернет заповнюють великими масивами даних і у зв'язку з цим виникла проблема пошуку потрібної інформації. Коли проблема почала стосуватись компаній яким необхідно отримувати цінну інформацію серед великих необроблених масивів даних, почали створювати та вдосконалювати різні моделі видобування інформації. Даною роботою нам вдалося детально дізнатися про створенні методі та технології Data Mining. Виходячи з мети вдосконалення роботи на ринках збуту товарів і послуг було детально оглянуто типи ринків збуту, їх особливості та проблеми з пошуком інформації, яка може нести в собі економічну привабливість для використання на підприємствах які займаються як продажем фізичних товарів чи продуктів виробництва, так і наданням послуг у різних сферах діяльності. Дослідивши проблему аналізу даних за допомогою технологій Data Mining на ринку збуту товарів та послуг нам вдалося дійти до таких висновків. Ринок збуту — в широкому сенсі, будь-який економічний простір для пропозиції товарів і послуг, що є кінцевою метою господарської діяльності підприємства. До найпопулярніших методів аналізу ринку товарів та послуг відносять: статистичний аналіз, комплексний підхід, інтелектуальний аналіз, кластерний аналіз, дерево рішень, нейронні мережі, пошук асоціативних правил та програмно-цільове планування. Кожен за методів має свої особливості та використовується для вирішення специфічних задач. На основі проведених досліджень було проведено програмну реалізацію пошукової роботи (web-crawler)

Досліджуючи проблему пошуку цінної інформації вдалося дійти до висновку, що будь-яка сфера діяльності людини вже потребує відсіювання зайвої інформації, починаючи з буденних соціальних мережі закінчуючи професійною діяльністю компаній-гігантів.

В результаті написання випускної кваліфікаційної роботи було створено та запропоновано до аналізу інформацію про користувачів платформи Steam від компанії Valve. Обираючи ігрову індустрію як ринок збуту було вирішено створити інструмент для пошуку інформації про користувачів (потенційних покупців ігор) за для досягнення мети створено пошуковий робот на мові програмування Python. Основне завдання полягало у записі інформації з web-сторінок про кількість ігор у користувача, його рівень профілю, кількість друзів та спільнот до яких долучений власник сторінки, інформація про VAC-блокування, наявність особливого фону сторінки. Зібрана інформація про псевдоніми користувачів та частота їх змінення. Відображено найпопулярніші слова, символи у псевдонімах, також отримано інформацію про відсоток прихованих та відкритих сторінок.

Дана інформація є цінною для компаній що займаються збутом ігор в ігровій індустрії. Так, маркетинговий відділ може розпоряджатися та аналізувати отриману інформацію з цілю збільшення продажів та отримання прибутку за рахунок кращої комунікації з кінцевим споживачем.

Вивчивши типи ринків збуту, проаналізувавши моделі та технології Data Mining у сучасному світі, та створивши інструмент для отримання цінної інформації ми досягли мети випускної кваліфікаційної роботи.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Барсегян А. А. Анализ данных и процессов / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров // 3-е изд. перераб. и доп. СПб.: БХВ-Петербург — 2009. — 512 с.
2. Большие данные и аналитика [Электронный ресурс]. — Режим доступа: <http://www-03.ibm.com/systems/ru/technicalcomputing/bigdata.html>
3. Будник М. М. Фінансовий ринок / М. М. Будник // Центр навчальної літератури. — 2009. — С. 201.
4. Вандер Д. П. Python для сложных задач. Наука о данных и машинное обучение / Д. П. Вандер // O'Reilly. — 2017. — 576 с.
5. Ваняшин А. За большими данными следит ПАНДА / А. Ваняшин, А. Климентов, В. Кореньков // Суперкомпьютеры. — 2013. — С. 56–61
6. Веклич О. О. Сектору екологічних товарів і послуг в Україні – спеціальний державний патронат / О. О. Веклич // Економіка України. – 2010. – С. 27–35.
7. Кондратенко Н. М. Маркетинг: Підручник та практикум / Н. М. Кондратенко // Юрайт 3-е видання. — 2017. — 542 с.
8. Лапішко З. Я. Формування інфраструктури регіонального ринку цінних паперів: автореф. дис. на здобуття наук. ступеня канд. екон. наук : спец. / З. Я. Лапішко // Львів — 2008. — С. 20.
9. Литвин В. В. Бази знань інтелектуальних систем підтримки прийняття рішень: монографія / В. В. Литвин // Національний університет “Львівська політехніка”. — 2011. — 240 с.
10. Носко В. П. Екорометрія / В. П. Носко // Підручник, книга друга — Видавництво Діло РАНХиГС — 2011. — 672 с.
11. Паклин Н. Б. Бизнес-аналитика: от данных к знаниям / Н. Б. Паклин, В. И. Орешков // – 2-е изд., испр. — СПб. : Питер — 2013. — 702 с.
12. Плєскач В. Л. Інформаційні системи і технології на підприємствах : підручник / В. Л. Плєскач, Т. Г. Затонацька // – Знання. — 2011. — 718 с.
13. Райх Р. Врятувати капіталізм. Як змусити вільний ринок працювати на людей / Р. Райх // Наш Формат. — 2018. — С. 50.

14. Серов Д. Аналитика “больших данных” – новые перспективы / Денис Серов // “StorageNews”, №1 (49) — 2012. — С. 12.
15. Ситник В. Ф. Интеллектуальный анализ данных (дейтамайнінг) / В. Ф. Ситник, М. Т. Краснюк // КНЕУ — 2007. — 376 с.
16. Чернявська О.В. Ринок праці / О.В. Чернявська // Центр навчальної літератури. — 2013. — С. 461.
17. Шаховська Н. Б. Організація великих даних у розподіленому середовищі / Н. Б. Шаховська, Ю. Я. Болюбаш, О. М. Верес // Обчислювальна техніка та автоматизація — 2014. — С. 147–155.
18. Шувар Б. Формування цільового ринку органічної продукції в Україні / Б. І. Шувар // Зб. наук.-тех. пр.: Науковий вісник НЛТУ України. — 2009. — С. 82–85.
19. Яу Н. Искусство визуализации в бизнесе. Как представить сложную информацию простыми образами / Н. Яу // 2013. — 352 с.
20. Allen B. D. Think Complexity / Allen B. D. // Green Tea Press. — 2017, — С. 3.
21. Big data от А до Я. Часть 3: Приемы и стратегии разработки MapReduce приложений [Электронный ресурс]. — Режим доступа: <https://habrahabr.ru/company/dca/blog/270453/>
22. Einav L. The Data Revolution and Economic Analysis / Liran Einav, Jonathan Levin // NBER Working PaperNo. 19035 — 2013. — 24 с.
23. Fisher L. The Perfect Swarm: The Science of Complexity in Everyday Life / L. Fisher // Basic Books. — 2009. — 376 с.
24. History and evolution of big data analytics [Электронный ресурс]. — Режим доступа: https://www.sas.com/en_us/insights/analytics/big-data-analytics.html
25. Ian H. Witten. Data Mining: Practical Machine Learning Tools and Techniques / Ian H. Witten, Eibe Frank, Mark A. Hall // 3rd Edition. Morgan Kaufmann. — 2011. — 664 с.
26. Inmon W. H. Big Data – getting it right: A checklist to evaluate your environment / W. H. Inmon. // DSSResources.COM — 2014. — С. 20.

27. Krum R. Cool infographics: effective communication with datavisualization and design / R. Krum // Indianapolis: Wiley — 2014. — 348 c.
28. Manyika J. Big data: The next frontier for innovation, competition, and productivity / Manyika James . McKinsey // Global Institute — 2011. — 156 c.
29. Marr B. Big Data: Using SMART Big Data, Analytics and Metrics to Make Better Decisions and Improve Performance / Bernard Marr. — John Wiley & Sons Ltd, 2015. — 256 c.
30. Mitchell R. 8 big trends in big data analytics / Robert L. Mitchell // Computerworld — 2014. — C. 567.