

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ
І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ
Факультет інформаційних технологій

УДК 004.9:005
«ПОГОДЖЕНО» «ДОПУСКАЄТЬСЯ ДО ЗАХИСТУ»
Декан факультету інформаційних технологій
Завідувач кафедри комп'ютерних наук

Глазунова О.Г., д.п.н., професор
Голуб Б.Л., к.т.н., доцент

2021 р. 2021 р.

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

на тему «Експертна система на основі платформи SYMS Marketplace»
Спеціальність 122 «Комп'ютерні науки»

(код і назва)

Освітня програма Комп'ютерний еколого-економічний моніторинг

(назва)

Орієнтація освітньої програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)

Гарант освітньої програми

кандидат технічних наук

(науковий ступінь та вчене звання)

Басараб Р.М.

(підпис)

(ПІБ)

Керівник магістерської кваліфікаційної роботи

кандидат технічних наук, доцент

(науковий ступінь та вчене звання)

Голуб Б.Л.

(підпис)

(ПІБ)

Виконав

(підпис)

Шевченко Д.В.

(ПІБ студента)

КИЇВ-2021

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ
І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ
Факультет інформаційних технологій

ЗАТВЕРДЖУЮ

Завідувач кафедри _____

К.Т.Н., доцент Голуб Б.Д.
(науковий ступінь, вчене звання) (підпис) (ПІБ)
" 20 " _____ року

ЗАВДАННЯ

ДО ВИКОНАННЯ МАГІСТЕРСЬКОЇ КВАЛІФІКАЦІЙНОЇ РОБОТИ СТУДЕНТУ
Шевченка Дмитра Віталійовича
(прізвище, ім'я, по батькові)

Спеціальність 122 «Комп'ютерні науки»

Освітня програма Комп'ютерний еколого-економічний моніторинг

Орієнтація освітньої програми освітньо-професійна

Тема магістерської кваліфікаційної роботи «Експертна система на основі платформи SYMS Marketplace»

затверджена наказом ректора НУБіП України від " 29 " жовтня 2020 р. №009

Термін подання завершеної роботи на кафедру 2021.11.30

Вихідні дані до магістерської кваліфікаційної роботи:

- 1) технічна документація
- 2) дані про категорії, товари, виробників та ціни, зібрані з платформи SYMS Marketplace
- 3) програмні продукти

Перелік питань, що підлягають дослідженню:

№ з/п	Питання, що підлягає дослідженню	Строк виконання	Примітка
1.	Аналіз предметної області.	21.09.2020-24.10.2020	
2.	Дослідження інструментів OLAP	30.11.2020-31.12.2020	
3.	Проектування системи	01.02.2021-06.03.2021	
4.	Дослідження та розробка алгоритмів аналізу даних, методів Data Mining	11.03.2021-22.05.2021	
5.	Дослідження отриманих результатів	01.09.2021-13.11.2021	
6.	Попередній захист	15.11.2021-10.12.2021	
7.	Захист	13.12.2021-18.12.2021	

Дата видачі завдання " 12 " листопада 2020 р.

Керівник магістерської кваліфікаційної роботи _____ Голуб Б.Д.
(підпис) (прізвище та ініціали)

Завдання прийняв до виконання _____ Шевченко Д.В.
(підпис) (прізвище та ініціали студента)

НУБіП України

ВСТУП	ЗМІСТ	
1 СИСТЕМНИЙ АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ		3
1.1 Аналіз процесів пов'язаних з реалізацією аналізу продажу товарів в SYMS Marketplace		7
1.2 Технічне завдання		15
1.3 Аналіз існуючих рішень		21
2 ТЕОРЕТИЧНЕ ПІДРУНТЯ		27
2.1 Експертна система аналізу даних		27
2.2 OLAP		28
2.3 Data Mining		30
2.4 Використані технології		32
3 РОЗРОБКА СИСТЕМИ		38
3.1 Архітектура системи		38
3.2 Платформа SYMS Marketplace		44
3.3 Сховище даних		50
4 РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ		53
4.1 Механізм вилучення, обробки і передачі даних		53
4.2 Використані методи Data Mining		55
4.3 Розрахунок КРІ		61
ВИСНОВОК		63
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ		64

НУБІП України

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

- BI - Business intelligence.
- KPI - Key Performance Indicators, ключовий показник ефективності(успішності).

НУБІП України

- MS - Microsoft.
- OLAP - online analytical processing, аналітична обробка в реальному часі.

- OLE DB - Object Linking and Embedding, Database.
- SQL - structured query language, мова структурованих запитів.

НУБІП України

- БД - база даних.
- ІС - інформаційна система.
- СД - сховище даних.

НУБІП України

НУБІП України

НУБІП України

НУБІП України

ВСТУП

Актуальність. З останніми роками, мережу інтернет почали використовувати набагато більше, кількість користувачів постійно збільшується. Всесвітня мережа інтернет стрімко розвивається та набирає обертів, а саме її можливості, які широко почали використовувати в різних сферах життя і діяльності людей. Особливо популярною стає електронна комерція.

Згідно зі звітом Shopify, у 2021 році половина всіх продажів в eCommerce відбувається на онлайн-маркетплейсах. Близько \$2 трильйонів доларів покупці витрачають щороку на найпопулярніших глобальних майданчиках.

Причини такого бурхливого розвитку є очевидними — трафік і оборот. Завдяки маркетплейс покупці отримують широкий асортимент товарів з можливістю порівняти і вибрати оптимальний варіант. Магазини мають стабільний додатковий канал реалізації товарів.

Маркетплейси стрімко набирають популярності серед споживачів, пропонуючи величезний асортимент та низькі ціни. Споживачі, у свою чергу, йдуть на маркетплейси не за брендами, а за товарами, які вирішують конкретні проблеми та задовольняють конкретні потреби.

Економія коштів, можливість доставки кур'єрів на будинок, великий асортимент, покупка брендів, не представлених в місті покупця — причини робити покупки онлайн безліч. Але які ж товари набувають найчастіше? Цей рейтинг дозволяє сформуванню уявлення про купівельну спроможність в мережі Інтернет.

Електронна комерція в 2021 році дає бізнесу неймовірні можливості заробити на хвилі попиту, що зростає, з боку споживачів, які цікавляться покупками в Інтернеті.

До кінця 2021 року загальний обсяг продажів інтернет-магазинів у всьому світі, за прогнозами, повинен досягти позначки в 2.5 трильйони доларів США. Зростання, порівняно з 2019 роком, складе +11%.

Бізнес в сфері інтернет-продажів - один з найпопулярніших видів заробітку в світі. При правильному виборі товару підприємець отримує серйозний дохід, здатний у декілька разів окупили всі витрати. Перед тим як створювати свою справу, потрібно обов'язково провести моніторинг ринку, і обчислити, які товари актуальні на цей момент.

Основне завдання будь-якого інтернет-магазину - продавати та приносити прибуток. Для цього власники бізнесу постійно покращують сайт, запускають контекстну рекламу та ретаргетинг, відправляють email-розсилки, проводять акції тощо. Аналітика допомагає зрозуміти, як всі ці зусилля впливають на кінцевий результат (зростання продажів) а які ні.

Також для того щоб користувачі інтернету були зацікавлені в саме вашому інтернет-магазині ви повинні постійно аналізувати його, та покращувати всі моменти які не дають вашому магазину розвитку. Основним показником в цьому є відслідковування кількості переглядів та кількість покупок.

Предмет та об'єкт дослідження. Об'єктом дослідження є експертна система яка створюється на основі платформи SYMS Marketplace.

Предметом дослідження є система моніторингу та аналізу перегляду та продажу товарів.

Мета дослідження. Метою роботи є дослідження використання сучасних методів Data Mining та OLAP технологій при побудові системи моніторингу та аналізу успішності перегляду товарів та їх продажу в розрізі часу/категорій та продавців.

Зміст поставлених завдань. Для реалізації такої експертної системи першочерговою задачею є накопичення даних. Платформа SYMS Marketplace дозволяє здійснювати таке накопичення. Адже вона система містить в собі досить велику структуру даних, і може мати в собі різні, як аналітичні, так і маркетингові показники.

В рамках даного дослідження ми створюємо систему з використанням технологій OLAP і Data Mining, що аналізує дані по товарам, збирає кількісні

та критеріальні дані, оброблює та робить висновки – рекомендації для подальшого прийняття рішення.

Завдання даної роботи:

- Дослідити предметну область.
- Провести системний аналіз OLAP-технологій.
- Побудувати архітектуру та БД об'єкту дослідження.
- На основі бази даних ми розроблюємо сховище даних, але проектування БД і СД не є самоціллю.
- Провести аналіз даних в реальному часі.
- Сформулювати отримані результати та зробити висновки щодо доцільності використання даних технологій

Методи дослідження. Для проведення аналізу даних та ключових показників ефективності використовується технологія OLAP; для формування нових гіпотез застосовуються алгоритми Data Mining.

Наукова новизна. Було досліджено використання інформаційних технологій для аналізу роботи маркетплейсів в сфері E-commerce, було проведено порівняння декількох інструментів аналізу даних.

Апробація результатів дослідження

- Шевченко Д.В. Дослідження технологій Data Mining, а саме задача класифікації за допомогою IR-алгоритма та методом Наївного Байєса в розрізі експертної системи на основі SYMS marketplace//Збірник матеріалів конференції «Теоретичні та прикладні аспекти розробки комп'ютерних систем» – Київ, © 2021. (тези доступні за посиланням:

https://docs.google.com/document/d/1BLi0UPBEKYNbMHbgsRvVTJv_iVIM0R/view)

- Шевченко Д.В. Експертна система на основі SYMS Marketplace//Збірник матеріалів конференції «Прикладні інформаційні системи, сучасні методи та засоби розробки». – Київ, – 2021. (тези доступні за посиланням:

https://docs.google.com/document/d/1kKFZQhwd2AKAafjrkbyqzaBK_iF9uXNjvfew)

Структура роботи. До складу пояснювальної записки входять: вступ, чотири розділи, висновки, та джерела. У вступі описуються основні аспекти, що характеризують предметну область проекту, коротко охарактеризовано сучасний рівень рішення поставленої задачі і взаємозв'язок з іншими роботами. Вказані мета дослідження, об'єкт, предмет та завдання, зазначено апробацію проведеної роботи та структура пояснювальної записки. В першому розділі проводиться аналіз постановки завдання та предметної області, розкривається постановка завдання, наводиться діаграма прецедентів та діаграма розгортання. В другому розділі розглядається інформаційне забезпечення, де описуються експертна система та використані технології в науковій роботі. В третьому розділі описана архітектура системи, яка включає в себе діаграми компонентів та пакетів. Показана топологія системи з описом вузлів системи. Розробка платформи для збору аналітичних даних SYMS Marketplace. Наводиться структура сховища даних. В четвертому розділі описується механізм вилучення, обробки і передачі даних з основної БД до сховища. Опис використаних методів для аналізу даних. Побудова звітностей та розрахунок KPI. У висновках наведено оцінку отриманих результатів роботи.

1 СИСТЕМНИЙ АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Аналіз процесів пов'язаних з реалізацією аналізу продажу товарів в SYMS Marketplace

Бізнес в сфері інтернет-продажів - один з найпопулярніших видів заробітку в світі. При правильному виборі товару підприємець отримує серйозний дохід, здатний у декілька разів окупили всі витрати. Перед тим як

створювати свою справу, потрібно обов'язково провести моніторинг ринку, і обчислити, які товари актуальні на цей момент.

Визначившись з товаром, перед бізнесменом постає вибір: власний сайт з нуля або інтернет-магазин на платформі маркетплейсу. У кожному з

напрямків є свої плюси і мінуси, але тенденція останніх декількох років вказує на те, що багато хто робить вибір на користь маркетплейсів.

Маркетплейс - це інтернет-майданчик, на якому виробник чи продавець викладає оголошення про продаж товарів. Покупець заходить на маркетплейс і вибирає товар серед різних варіантів. Він може дивитися на бренд, ціну, вартість доставки і т.д. Головне, що тут усі товари показані у загальному списку, а не поділені на сторінки різних виробників.

Незрівнянний плюс для покупців у тому, що можна вибрати найкращий варіант у різних постачальників. Не має значення навіть країна, звідки прибуватимуть товари. Головне, щоб варіант повністю відповідав очікуванням покупця.

Для продавця також маркетплейс відіграє важливу роль. Завдяки цьому посереднику власник фірми або продавець не повинен купувати свій магазин або створювати власний сайт, щоб продавати свої товари.

Існують різні класифікації маркетплейсів. Найпоширеніша класифікація з галузі торгівлі:

- Товари – це такі маркетплейси як AliExpress, Amazon і т.д.
- Послуги - це майданчик для фрілансерів, доставка їжі і т.д.
- Інформація – майданчик для надання інформації від різних людей іншим людям, наприклад, YouTube,

Інвестиції – маркетплейси, які допомагають людям купувати інвестиції та продавати їх, наприклад, Kickstarter та Indiegogo



Рис. 1.11 Типи маркетплейсів

Ці чотири види маркетплейс також поділяються на кшталт пропонуванних послуг і товарів.

- Вертикальні маркетплейс – це ті, що продають товари чи послуги в одному сегменті. Наприклад, це Убер чи Уамола. У першому випадку лише послуги таксі, у другому випадку все, що пов'язане з сльгом
- Горизонтальні – товари чи послуги налаштовані одному напрямку, але можуть бути різного типу.
- Глобальні – вони торгують абсолютно різними речами, наприклад, Wildberries або AliExpress

Також різняться види маркетплейсу на кшталт учасників торгівлі.

- C2C, тобто. Customer to Customer переводиться як споживач споживачеві. Це означає, що продаватиме і купуватиме фізична особа. Тут відсутній великий бізнес та середній бізнес. Максимум свої послуги може пропонувати самозайняте або мевелике ПІ. Найчастіше ми користуємося такими майданчиками, як Юла, Авіто та Airbnb. Ці маркетплейс дозволяють продавати і купувати, здавати житлові приміщення без зайвої паперової тяганини, створення ІП або ТОВ для підприємництва. Великого доходу досягти складно, але торгувати своїми речами чи товарами уроздріб можна. Покупцям не завжди

вигідно купувати за такою схемою, бо повернути товар чи замінити практично неможливо. Продавець не несе жодної юридичної відповідальності

- B2C розшифровується як Business to Customer. Переклад - "Бізнес споживачеві". Це класична схема, коли людина купує товари у великого та малого бізнесу на Аліекспрес, Казанекспрес, Беру і т.д. У разі товари будуть абсолютно новими, з усіма гарантіями. Якщо товар покупця не влаштує, він звертається в претензію на маркетплейс, а той вже серйозно спілкується з постачальником і виробником

- B2B, Business to Business. Схема бізнес бізнесу допомагає продавцеві купувати товар у постачальника за меншу ціну, а потім продавати його на маркетплейсі. У разі покупець ще задіяний і продавець, який купує товар з виробництва чи в постачальника сам є покупцем у разі

Проблеми та переваги

Основні проблеми які зустрічають користувачі під час пошуку товарів в Інтернеті це :

- велика затрата часу на пошук в різних магазинах;
- проблема з асортиментом;
- проблема з доставками;
- проблеми з замовленням;
- помилки в структурі сайту.

Переваги платформи такі:

- пізнаваність і довіра клієнтів;
- перегляд рейтингів та відгуків;
- безпека;
- великий асортимент товарів.

Аналітика даних

Бізнес-аналітика — це процес виявлення, інтерпретації та інформування про знайдені закономірності даних, використання будь-яких засобів що

зможуть допомагати компаніям з аналізом даних які були отримані з різних точок, середовищ. Ця аналітика також надає додаткові можливості для досягнення бажаних результатів, таких як оптимізація, зниження витрат та

взаємодія із замовниками. Ті, хто успішно використовує аналітику даних, не покладаються на інтуїцію і віддають перевагу результатам, заснованим на аналізі даних. В ідеальному сценарії ділові керівники створюють об'єктивну та неупереджену методологію і таким чином одержують результати без упереджених думок та минулого досвіду.

Сьогодні кожна компанія прагне отримати якомога більше від аналітики, використовує більше даних для прискороного та глибокого аналізу для все більш широкого кола людей, і все це за менші гроші. Для досягнення цих цілей потрібна надійна платформа, яка підтримує весь процес аналізу з необхідною безпекою, гнучкістю та надійністю. Вона повинна допомогти користувачам

проводити аналіз у режимі самообслуговування без втрати загального контролю. Вона також має бути простою в адмініструванні. Але як можна отримати такі переваги для системи корпоративного рівня без суттєвих витрат та не погіршити саму інфраструктуру? На даний момент бізнес-аналітика є скрізь де тільки можна зібрати дані, оскільки кожна компанія прагне підвищити свої показники, а отже, аналізуватиме дані для прийняття кращих рішень.

За допомогою бізнес-аналітики з персоналізацією, машинним навчанням і глибокими знаннями галузі компанії можуть отримувати актуальні результати аналізу даних по всіх додатках, сховищах і озерах даних. Бізнес-аналітика повинна бути комплексним процесом, який призводить до конкретних дій. Коли результати аналізу отримані, компанія може зайнятися переоцінкою, змінами та перестроюванням своїх процесів. Головне тут – довести до конкретної дії.

Цінність аналітики для бізнесу:

1. Новий підхід до роботи

Природа бізнесу змінюється, і водночас з'являються нові напрями конкуренції. Сьогодні співробітники добре знаються на технічних питаннях і не хочуть, щоб підготовка до роботи займала багато часу. Забезпечуйте необхідну швидкість та зручність для своїх користувачів, одночасно підтримуючи високі стандарти якості та безпеки даних. Централізована аналітична платформа, де ІТ відіграє ключову роль, має бути фундаментальною частиною Вашої стратегії бізнес-аналітики. Запорукою успішного впровадження інновацій є поєднання проектів, що виконуються бізнес-підрозділами та ІТ.

2. Відкрийте нові можливості

Аналіз даних сприяє розвитку нових технологій і дає можливості для отримання нової додаткової вартості даних. Ці сучасні можливості дозволяють виявляти приховані моделі даних завдяки можливостям прогнозування, самонавчання та адаптації. Технології є інтуїтивними і включають приголомшливі візуалізації даних, які допомагають миттєво аналізувати мільйони рядків і стовпців. Сучасні аналітичні рішення зручні та можуть використовуватися на мобільних пристроях. Можна отримувати доступ до потрібних даних у потрібний час із мінімальним освоєнням принципів роботи і навіть без них.

3. Візуалізуйте свої дані

Звичайно, Ви хочете бачити, про що сигналізують дані раніше конкурентів. Аналітика даних забезпечує чітку та ясну картину ситуації у бізнесі. Аналізуючи особисті, корпоративні та великі дані, ви можете швидко зрозуміти їх зміст, поділитися ним із колегами і зробити все це за лічені хвилини.

Для кращого розуміння вимог до системи було побудовано діаграму варіантів прецедентів. Діаграма варіантів прецедентів – це тип поведінкової діаграми UML, часто використовується для аналізу різних систем. Вони

НУБІП України

Дозволяють візуалізувати різні типи ролей у системі та те, як ці ролі взаємодіють із системою.

Важливість використання діаграми прецедентів.

Діаграми прецедентів використовуються для збору вимог до використання системи. Залежно від ваших вимог, ви можете використовувати ці дані різними способами. Нижче наведено декілька способів їх використання.

- Ідентифікація функцій та як з ними взаємодіють ролі – основне призначення діаграм сценаріїв використання.

- Для представлення системи на високому рівні – Особливо корисно при поданні її керівникам або зацікавленим сторонам. Ви можете виділити ролі, які взаємодіють із системою, та функціональні можливості, що надаються системою, не заглиблюючись у внутрішню роботу системи.

- Ідентифікація внутрішніх та зовнішніх факторів – Це може здатися простим, але у великих складних проектах система може бути ідентифікована як зовнішня роль в іншому випадку використання.

Використовувати діаграми корпусу складаються з чотирьох об'єктів:

- актор;
- випадок використання
- система
- пакет.

Актор використовує діаграму прецедентів - це будь-яка сутність, яка виконує роль в одній даній системі. Це може бути людина, організація або зовнішня система і, як правило, малюється як скелет, показаний нижче.

НУБІП України

Випадок використання являє собою функцію чи дію всередині системи. Вона намальована як овал та названа функцією.

Система використовується для визначення сфери застосування та намальована у вигляді прямокутника. Це необов'язковий елемент, але

корисний під час візуалізації великих систем. Наприклад, ви можете створити всі випадки використання, а потім використовувати системний об'єкт для визначення сфери застосування вашого проекту. Або навіть можна використовувати його, щоб показати різні області, охоплені в різних релізах.

Пакет є ще одним додатковим елементом, який є надзвичайно корисним у складних діаграмах. Подібно до діаграм класів, пакети використовуються для угруповання випадків використання. Вони намальовані, як показано нижче.

Діаграма прецедентів є графом, що складається з множини акторів, прецедентів (варіантів використання) обмежених границею системи (прямокутник), асоціацій між акторами та прецедентами, відношень серед прецедентів, та відношень узагальнення між акторами. Діаграми прецедентів відображають елементи моделі варіантів використання.

Діаграма прецедентів для системи SYMS Marketplace наведена на рис.1.1.2. Проектована система представляється у вигляді безлічі сутностей чи акторів, що взаємодіють із системою за допомогою так званих варіантів використання. Іншими словами, кожен варіант використання визначає деякий набір дій, який виконує система при діалозі з актором.

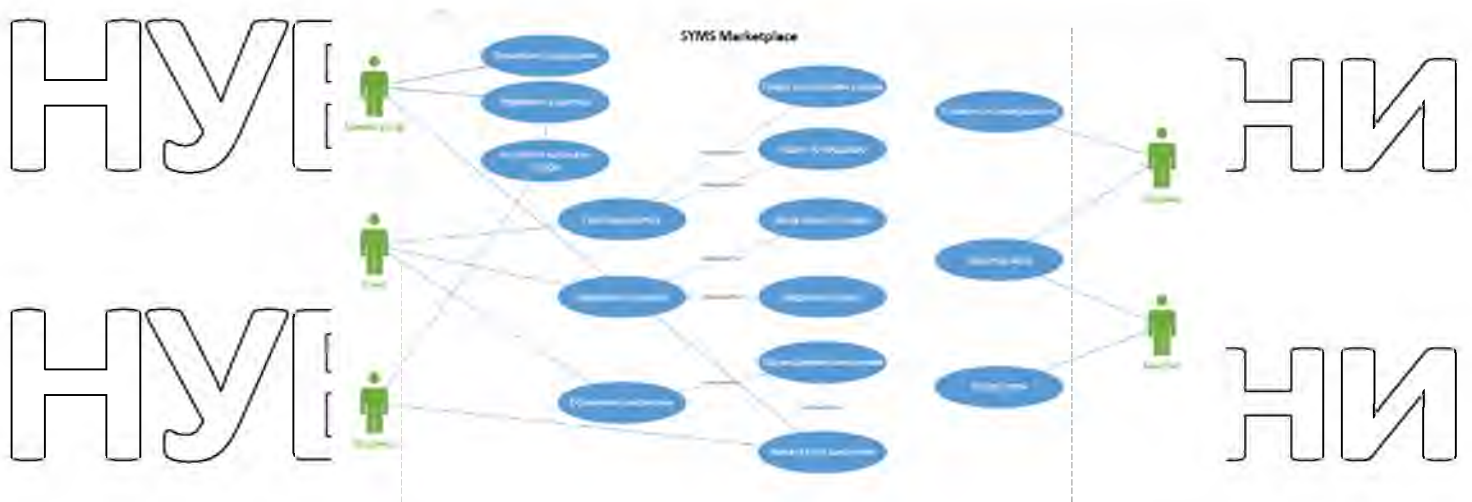


Рис. 1.1.2 Діаграма прецедентів

З діаграми видно, що з системою працюють п'ять основних акторів: клієнт, продавець, адміністратор, керівник та аналітик. Задача аналітика опрацювати ці дані, створити аналітику та донести свої результати до керівника. Адміністратор слідкує за взаємодією клієнтів та продавців в системі. Клієнт має можливість робити замовлення а продавець заповнювати сайт своїми пропозиціями для продажу товарів.

Кожен з них має свої варіанти використання, зробимо їх детальний опис в таблиці 1.1.1.

Таблиця 1.1.1 Опис прецедентів

Основний актор	Назва прецеденту	Опис прецеденту
Клієнт	Перегляд каталогу	Клієнт має можливість переглядати всі доступні в каталогах товари.
Клієнт	Управління корзиною	Є можливість повністю управляти своєю корзиною.
Клієнт	Оформлення замовлення	Оформлення замовлення по відібраним товарам.
Адміністратор	Управління продавцями	Адміністратор може керувати продавцями.
Адміністратор	Управління клієнтами	Адміністратор може керувати клієнтами.
Адміністратор	Управління статусом замовлення	Має можливість змінювати статус замовлення.
Продавець	Керування каталогом товарів	Керування каталогом товарів, а саме наповненням

Продавець	Зміна статусу замовлення	та видаленням своєї продукції. Надана можливість відмінювати або підтверджувати замовлення.
Аналітик	Формування звітів	Обробка даних отриманих зі сховища для надання інформації керівнику
Аналітик	Перегляд та аналіз результатів	Проведення аналізу отриманої інформації для подальшого надання її керівнику
Керівник	Перегляд звітів	Дослідження результату аналізу аналітика
Керівник	Ухвалення рішень про подальший розвиток системи	На основі отриманих результатів прийняття подальших рішень.

1.2 Технічне завдання

Функцією кожної програмної системи є визначене оброблення даних і вивід отриманих результатів. З метою побудови правильної програмної системи до неї, потрібно формулювати вимоги до умов виконання функції та обробки даних. Ці вимоги є предметом практичного контракту між замовником і розробником системи.

Ці вимоги до можна поділити на дві групи, а саме:

- функціональні;
- нефункціональні.

Функціональні вимоги - це перелік тих функцій або сервісів, які система надає користувачам. Функціональні вимоги можуть включати обчислення, технічні деталі, маніпуляції з даними та їх обробку, а також інші конкретні функції, які визначають, що система повинна виконувати. З цього можна виділити такі основні функції:

- реєстрація користувачів, та зберігання основних даних;

• редагування основних даних користувача з можливістю зміни пароля для входу;

• автентифікація користувачів;

• додавання, редагування та видалення каталогу товарів;

• можливість переглядати товари за категоріями;

• пошук товарів в системі;

• кошик товарів з можливістю оформлення замовлення;

• перегляд минулих замовлень;

• сторінка перегляду товарів з повним описом;

• модуль для збору аналітичних даних за товаром в заданий проміжок часу (кількість переглядів/кількість покупок)

Нефункціональні вимоги – визначають умови виконання функцій в середовищі які не напряму пов'язаними з функціями. Вони відображають потреби користувачів по мірі їх виконання. Цім вимогам також притаманні принципи взаємодії з середовищем і можливість визначати показники часу, захисту даних з урахування стандартів.

Складені такі нефункціональні вимоги:

• інтерфейс повинен бути зручним та простим у користуванні;

• для роботи з сервісом потрібне стабільне підключення до мережі інтернет;

• контроль правильності додавання інформації;

• безпека та захист даних;

• відмова стійкість;

• одночасність доступу користувачів до системи;

• кінцева розробка повинна бути безкоштовною.

Діаграма розгортання – це тип UML-діаграми, яка показує архітектуру виконання системи, включаючи такі вузли, як апаратні або програмні середовища виконання, а також проміжне програмне забезпечення, яке їх з'єднує.

Діаграми розгортання зазвичай використовуються для візуалізації фізичного апаратного та програмного забезпечення системи. Використовуючи його, можна зрозуміти, як система буде фізично розгорнута на апаратному забезпеченні.

Діаграми розгортання допомагають моделювати апаратну топологію системи порівняно з іншими типами UML-діаграм, які переважно описують логічні компоненти системи. Діаграма розгортання, зображена на рис. 1.2.1.

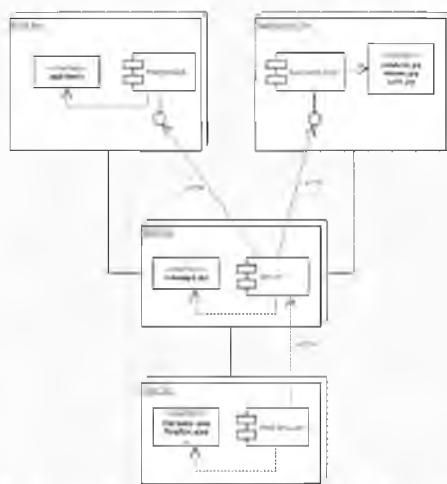


Рис. 1.2.1 Діаграма розгортання.

- Web Tier, Data Tier та Application Tier, які означають серверну частину програмної системи.
- User Tier який означає клієнтську частину.

Системні вимоги (system requirements) - це опис характеристик, яким повинен відповідати комп'ютер.

Для того, щоб на ньому могло використовуватися визначене програмне забезпечення. Вони визначають обмеження на створення продукту. Це визначення зовнішніх умов виконання функцій в системі разом з вимогами до апаратних підсистем. Ці вимоги надають обмеження при розробці системної архітектури, за допомогою візуального відображення та їх функціонування.

Мінімальні апаратні вимоги:

- частота процесора: не менше 1.8 Гц;
- оперативна пам'ять: не менше 1024 МБ.

Підтримка таких операційних систем:

- Microsoft;
- MacOS.

Підтримуючі браузери для операційних системних:

- Google Chrome 20.0 і новіші;
- Mozilla Firefox 35.0 і новіші;
- Microsoft Internet Explorer 6.0;
- Opera 15.0 більш новіші.

Постановка завдання. На етапі постановки завдання ми маємо визначити яких цілей хочемо досягти в результаті роботи над дослідженням та яким способом.

Для створення платформи SYMS Marketplace потрібно виконати такі завдання:

- побудувати відповідно до проекту структуру СУБД;
- розробити інтерфейс програмного забезпечення;
- реалізувати програмним кодом.

Платформа повинна мати реалізований наступний функціонал:

- Сторінка реєстрації користувача в системі яка містить в собі форму для введення з такими полями:

– ім'я;

– прізвище;

– електронна пошта;

– пароль;

– номер телефону;

– отп-пароль.

НУБІП УКРАЇНИ

- Сторінка авторизації користувача на якій розміщена форма з такими полями:
 - пароль;

НУБІП УКРАЇНИ

- номер телефону;
- отп-пароль.
- Головна сторінка платформи:

НУБІП УКРАЇНИ

- відображається слайдер з зображеннями спеціальних пропозицій;
- відображаються новинки товарів.
- Сторінка категорії:

НУБІП УКРАЇНИ

- відображаються товари за обраною категорією;
- реалізований пагінатор для перегляду товарів;
- доданий фільтр для кращого пошуку та фільтруванню.

НУБІП УКРАЇНИ

- Сторінка перегляду товару:
 - відображається вся інформація за товаром;
 - відображається опис магазину та його сервісних послуг;
 - слайдер для відображення фото товару.

НУБІП УКРАЇНИ

- Кошик товарів:
 - можливість з будь якого місця на платформі додавати товар до кошика;
 - редагувати та видаляти товар з кошика;

НУБІП УКРАЇНИ

- оформлення заявки.
- Історія замовлень:

НУБІП УКРАЇНИ

– можливість перегляду раніше створених замовлень.

- Налаштування профілю:

– можливість редагувати дані користувача;

– завантаження фото до профілю.

НУБІП УКРАЇНИ

Ми маємо концентрувати необхідні дані в базі даних (сукупність даних, організованих відповідно до концепції, яка описує характеристику цих даних

і взаємозв'язки між їх елементами). На основі бази даних ми розроблюємо сховище даних, але проектування БД і СД не є самоціллю. Оперативна БД за

НУБІП УКРАЇНИ

допомогою OLTP-технології використовується в автоматизованих системах обробки інформації (ІС). СД використовується у системах аналізу даних.

Технологія, яка у цьому випадку застосовується, має назву OLAP – On-Line Analytical Processing, тобто технологія аналізу даних у режимі реального часу

НУБІП УКРАЇНИ

(оперативний аналіз даних).

Сховище даних має базуватися на основі бази даних, що зберігає інформацію стосовно товарів, продавців, категорій та часових проміжків.

НУБІП УКРАЇНИ

Сховище даних має містити інформацію про часовий вимір, інформацію про товар, продавця їх кількість перегляду та купленого товару. Сам товар має містити інформацію до якої з категорій він відноситься.

Сховище даних має мати можливість надавати всю потрібну інформацію для аналізу даних, зокрема, для відповіді на питання:

НУБІП УКРАЇНИ

- Середня кількість переглядів за товаром;
- Середня кількість куплених товарів по продавцям;
- Середні значення купівлі/переглядів у вказаному проміжку часу.

НУБІП УКРАЇНИ

1.3 Аналіз існуючих рішень

1. Jim MCGAW. Beginning Django E-Commerc. — Jim McGaw, 2009 .

409 с – книга націлена на розробку веб-сайту електронної комерції за

допомогою Django, який є найпопулярнішою платформою веб-розробки на мові програмування Python.

Висвітлені теми включають в себе такі основні моменти: розробка кошику для покупок; можливість створення замовлення та оплати; найкращі практики пошукової оптимізації; розробка пошуку на сайті.

2. Пат. US11062329 (B1) USA Cross device tracking of e-commerce sales /Hust Andrew Michael, Lems Gregory Spencer – № US 2011/132,981 A, Заявлено 2011-11-16; Опубл. 2018-05-15. — 4 с. – Надається система електронної комерції, яка відстежує транзакції покупки на кількох клієнтських пристроях.

Система електронної комерції зберігає інформацію про стрибки, яка описує, коли клієнт знайомиться з продуктом постачальника через філію, яка рекламує продукти постачальника. Система електронної комерції визначає на основі збереженої інформації про хміль, який афілійований або афілійований партнер компенсуватиме продаж продукту. Це дозволяє системі електронної комерції визначати після продажу, чи потрібно компенсувати додаткові філії за продаж продукту на основі інформації про стрибки.

3. Олександр Дружков. ««Швидкий» та «повільний» пошук: два різних підходи для пошуку на сайті» [Електронний ресурс]. – Режим доступу: <https://habr.com/ru/post/70656/> - стаття в якій автор розповідає про два підходи до реалізації пошуку інформації які він використовує на своєму сайті. А саме такі як «швидкий» та «повільний»

4. Користувач ресурсу – Razoornick. «Як ми автоматизували великий інтернет-магазин і стали зіставляти товари автоматично» [Електронний ресурс]. – Режим доступу: <https://habr.com/ru/post/456604/> - в даному електронному ресурсі розповідається як автоматизували інтернет-магазин.

Описано про співставлення та пошук в категоріях по товарам, атрибутам. Проведені дослідження на швидкість та точність співставлення. Проводиться пошук неправильних цін, та робота з базою даних.

5. Костянтин Єлістратов. «Проектування інтернет-магазинів» [Електронний ресурс]. – Режим доступу: <https://habr.com/ru/post/466047/> – Каталог найголовніше в інтернет-магазині. Від того, як він спроектований,

залежить все. Зручність пошуку товару покупцем. Індексція пошуковими системами. Конверсія. Окупність. Помилки в розробці неприємні. Помилки в проектуванні фатальні. Від якості проектування вашого інтернет-магазину залежить буквально все. Вмістяться чи в каталог всі розділи і товари? Чи зможе пошукова система проіндексувати їх коректно? Чи буде покупцеві зручно вибирати товари і шукати важливу інформацію?

Помилки в проектуванні дуже складно виправити на готовому продукті. Іноді ціна виправлення перевищує вартість розробки сайту. Для якісного проектування важливо розуміти обсяг контенту, який буде на сайті, а також шляхи користувачів по сайту і необхідний їм інструмент

6/ Софія Лобанова, Марія Новикова. «Як поліпшити пошук по інтернет-магазину» [Електронний ресурс]. – Режим доступу: <https://usabilitylab.ru/blog/ecommerce-search/> - Простий і зрозумілий інструмент

пошуку дуже важливий для успіху інтернет-магазину. Якщо користувачі не зможуть знайти товари, то не зможуть їх купити. Через погане пошуку користувач може піти на сайт конкурента. Тут описано, які моменти треба враховувати, щоб пошук на сайті дійсно виконував свою функцію.

7. Пат. US2020279213 (A1) CIAA Systems & Methods for an internal private online marketplace / Donald Douglas-Hamilton – №US 2,020,279,213 A1; Заявлено 02.28.20; Опубл. 09.03.20. – 75 с. – Представлені системи, методи та статті виробництва для внутрішнього приватного інтернет-ринку. Деякі варіанти здійснення, описані в цьому документі, можуть включати метод, який може включати ідентифікацію одного або більше елементів, перелічених в одній або декількох базах даних, і генерацію протоколу передачі між цільовим об'єктом, який бажає володіти одним або більше елементами з цим суб'єктом-джерелом, який має або відповідає за один або декілька об'єктів. Бажані

предмети можуть бути ідентифіковані за ознаками. Також розкриваються різні інші методи, системи та машиночитані носії інформації.

8. Пат. US10332181 (B1) США Ranking search results and recommendations / Mcallister Vevina Mariam – №US 10,332,181 B1; Заявлено

15.04.2015; Опубл. 25.06.2015. – 15 с. – Патент про те коли користувач

вибирає товар в інтернет-магазині та надає пошуковий запит до нього, то результати пошуку або рекомендації, отримані від інтернет-магазину у

відповідь на пошуковий запит, можуть відображатися на основі того, чи

задовольнять результати пошуку або рекомендації вимога, встановлена джерелом обраного товару.

Така вимога може стосуватися мінімального порогу витрат або мінімальної кількості предметів. Будь-які результати пошуку або

рекомендації, які, якщо їх обрано, спричинять виконання вимог або

перевищення, можуть бути переважно показані користувачеві вище або більш

помітним чином, ніж результати пошуку чи рекомендації, які, якщо вони вибрані, не спричинять вимоги дотримуватися або перевищувати.

9. Пат. US2018260738 (A1) США Adaptive search refinement /Gabbai

Jonathan, Tenniswood James, Chakravarti Nandita – №US 10,332,181 A1;

Заявлено 08.05.2018; Опубл. 13.09.2018. – 18 с. – В цьому патенті

запропоновано комп'ютерний метод адаптивного уточнення пошуку. Цей

спосіб може включати отримання вказівки на взаємодію користувача з

інтернет-магазином, а у відповідь на отримання індикації - надання уточнення

пошуку за допомогою візуального керування для побудови пошукового запиту

в магазині.

Надання уточненого пошуку з візуальним керуванням може включати

надання варіантів уточнення пошуку, отримання вказівки на схвалення або

несхвалення одного або декількох варіантів уточнення пошуку та повторення

надання варіантів уточнення пошуку та отримання вказівки. Для кожної

ітерації надання безлічі варіантів уточнення пошуку принаймні деякі варіанти

НУБІП УКРАЇНИ

уточнення пошуку можуть бути різними і визначатися на основі раніше отриманих вказівок як про схвалення, так і про несхвалення.

Спосіб адаптивного уточнення пошуку може додатково включати надання результатів пошуку на основі пошукового запиту.

НУБІП УКРАЇНИ

10. Пат. US9619526 (B1) США Increasing the relevancy of search results across categories / Hoover Thomas Jay, Sanders Adam Callahan, Schroedl Stefan – № US 9,619,526 B1; Заявлено 06.11.2015; Опубл. 11.04.2017. — 18 с. - Метод

та системи, що персоналізують результати пошуку, представлені користувачеві на електронному ринку. Отримано пошуковий запит та отримано набір результатів пошуку, пов'язаних із пошуковим запитом. Окремі записи результатів пошуку пов'язані з типом категорії. Результати пошуку сортуються за кількістю груп категорій відповідно до типу категорії.

НУБІП УКРАЇНИ

Вибирається запис із кожної групи категорій та визначаються бали для вибраних записів. Найбільш оцінений запис призначається найвищій доступній позиції у результуючому наборі результатів пошуку. Бали за записи, які залишаються непризначеними, збільшуються. Для групи категорій, що відповідає призначеному запису, обирається та зараховується інший запис.

НУБІП УКРАЇНИ

Найвищий бал присуджується наступній найвищій позиції. Процес може тривати рекурсивно, доки всі або порогове число записів не будуть присвоєні результуючому списку.

НУБІП УКРАЇНИ

11. Пат. US2020210502 (A1) США User interface for search results / Gan Yu, Liu Xiaowei, Gao Huiji, Long Bo – № US 2020/0210502 A1; Заявлено 26.12.2018; Опубл. 02.07.2020. — 16 с. - Інтернет-система та спосіб включають

отримання пошукового запиту, що включає принаймні один пошуковий термін, при цьому пошуковий запит пов'язаний з членом онлайн-системи. Тег даних застосовується окремо до кожного окремого пошукового терміну пошукового запиту.

НУБІП УКРАЇНИ

Стан неоднозначності пошукового запиту визначається на основі принаймні деяких дій, що зберігаються в електронному сховищі даних, також

сконфігурованому для збергання елементів вмісту онлайн-системи, включаючи дані профілю учасника.

Розподіл ймовірностей категорій елементів вмісту визначається на основі тегів даних та принаймні деяких дій та, якщо пошук неоднозначний, даних профілю учасника. Принаймні один елемент вмісту пов'язаний з категорією елемента вмісту, що має найбільшу ймовірність розподілу ймовірностей, і користувальницький інтерфейс відображає принаймні один елемент вмісту.

12. Пат. RU2011132437 (A) РФ Information search method (versions) and computer system for realising said method / Литвин Семен Соломонович, Рубин Михайл Семенович, Колчанов Станислав Аркадьевич, Соколов Євгеній Львович, Смирнов Александр Валентинович, Іванькович Іван Сергєевич – № RU 2011/132,437 A; Заявлено 01.08.2011; Опубл. 10.02.2013. — 4 с. – Цей винахід відноситься до функціонально-орієнтованого пошуку в базах даних, що містять матеріали, класифіковані за заголовками однієї з класифікаційних систем, наприклад, міжнародної класифікації патентів, і може бути використане для різних інформаційних завдань, зокрема для пошуку інженерних рішень в різних галузях науки та техніки.

У відповідь на визначене користувачем завдання у вигляді опису слова функції, яка включає принаймні слово, що характеризує дію, та/або слово, що описує параметр, зміна якого характеризує результат згаданої дії, для пошуку для інформації про можливе виконання якого здійснюється пошук інформації, система забезпечує опис слів функціональних аналогів, і у відповідь на їх вибір користувачем представляє заголовок системи класифікації, за якою пошук інформації може бути здійснений після цього.

Заголовки визначаються модифікованою системою класифікації, отриманою заздалегідь, додаючи до заголовків системи класифікації поле, що містить принаймні одне слово, яке характеризує дію, та/або слово, що описує параметр, зміна якого характеризує результат цього дія, що відповідає змісту

заголовка системи класифікації, де заголовок визначається на основі відповідності принаймні одного із зазначених слів, що містяться в додатковому полі модифікованої системи класифікації, принаймні одним словом із слова опис вибраних функціональних аналогів.

НУБІП України

НУБІП України

НУБІП України

НУБІП України

2 ТЕОРЕТИЧНЕ ПІДРУНТЯ

2.1 Експертна система аналізу даних

Технологію побудови експертних систем найчастіше називають інженерією знань. Як правило, цей процес вимагає специфічної форми взаємодії творця експертної системи, якого називають інженером знань, та одного або кількох експертів у деякій предметній області. Інженер знань «витагує» з експертів процедури, стратегії, емпіричні правила, які вони використовують при вирішенні завдань, та вбудовує ці знання на експертну систему.

НУБІП України

Експертні системи призначені для аналізу даних, що містяться в базах знань, та видачі рекомендацій на запит користувача. Використовуються у випадках, коли вихідні дані добре формалізуються, але прийняття рішення потрібні спеціальні, великі знання. Експертні системи – це складні програмні комплекси, що акумулюють знання фахівців у конкретних предметних галузях та тиражують цей емпіричний досвід для консультацій менш кваліфікованих користувачів.

Серцевину експертної системи становить база знань (БЗ), яка накопичується у процесі побудови ЕС. Знання повинні бути виражені в явному вигляді та організовані так, щоб спростити прийняття рішень. Важливість цієї особливості експертної системи неможливо переоцінити.

Наслідки цього процесу за межі побудови програми, призначеної на вирішення деякого класу завдань. Причина у тому, що знання, як основа ЕС, є явними та доступними, що й відрізняє ці системи від більшості традиційних програм. БЗ мають таку ж цінність, як і будь-який великий обсяг знань, які можуть широко поширюватися за допомогою книг та лекцій.

Корисною рисою експертних систем є наявність у них прогностичних можливостей. Експертна система може функціонувати як теорію обробки інформації або моделі розв'язання задачі у заданій області, даючи очікувані відповіді у конкретній ситуації та показуючи, як зміняться ці відповіді нових ситуаціях. Експертна система може пояснити докладно, як нова ситуація призвела до змін. Це дозволяє користувачеві оцінити можливий вплив нових фактів або інформації та зрозуміти, як вони пов'язані з рішенням. Аналогічно, користувач може оцінити вплив нових стратегій чи процедур на рішення, додаючи нові правила або змінюючи існуючі.

2.2 OLAP

Термін OLAP слугує для опису моделі подання даних і відповідно технології їхньої обробки в сховищах даних. В OLAP застосовується багатомірне подання агрегатних даних для забезпечення швидкого доступу до

стратегічно важливої інформації з метою поглибленого аналізу. Додатки OLAP повинні мати такі основні властивості:

- багатомірне подання даних;
- підтримка складних розрахунків;
- правильний облік фактору часу.

Додатки OLAP опираються на сховища даних, одержуючи від них актуальні дані, що дозволяє контролювати цілісність корпоративних даних.

Таким чином, OLAP - технологія оперативної аналітичної обробки даних, що використовує методи і засоби для збору, зберігання та аналізу.

Основне призначення OLAP-систем багатовимірних даних з метою підтримки процесів прийняття рішень підтримка аналітичної діяльності, довільних запитів користувачів-аналітиків. Мета OLAP-аналізу - перевірка

гіпотез.

Всі дані в OLAP-системах поділяються на три категорії: детальні, агреговані, метадані (рис. 2.2.1).

1) Детальні дані — дані, що переносяться безпосередньо з OLTP-підсистем. Відповідають елементарним подіям, фіксується у OLTP-системах.

Поділяються на:

- виміри - набори даних, необхідні для опису подій (товар, продавець, покупець, магазин, ...)
- факти - дані, що відображають сутність події (кількість проданого товару, сума продажів, ...)

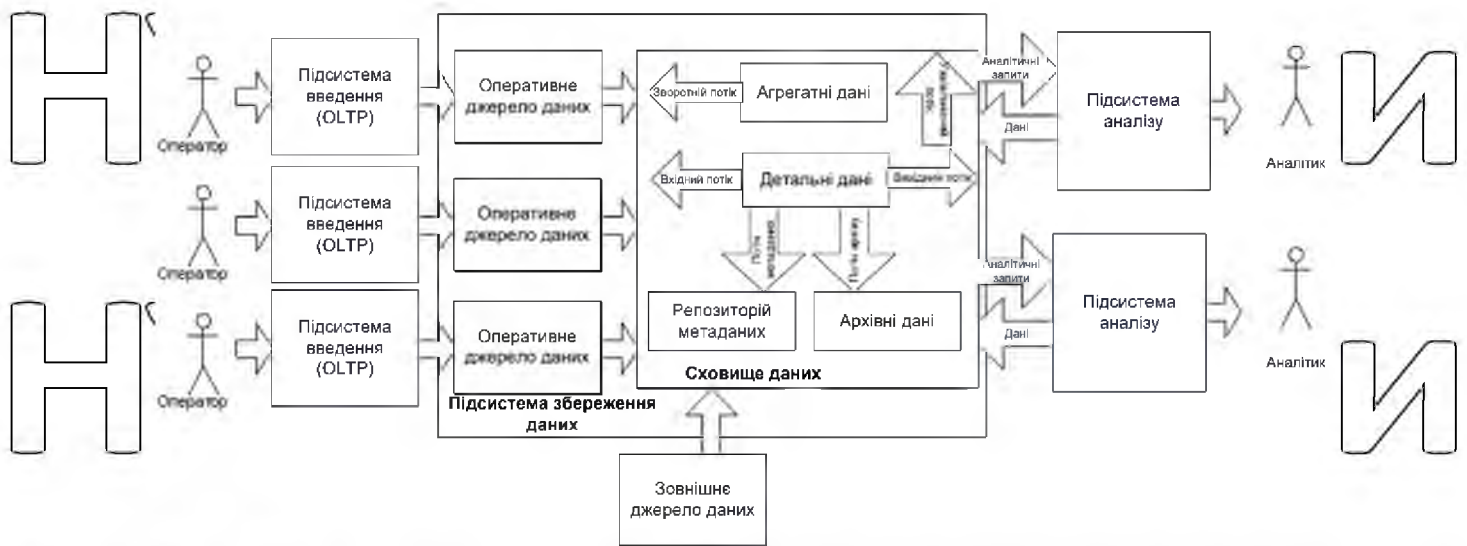


Рис. 2.2.1 Категорії даних та інформаційні потоки в OLAP

2) Агреговані (узагальнені) дані - дані, отримані на підставі детальних шляхом підсумовування за певними вимірами.

3) Метадані — дані про дані, що містяться в СД. Можуть описувати:

- об'єкти предметної області, інформація про яких міститься в СД;
- категорії користувачів, що використовують дані в СД;
- місця і способи зберігання даних;
- дії, що виконуються над даними;
- час виконання різноманітних дій над даними;
- причини виконання різних дій над даними.

Для коректного функціонування OLAP-систем необхідно виконання певних етапів. Ці етапи дозволяють зібрати інформацію у СД, яка і забезпечить аналіз процесів, що відбуваються у корпорації. Етапи: вилучення та перетворення, очищення даних, завантаження, оновлення та управління метаданими.

2.3 Data Mining

Data Mining - це автоматизований пошук даних, заснований на аналізі великих масивів інформації. За мету береться ідентифікація тенденцій та патернів, яка за звичайного аналізу неможлива. Для сегментації даних та

НУБІП УКРАЇНИ

оцінки ймовірності наступних подій використовуються складні математичні алгоритми.

Основні складові data mining:

- Автоматичне виявлення патернів. І тому вибудовуються спеціальні моделі, з урахуванням яких задіяні певні алгоритми аналізу та аналізу даних

НУБІП УКРАЇНИ

- Прогнозування можливих результатів. Відповідно до деяких форм data mining можна встановлювати звані “правила” чи “обмеження”, щоб підвищити рівень точності прогнозу події.

НУБІП УКРАЇНИ

- Генерування релевантної інформації. Також є форми data mining, які можуть допомогти сегментувати групи споживачів з урахуванням необхідних характеристик, що значно підвищує точність націлювання

НУБІП УКРАЇНИ

- Фокусування уваги великих масивах даних. Рекламу

Статистичні методи зазвичай працюють з невеликими обсягами інформації, так званими вибірками, що суттєво спотворює прогнозованість

НУБІП УКРАЇНИ

результатів.

Алгоритми data mining побудовані таким чином, щоб працювати з величезними масивами інформації.

НУБІП УКРАЇНИ

Для обробки зібраної інформації використовується інша структура – online analytical processing. Це швидкий онлайн аналіз даних, зібраних із різних майданчиків. Таким чином, взаємодія цих двох структур допомагає зробити збір та аналіз даних ультра швидким.

НУБІП УКРАЇНИ

Data mining не принесе користі без обдуманого керування. Сам процес дасть результати лише у разі чіткого розуміння специфіки самого бізнесу, даних, методів аналітики. Data mining дозволяє розкрити нові способи

НУБІП УКРАЇНИ

застосування даних компанії, але необхідний безпосередній аналіз визначення цінності знайдених патернів.

Важливо пам'ятати, що спрогнозовані взаємовідносини споживачів із брендом не обов'язково є причиною його дій. Таким чином, data mining

допоможе ідентифікувати групу споживачів, які, ймовірно, куплять продукт, але приналежність до цієї групи не буде причиною їх купівлі.

Усього існує 4 етапи видобутку даних:

- визначення проблеми (завдання);
- збір даних та підготовка до аналізу;
- побудова моделі та оцінка;
- використання отриманих знань на практиці.

Визначення проблеми – це чітке позначення цілей та вимог. Спочатку варто визначити проблему, яку потрібно вирішити. Наприклад, як продати більше товару споживачеві. Далі слід інтерпретувати це питання у зрозумілу для data mining мету – яких споживачів найімовірніше зацікавить цей продукт.

Збір даних та підготовка до аналізу:

- Пошук та збір даних;
- Визначити, наскільки релевантною є зібрана інформація для вирішення бізнес проблеми (визначеної на першому етапі);

- Зняти шум – усунути непотрібну інформацію;
- Ідентифікувати закономірності та патерни;
- Побудова таблиці, де буде структуровано зібрана інформація, необхідна майбутньої моделі аналізу.

Побудова моделі аналізу та оцінка:

• Коригування параметрів та технологій побудови моделі для створення оптимізованої системи аналізу проблеми;

- Співвідношення моделі і проблеми, яку вона покликана вирішити.

Якщо вони відрізняються, повертаємось на етап вище.

Використання отриманих знань на практиці - вивчення специфіки моделі, яка принесла певні результати (або не принесла, теж важливо розуміти, що було зроблено неправильно, щоб уникнути повторення помилок).

Сам процес видобутку даних нескінченний. Це кругообіг взаємодоповнюючих та послідовних кроків, які допомагають ідентифікувати, вирішити та визначити нове завдання.

2.4 Використані технології

Для розробки платформи SYMS Marketplace були використано такі технології як:

- мова програмування Python;
- framework Django;
- Django REST framework;
- PostgreSQL;
- Vue.js;
- Material Design Framework – Vuetify.

Python - мова програмування яка відноситься до високого рівня з динамічною типізацією (можна вказати строгою динамічною) яка в свою чергу є інтерпретованою.

Ця мова була розроблена в 1990 році програмістом Guido van Rossum.

Ця мова програмування виглядає досить привабливо через свої структури даних, її динамічне зв'язування (в процесі виконня коду) та засіб поєднувати компоненти.

Python має підтримку модулів та пакетів, що надає можливість модульності та використовувати код повторно. Інтерпретатор в мові програмування Python та `standard libraries` доступні як в скомпільованому вигляді для швидкого використання так і у вихідному коді який розміщений на платформах, доступних для будь якого розробника.

Також ця мова має підтримку декількох парадигм стосовно програмування, а саме такі як об'єктно-орієнтована, процедурна, функціональна та аспектно-орієнтована.

Framework Django це framework для створення веб додатків за допомогою мови високого рівня програмування Python. Він був створений на початку 2005 року, коли веб-розробники з газети Lawrence Journal-World почали використовувати Python в якості мови для створення веб-платформ. А вже в 2008 році вийшов публічний перший реліз фреймворку. На сьогоднішні він продовжує розвиватися.

Django досить популярний. Він використовується на багатьох сайтах, в тому числі таких, як Pinterest, PBS, Instagram, BitBucket, Washington Times, Mozilla і багатьох інших.

Vue.js це прогресивна структура для створення інтерфейсів користувача. На відміну від інших монолітних фреймворків, Vue розроблений з нуля, щоб його можна було поступово використовувати. Основна бібліотека зосереджена лише на шарі перегляду, і її легко підібрати та інтегрувати з іншими бібліотеками або існуючими проектами. З іншого боку, Vue також ідеально здатний використовувати складні односторінкові програми в поєднанні з сучасними інструментами та допоміжними бібліотеками.

Vuetify — це повна структура інтерфейсу користувача, побудована на основі Vue.js. Мета проекту — надати розробникам інструменти, необхідні для створення багатих і привабливих користувацьких можливостей. На відміну від інших фреймворків, Vuetify розроблено з нуля так, щоб його було легко

навчитися та корисно опанувати завдяки сотням ретельно розроблених компонентів.

Vueify використовує мобільний підхід до дизайну, що означає, що ваша програма працює «з коробки» — на телефоні, планшеті чи настільному комп'ютері.

PostgreSQL - вільна об'єктно-реляційна система управління базами даних. Існує в реалізаціях для безлічі UNIX-подібних платформ, включаючи AIX, різні BSD-системи, HP-UX, IRIX, Linux, macOS, Solaris / OpenSolaris, Tru64, QNX, і також Microsoft Windows.

MySQL є найпопулярнішою у світі базою даних з відкритим вихідним кодом. Незалежно від того, чи є ви веб-власністю, що швидко розвивається, постачальником технологій або великим підприємством, MySQL може рентабельно допомогти вам надати високопродуктивні, масштабовані програми баз даних.

Служба аналізу SQL Server (SSAS) — це багатовимірний OLAP-сервер, а також аналітичний механізм, який дозволяє нарізати та нарізати великі обсяги даних. Він є частиною Microsoft SQL Server та допомагає виконувати аналіз з використанням різних вимірювань. Він має 2 варіанти багатовимірного та табличного. Повна форма SSAS — це служба аналізу SQL Server.

Microsoft Integration Services - це платформа для створення рішень для інтеграції даних і перетворення даних на рівні підприємства. Використовуйте служби інтеграції для вирішення складних бізнес-завдань шляхом копіювання або завантаження файлів, завантаження сховищ даних, очищення та інтелектуального аналізу даних, а також керування об'єктами та даними SQL Server.

Business intelligence (BI) - це термін-метафора, який не має дослівного перекладу й тлумачення і позначає ієрархічно-синергетичний комплекс концепцій, технологій і програмних засобів аналізу первинних даних і візуалізації його результатів для підтримки рішень. Business Intelligence

НУБІП УКРАЇНИ

об'єднує технології реляційних (у справжній час і нереляційних баз даних — NoSQL) разом з найбільш зрілими технологіями Artificial Intelligence та передовими технологіями традиційної статистики й візуалізації результатів аналізу.

НУБІП УКРАЇНИ

Зазвичай програмні продукти типу ВІ використовують дані, які зберігаються у сховищі даних.

Системи типу ВІ обов'язково містять такі наступні три основні категорії функцій:

- можливість інтеграції;
- представлення інформації;
- аналіз даних.

НУБІП УКРАЇНИ

Алгоритм 1-Rule. До кожного можливого значення кожної незалежної змінної формується правило, яке класифікує об'єкти з навчальної вибірки. При цьому в правій частині вказується значення тієї залежної змінної, яка найчастіше зустрічається у об'єктів з обраним значенням незалежної змінної, але не відноситься до обраного класу. Якщо незалежні змінні мають речовий тип, кількість можливих значень може бути нескінченно. Для вирішення цієї проблеми всю область значень розбивають на інтервали, тому щоб кожен з них відповідав певному класу.

НУБІП УКРАЇНИ

Метод Наївного Байєса — метод класифікації, що ґрунтується на застосуванні теореми Байєса з сильним припущенням, що всі предиктори незалежні один від одного. Простіше кажучи, передбачається, що наявність функції у класі не залежить від присутності будь-якої іншої функції у тому ж класі.

НУБІП УКРАЇНИ

KPI (key performance indicators, ключові показники ефективності) - показники діяльності підрозділу (підприємства), які допомагають організації в досягненні стратегічних і тактичних (операційних) цілей. Використання ключових показників ефективності дає підприємству можливість оцінити своє поточний стан і допомогти в оцінці успішності реалізації обраної стратегії.

KPI дозволяють виробляти контроль ділової активності співробітників, підрозділів і компанії в цілому. Для терміну «key performance indicators (KPI)» часто використовується український переклад «ключові показники ефективності» (КПЕ), однак правильним перекладом вважається варіант «ключові показники результатів діяльності».

За допомогою KPI можна створити досконалу та ефективну систему мотивації і стимулювання співробітників компанії, тому KPI і мотивація персоналу стали нерозривними поняттями.

Залежно від стратегії компанії розрізняють різні KPI. В основному їх застосовують для визначення результативності роботи адміністративно-управлінського персоналу.

Ключові показники ефективності можна розділити на:

- запізнілі - відображають результати діяльності після закінчення періоду;
- випереджаючі - дають можливість управляти ситуацією в межах звітного періоду з метою досягнення заданих результатів по його закінченні.

До запізнілих відносяться фінансові показники. Вони, в силу своєї специфіки, не можуть описувати поточну ефективність підрозділів і компанії в цілому.

Оперативні (випереджальні) показники дозволяють оцінити саме поточну діяльність підрозділів і компанії в цілому, паралельно і побічно відповідаючи на питання про те, які грошові потоки можуть бути в майбутньому, а також яка якість процесів і продукції, ступінь задоволеності замовників.

В результаті введення KPI можуть бути отримані наступні позитивні результати:

- Підвищення ефективності роботи підприємства.
- Опис головних факторів і показників, від яких залежить успішність роботи підприємства.

НУБІП України

- Планування від цілей.
- Розробка реалістичних нормативів.
- Прозора оцінка ефективності роботи персоналу.

НУБІП України

3 РОЗРОБКА СИСТЕМИ

3.1 Архітектура системи

НУБІП України

Будь-яка інформаційна система часто складається з трьох функціональних частин – модулі для повноцінного зберігання даних, їх правильної обробки та інтерфейс з яким буде взаємодіяти користувач. Кожна частина може бути реалізована окремо від іншої але в подальшому відповідно працювати в поєднанні.

НУБІП України

Маючи такі вимоги до системи найбільш правильним буде використання клієнт-серверної архітектури яка дозволяє нам розділити ці три функціональні частини в два модулі.

НУБІП України

На сервері буде розташоване програмне забезпечення що відповідає за зберігання даних. Інтерфейс користувача знаходиться на стороні клієнта. Переважна частина обробки даних буде відбуватися на сервері, але на стороні клієнта вона також буде, тому вона розділиться між ними двома.

НУБІП України

Архітектура системи описує її основні компоненти, їх зв'язки (структури) та те, як вони взаємодіють між собою.

Архітектура є основою для системи:

- вона забезпечує абстраговане уявлення про систему для управління її складністю, встановлення механізму зв'язків та координації між компонентами;

НУБІП України

НУБІП УКРАЇНИ

- вона визначає структуроване рішення для задоволення всіх технічних та експлуатаційних вимог, оптимізує загальні атрибути якості, такі як продуктивність та безпека;

НУБІП УКРАЇНИ

- вона передбачає формування переліку важливих рішень щодо структури, які пов'язані з розробкою програмного забезпечення.

Кожне з цих рішень може мати значний вплив на якість, ремонтпридатність, продуктивність та загальний успіх кінцевого продукту.

Ці рішення включають:

НУБІП УКРАЇНИ

- вибір структурних елементів та їх інтерфейсів, з чого саме і формується система;
- вибір поведінки, яка визначається взаємодією між цими елементами;

НУБІП УКРАЇНИ

- об'єднання цих структурних та поведінкових елементів у великі підсистеми;
- узгодження системних архітектурних рішень з цілями бізнесу;

НУБІП УКРАЇНИ

- вибір архітектурних стилів, які визначають структуру складових системи.

НУБІП УКРАЇНИ

Для відображення структури програмного забезпечення найкраще підходить діаграма компонентів (рис. 3.1.1), вона найбільш повно показує усі частини програмної системи та їхні взаємозв'язки. На цій діаграмі ми можемо побачити, що є як клієнтська частина так і серверна.

НУБІП УКРАЇНИ

Клієнтська частина:

- `main_client.html`;

НУБІП УКРАЇНИ

- `main_client.dll`.

Серверна включає в себе :

- `main_server.py`;

НУБІП України

- server.dll;
- database.db;
- help.py.

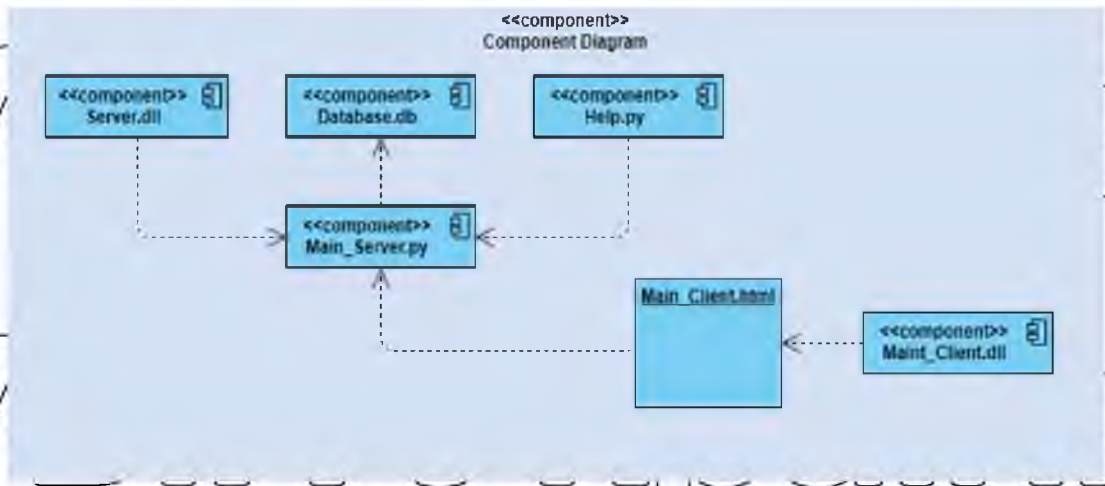


Рис. 3.1.1 Діаграма компонентів.

Програмна система працює за шаблоном MVT (Модель-Уявлення Шаблон), він передбачає поділ системи на три взаємопов'язані частини.

- Model: описує дані, які використовуються в додатку. Окремі класи, як правило, відповідають таблицям в базі даних.

- View: отримує запит, обробляє його і відправляє у відповідь користувачеві деяку відповідь. Якщо для обробки запиту необхідно звернення до моделі і бази даних, то View взаємодіє з ними. Для створення відповіді може застосовувати Template або шаблони. В архітектурі MVC цього компоненту відповідають контролери (але не уявлення).

- Template: представляє логіку уявлення у вигляді згенерованої розмітки html. У MVC цього компоненту відповідає View, тобто уявлення.

Для більшого розуміння роботи системи можна глянути на діаграму пакетів (рис. 3.1.2), де ми можемо побачити як взаємодіють і залежать між собою компоненти системи.

НУ

НУ

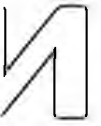
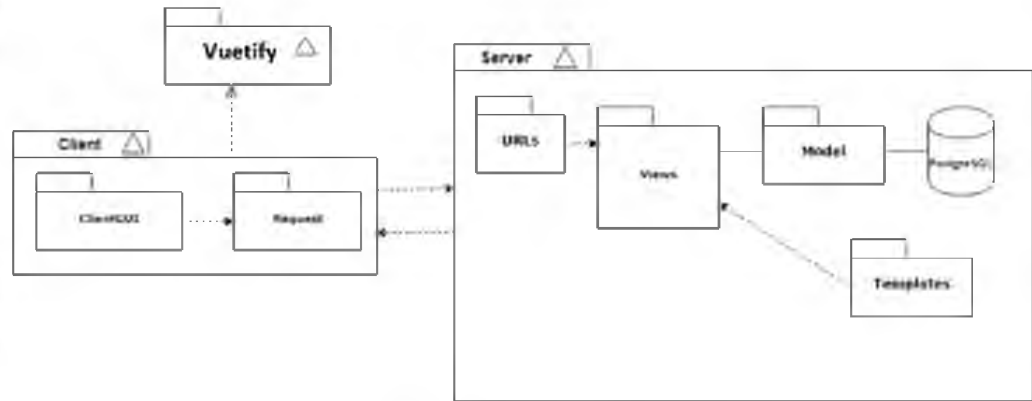


Рис. 3.1.2 Діаграма пакетів

Архітектура системи — це сукупність зв'язків між частинами системи. Існують різні визначення системної архітектури, і різні організації описують її різними способами. Іншими словами, архітектура системи – це сукупність її основних функціональних елементів та засобів забезпечення їх взаємодії один з одним, з користувачем та з системним персоналом.

Архітектуру системи зображено на рис. 3.1.3

НУ

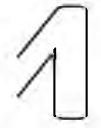
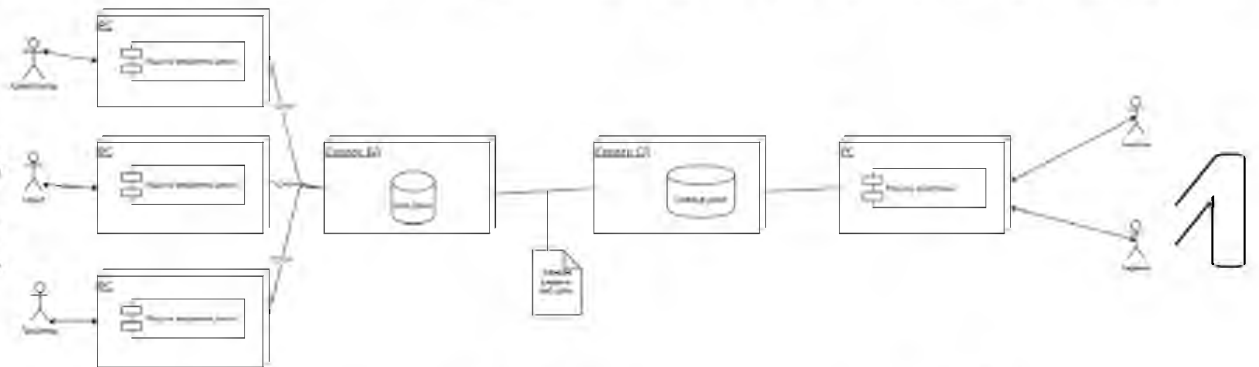


Рис. 3.1.3 Топологія системи

Інформаційно-аналітична підсистема аналізу складається з клієнта, який робить замовлення товару, адміністратор – оновлює, керує та додає інформацію в системі, продавець – керує каталогом своїх товарів. Дані три об'єкта відповідають за наповнення Баз даних, після чого частина даних копіюється в сховище даних оскільки не вся інформація потрібна аналітику для аналізу поставлених задач. В кінці ми маємо підсистему аналізу або так

НУБІП УКРАЇНИ

звану OLAP технологію, за допомогою якої буде відбуватися аналіз даних, що зберігається в сховищі.

Логічна модель – абстрактна структура в інформаційній області, в якій є погляд в загальному на дані, предметна область з певною моделлю даних.

Можна також її називати концептуальною моделлю. Такі дані виступають в ролі моделі з типом сутність зв'язок.

ERwin – це такий програмний пакет, який дозволяє створювати саме логічну модель даних з типом сутність-зв'язок в стандартах IDEF1x. Чим це добре, що при розробці в цьому стандарті можна побудувати як логічну так і фізичну на її основі. Побудувати фізичну БД можна майже для будь-яких СКБД. Головною його перевагою є те, що можна спокійно робити зворотнє перетворення з однієї моделі в іншу. Що є дуже непоганою якістю.

Реляційна модель даних — це та ж сама логічна модель даних. На сьогоднішній момент ці моделі є головним стандартом в світі розробки, на них спираються майже всі новітні СУБД. В таких моделях ми можемо побачити найбільший абстрактний рівень.

Нормалізація – являється поетапним послідовним перетворенням початкової БД в нормальну форму. Якщо використовувати наступну НВ, то обов'язково має бути виконана попередня форма.

1 НФ - доведення відношення стосовно неї нескладна операція. Потрібно поглянути на відношення та розділити атрибути на потрібні відповідні рядки та стовпці. Таку дію потрібно проводити можливо і не один раз доки всі атрибути не стануть атомарними. На даний момент всі значення кожного з атрибутів є атомарними. З цього зробимо висновок що відношення перебуває в 1 НФ.

ER-діаграма, що створена в додатку ER-win, представлена на рис. 3.1.4

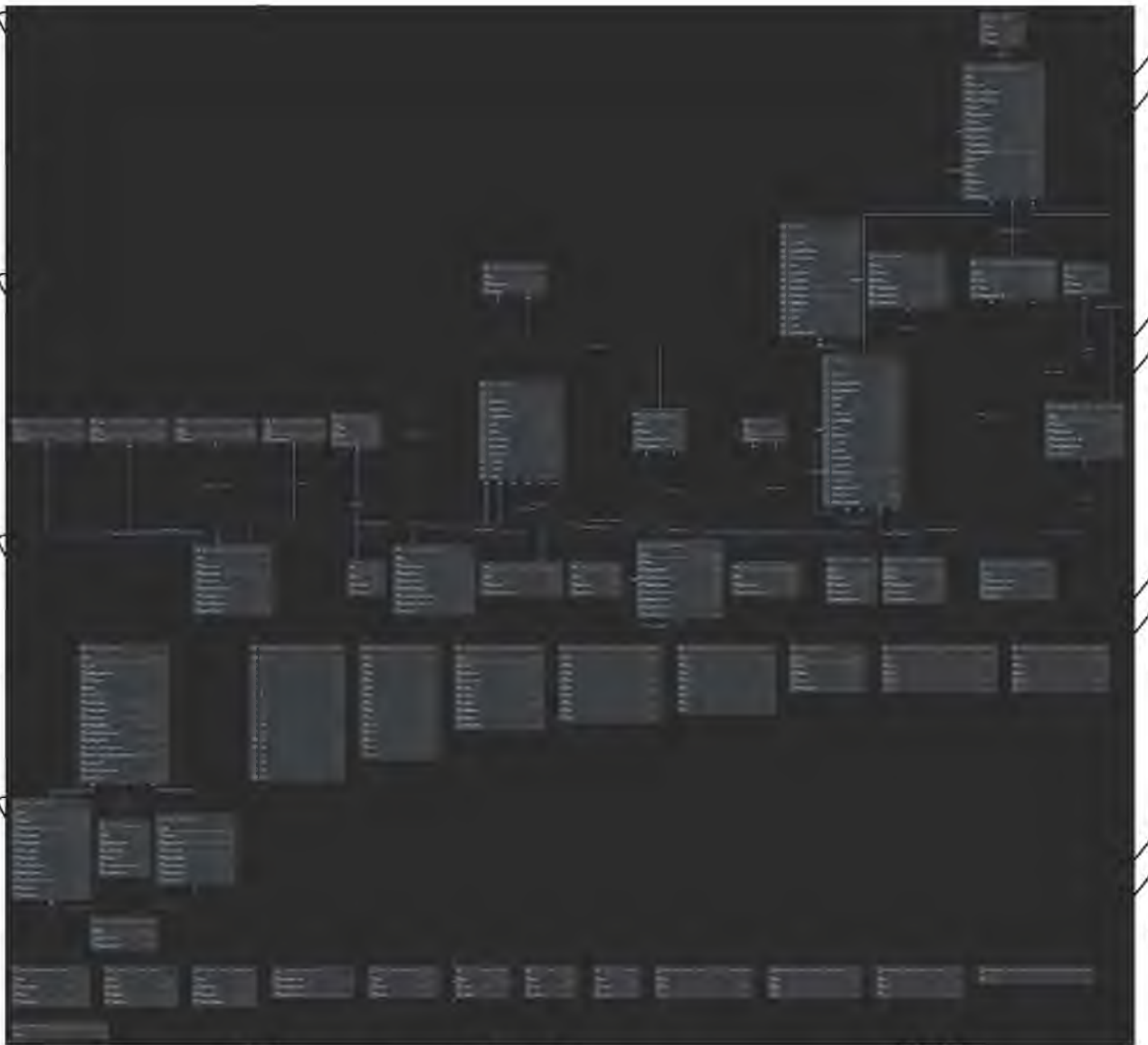


Рис. 3.1/4/ER діаграма

Розроблена база даних зберігає в собі будь-яку інформацію потрібну для роботи системи на накопичення даних для подальшого аналізу. Основники моделями в системі є: Продукт, Категорія, Продавці, Користувач з правами доступу, Замовлення та часові виміри.

Дані до сховища надходять з основної бази даних, що має близько 40 таблиць.

Реляційний зв'язок перебуває в 3НФ, так як воно перебуває в 2 НФ й не містить транзитивних функціональних залежностей непервинних атрибутів від можливих ключів. В моєму випадку досягнення 3НФ можна вважати достатнім для того щоб створити надійну схему бази даних, так як

форми з вищих порядків використовують в більш складних системах й інших підходах.

3.2 Платформа SYMS Marketplace

Для побудови експертної системи було розроблено платформу SYMS Marketplace для подальшого збору даних для аналізу. Інтерфейс програмного забезпечення зображений на рис. 3.2.1-3.2.10.

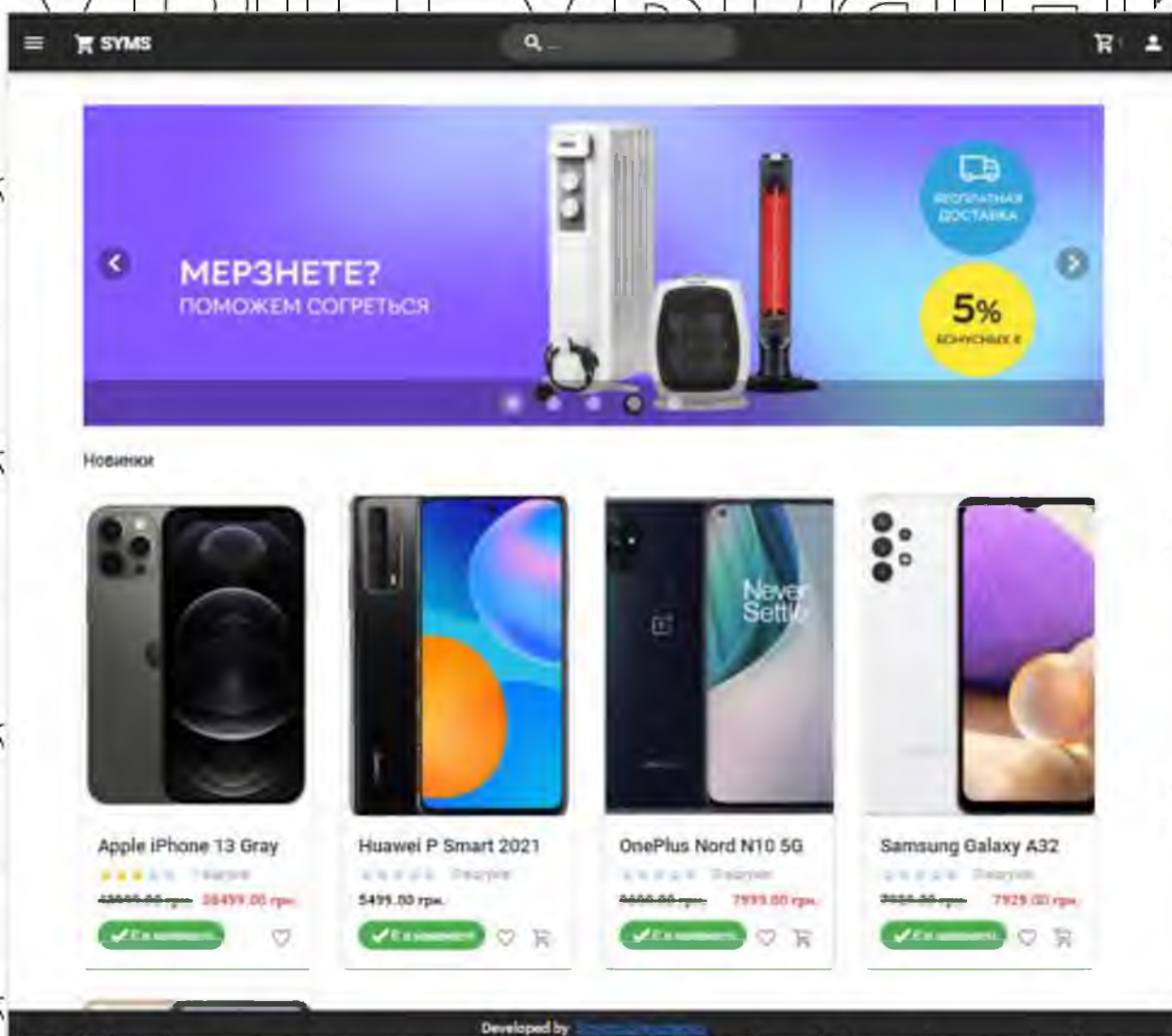


Рис. 3.2.1 Головна сторінка платформи

На рис. 3.2.1 зображена головна сторінка платформи. Що розміщує в собі ця сторінка:

- слайдер з спеціальними пропозиціями;
- відображаються останні новинки;
- можливість обрати потрібну категорію;

НУБІГ Україна

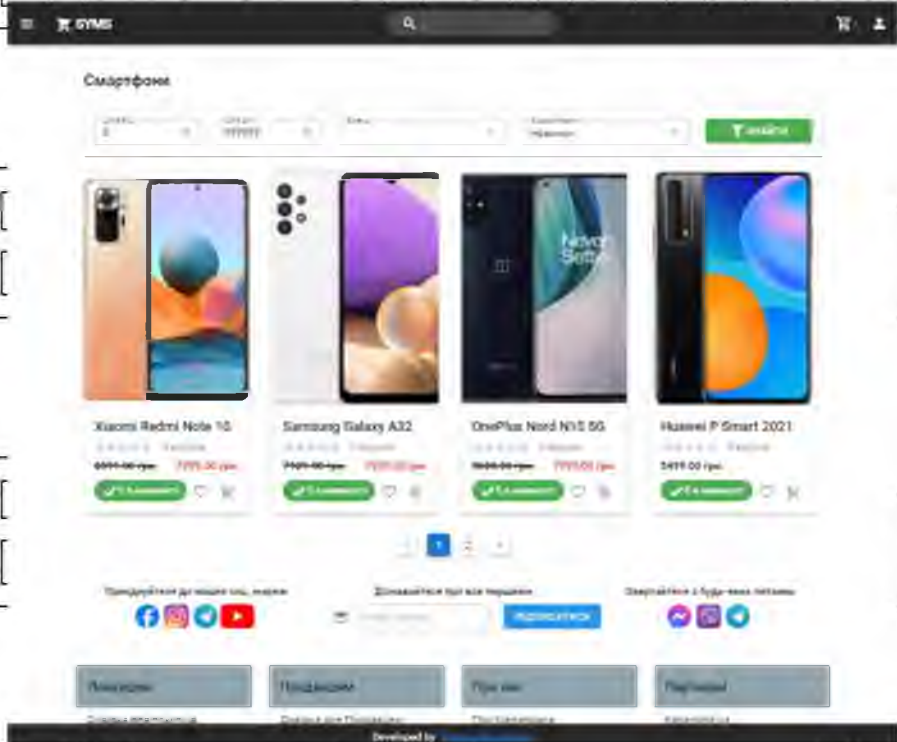


Рис. 3.2.2 Сторінка категорії

На рис. 3.2.2 зображена сторінка категорії. Що розміщує в собі ця сторінка:

- відображення товарів за категорією з влаштованим пагінатором;
- можливість фільтрувати за ціною, брендом;
- можливість сортування за новизною або ціною.



Рис. 3.2.3 Модальне вікно відображення історії замовлень

На рис. 3.2.3 зображено модальне вікно відображення історії. Що розміщує в собі ця сторінка:

- відображення раніше створених замовлень.

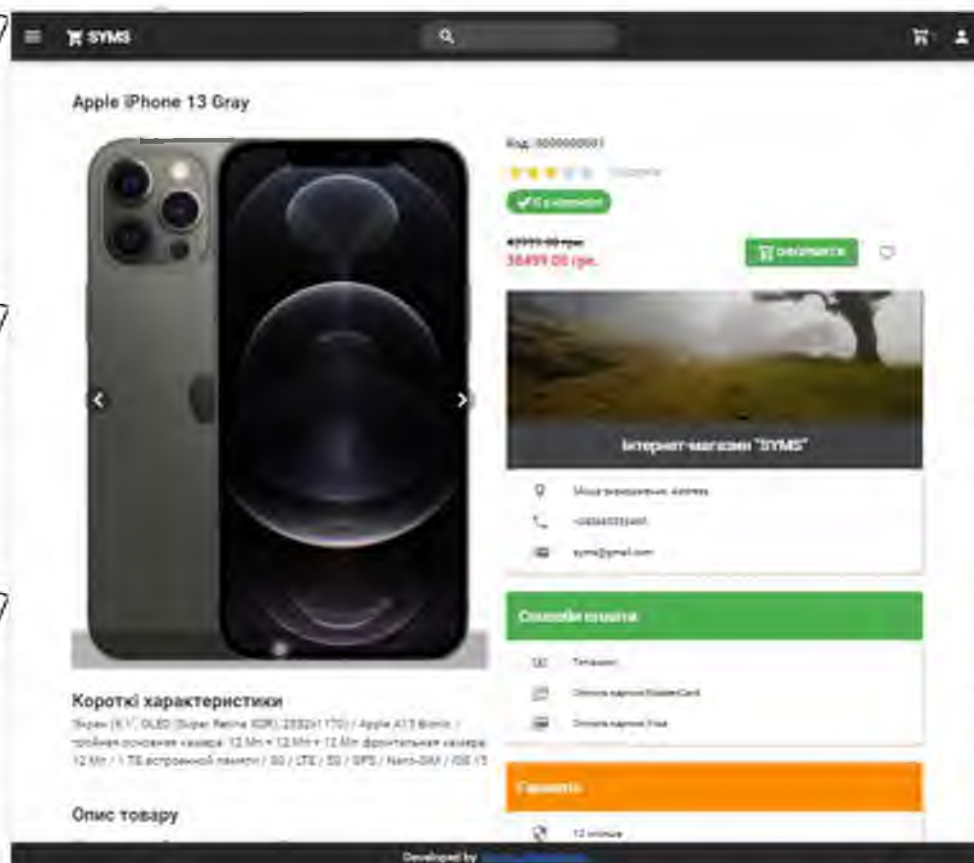


Рис. 3.2.4 Сторінка товару

На рис. 3.2.4 зображено сторінку перегляду товару. Що розміщує в собі ця сторінка.

- відображення інформації за обраним товаром;
- відображення відгуків до товару;
- можливість переглянути опис продавця.

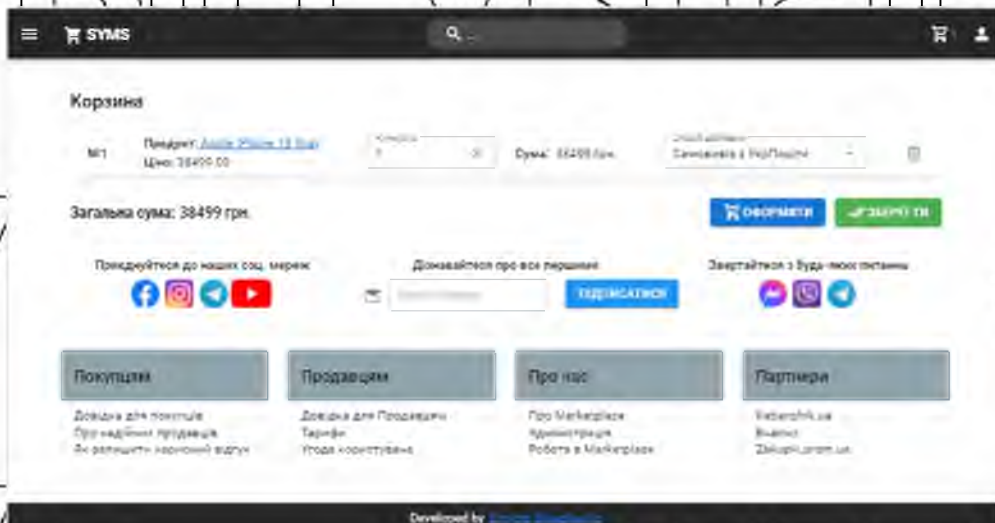


Рис. 3.2.5 Сторінка кошику товарів

На рис. 3.2.5 зображено сторінку кошику товарів. Що розміщує в собі ця сторінка:

- відображення раніше обраних товарів;
- можливість редагування кількості та способів доставки;
- оформлення заявки.

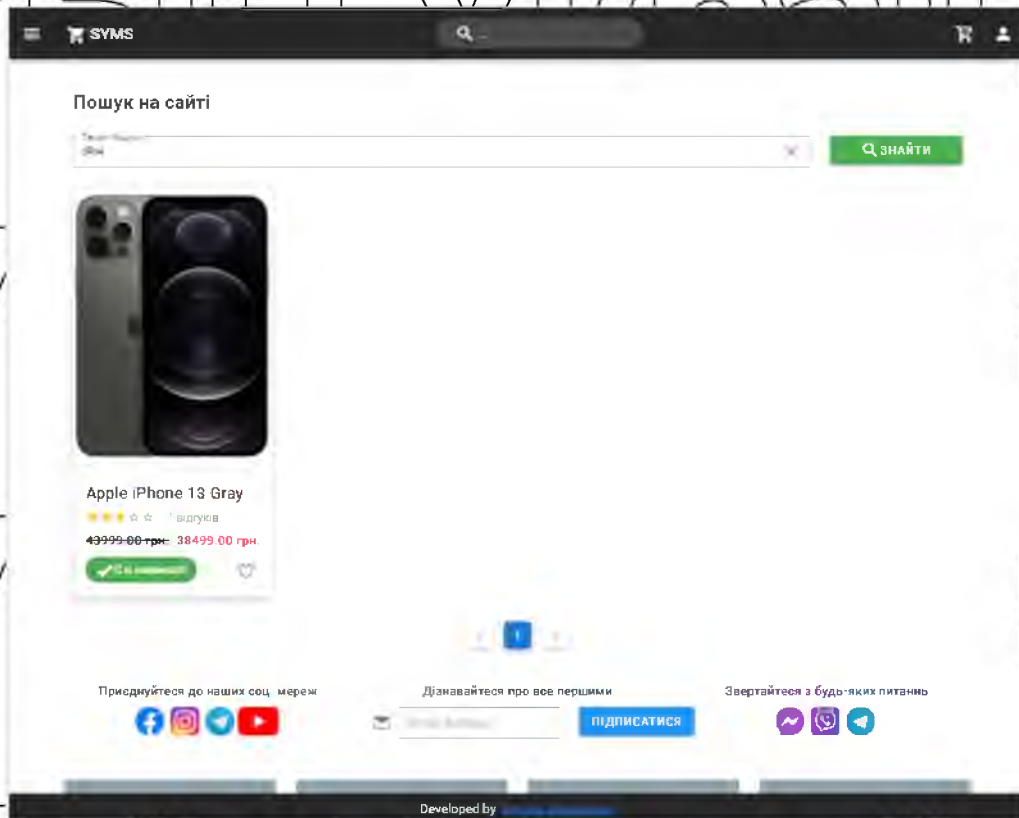


Рис. 3.2.6 Сторінка пошуку товарів

На рис. 3.2.6 зображено сторінку кошику товарів. Що розміщує в собі ця сторінка:

- відображення товарів які відповідають заданому пошуку;
- можливість редагувати свій текст для повторного пошуку.

Рис. 3.2.7 Модальне вікно редагування особистих даних

На рис. 3.2.7 зображено модальне вікно редагування особистих даних. Що розміщує в собі ця сторінка:

- можливість редагування ім'я та прізвища;
- можливість встановлення нового пароля;
- можливість оновити фото свого профілю.

Рис. 3.2.8 Сторінка адміністративного перегляду товарів

На рис. 3.2.8 сторінка адміністративного перегляду товарів. Що розміщує в собі ця сторінка:

- перегляд всіх товарів в системі;
- можливість перейти до управління ним.

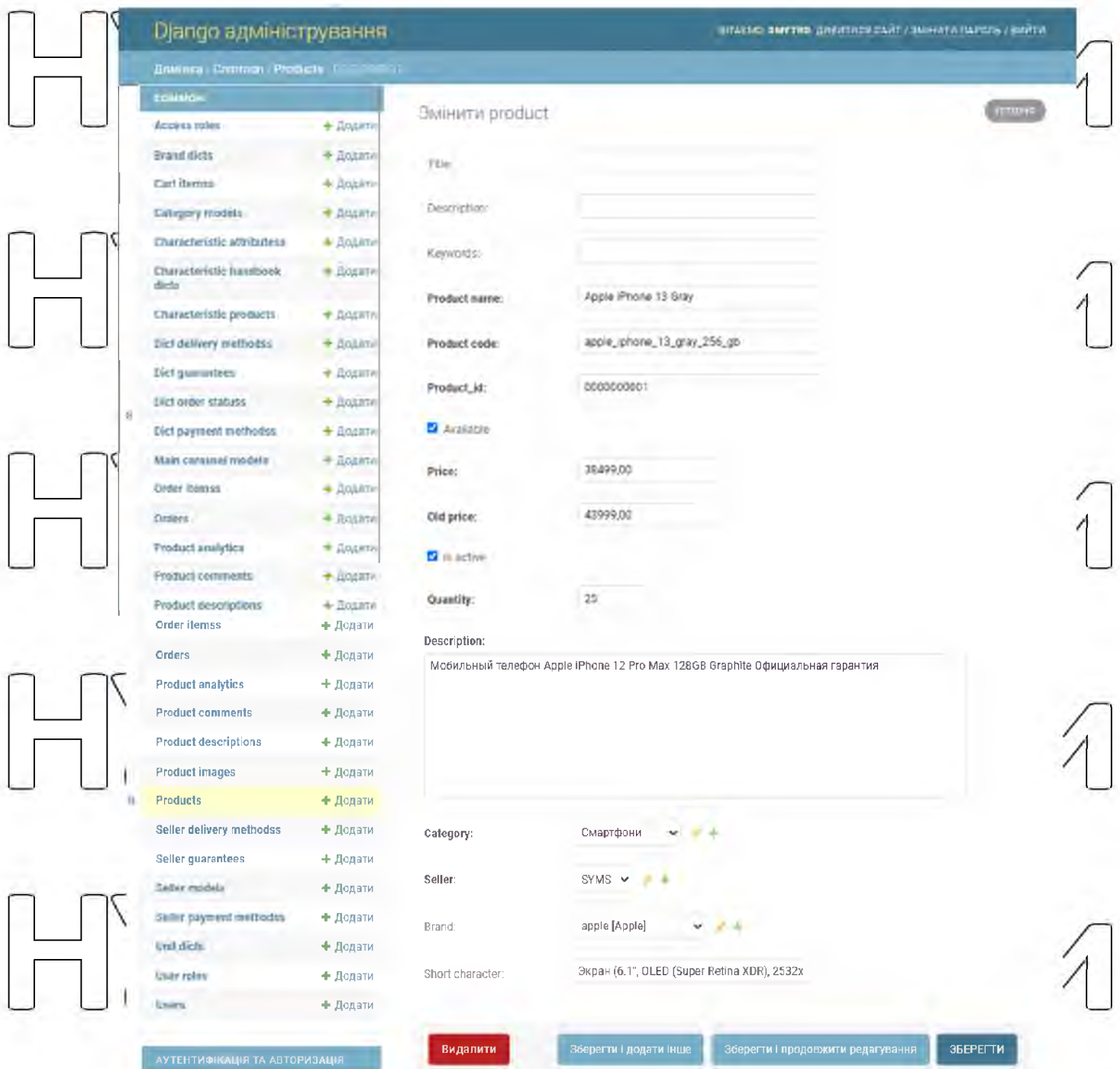


Рис. 3.29 Сторінка адміністративного редагування товару

На рис. 3.2.9 сторінка адміністративного редагування товару. Що розміщує в собі ця сторінка:

- можливість редагування доступних атрибутів товару;
- можливість видалення товару.

НУБІП України



Рис. 3.2.10 Сторінка адміністративного редагування даних

На рис. 3.2.10 зображено модальне вікно редагування особистих даних. Що розміщує в собі ця сторінка:

- можливість управління даними в системі.

3.3 Сховище даних

Сховище даних - предметно орієнтований, інтегрований незмінний набір даних, що підтримує хронологію і здатний бути комплексним джерелом достовірної інформації для оперативного аналізу та прийняття рішень.

Іншими словами, сховище даних – це дуже велика інформаційна корпоративна база даних, спеціально розроблена та призначена для підготовки звітів та аналізу інформації з метою підтримки прийняття рішень в організації.

Дані до сховища потрапляють із транзакційних систем і після

завантаження не змінюються. Як правило, процес завантаження інформації в сховище даних пов'язаний зі значними витратами часу і виконується за певним розкладом, наприклад щодня в нічний час. Зазвичай процес завантаження складається з наступних кроків.

Зазвичай процес завантаження складається з наступних кроків:

- Вилучення даних – дані з транзакційних систем поміщаються в проміжну область сховища.

- Перетворення та завантаження даних – дані з проміжної області перевіряються, перетворюються до одного виду, після чого переміщуються з проміжної області у сховище даних.

- Агрегація даних – цьому етапі відбувається розрахунок підсумкових значень, щоб скоротити час для формування звітів і аналізу даних.

Базова структура дозволяє кінцевим користувачам сховища безпосередньо отримувати доступ до зведених даних, отриманих із вихідних систем, створювати звіти та аналізувати ці дані. Ця структура корисна для випадків, коли джерела даних походять з тих самих типів систем баз даних.

В основі концепції сховища даних (СД) лежить розподіл інформації, що використовують в системах оперативної обробки даних (OLTP) і в системах підтримки прийняття рішень (СППР).

В осередках якого зберігаються аналізовані числові показники (наприклад, обсяги обороту, витрат, інвестицій і т.д.), час, група продукції, регіон, тип процесу (тип клієнта та інших.). При зберіганні ознаки аналізу відокремлюються фактичних даних, утворюючи так звану інвертовану організацію зберігання даних чи структура даних типу "зірка"

Структура сховища даних представлена на рис. 3.3.1.



Рис. 3.3.1 Структура сховища даних

Для зберігання даних ми використали структурну схему, де маємо 3 виміри та одну таблицю фактів:

- CategoryDim – містить інформацію про категорію товару;
- ProductDim – містить інформацію про товари;
- DateDim – фіксує дату;
- SellerDim – містить інформацію про продавців;
- Fact – до таблиці фактів мігрують всі ключі з таблиць вимірів, окрім цього ми зберігаємо інформацію про кількість переглядів та кількість проданих товарів.

4 РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

4.1 Механізм вилучення, обробки і передачі даних

За допомогою SQL Server Business Intelligence, в якому формувався куб, створювались потоки даних для отримання даних з джерел на заповнення створеного кубу.

Одним із кроків є створення Data Source View (рис. 4.1.1). Під Data Source View розуміється зріз джерела, яке буде використовуватися для заповнення сховища, при цьому в нього можуть входити як таблиці, так і уявлення (view) реляційної бази - джерела даних.

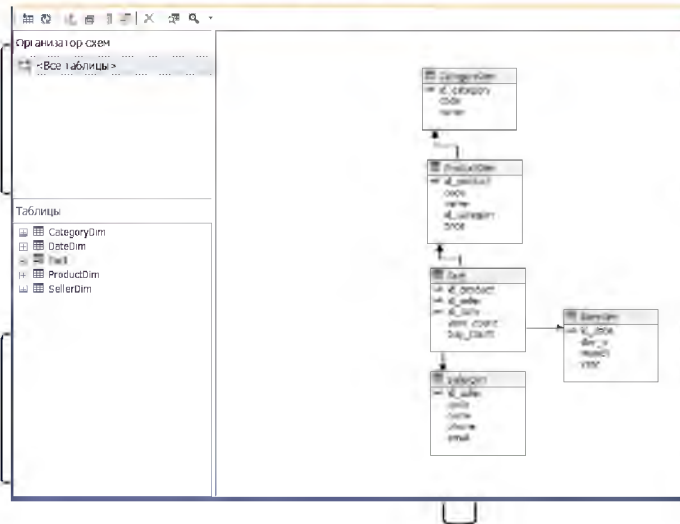


Рис. 4.1.1 Результат Data Source View

Після було розгорнуто куб. На рис. 4.1.2 можна побачити всі розгорнуті виміри, рис. 4.1.3 відображає повномасштабну схему кубу.

- Источники данных
 - SYMS- Marketplace-2014.ds
- Представления источников данных
 - SYMS- Marketplace-2014.dsv
- Кубы
 - SYMS- Marketplace-2014.cube
- Измерения
 - Category Dim.dim
 - Date Dim.dim
 - Product Dim.dim
 - Seller Dim.dim
- Структуры интеллектуального анализа данных
- Роли
- Сборки
- Разное

Рис. 4.1.2 Результат розгортання кубу

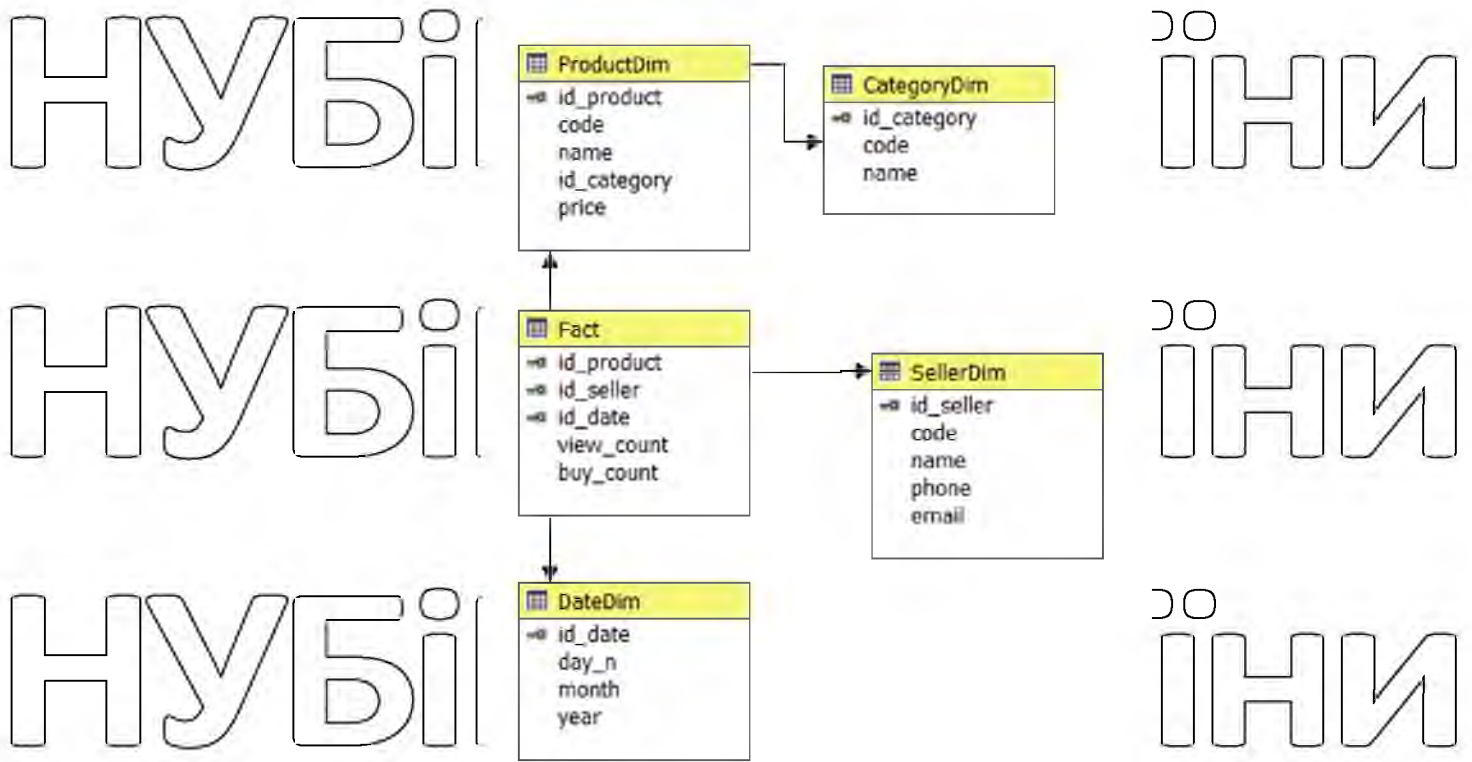


Рис. 4.1.3 Розгорнутий куб

Отримання даних з джерела та заповнення згенерованого кубу було виконане за допомогою Data Flow. Для цього було використано службу SSIS (SQL Server Integration Services)

Було створено задачу потоку даних для першого (рис. 4.1.4), другого (рис. 4.1.5) та третього рівня (рис. 4.1.6). На скріншотах представлено запуснений проект, де видно позитивний результат виконання та загальну структуру (рис. 4.1.7).



Рис. 4.1.4 Задача потоку даних першого рівня



Рис. 4.1.5 Задача потоку даних другого рівня



Рис. 4.1.6 Задача потоку даних третього рівня



Рис. 4.1.7 Загальна структура та результат

4.2 Використані методи Data Mining

Алгоритм 1-Rule. В аналізі даного алгоритму було визначено два класи за якими ми будемо класифікувати дані:

Висока продуктивність (H) – коли рівень продаж є більшим за вказаний показник.
 Низька продуктивність (L) – коли рівень продаж є нижчим за вказаний показник.

Незалежні змінні – Продавець, Категорія, Продукт.
 Залежна – Рівень продажу

Для відображення результатів задачі класифікації за цим алгоритмом було розроблено аналітичний модуль, який зображено на рис. 4.2.1 – 4.2.3.

Задача класифікації

Є два класи за якими проводиться класифікація. H - Високий рівень продажу. L - Низький рівень продажу.

Вхідний набір даних	Продавець	Категорія	Продукт	Рівень продажу
1	Rozetka	Компютери та ноутбуки	Aser Aspire 7	H
2	Kozachok	Смартфони	Samsung A30	L
3	Natali	Смартфони	IPhone 12	L
4	EKO	Компютери та ноутбуки	Notebook lenovo 500	L
5	Prom	Смартфони	IPhone 12	H
6	Kozachok	Компютери та ноутбуки	Notebook lenovo 500	L
7	EKO	Спорт	trenajery 2021	L
8	Ericentr	Компютери та ноутбуки	Aser Aspire 7	L
9	ATB	Смартфони	Samsung A30	L
10	Prom	Компютери та ноутбуки	Notebook lenovo 500	H
11	Kozachok	Косметика	perfume 2020	L
12	EKO	Смартфони	Sony pro	L
13	ATB	Смартфони	IPhone 12	L

Рис. 4.2.1 Алгоритм J-Rule

Задача класифікації

Є два класи за якими проводиться класифікація: Н - Високий рівень продажу, L - Низький рівень продажу.

Відний набір даних	Продавець	Категорія	Продукт	Рівень продажу	Всього значень	Задовольняють	Ймовірність
	Rozetka			H	4	4	1.0
	Prom			H	10	3	0.4
	Kozachok			L	8	2	0.2
	ATB			L	3	0	0.0
	Omega			L	3	1	0.3
	METRO			H	5	2	0.6
	EKO			L	5	1	0.2
	Epicentr			L	5	1	0.4
	Sportlike			L	3	1	0.3
	DanceShop			L	3	0	0.0
	Forsage			L	1	0	0.0
	Natali			L	3	1	0.3
	romashka			L	0	0	0

Рис. 4.2.2 Алгоритм 1-Rule

Задача класифікації

Є два класи за якими проводиться класифікація: Н - Високий рівень продажу, L - Низький рівень продажу.

Відний набір даних	Продавець	Категорія	Продукт	Рівень продажу	Всього значень	Задовольняють	Ймовірність
	Смартфони			L	19	1	0.1
	Компютери та ноутбуки			L	12	1	0.2
	Товари для дому			L	8	3	0.6
	Косметика			H	2	1	0.5
	Одяг			H	13	4	0.6

Рис. 4.2.3 Алгоритм 1-Rule

Необхідно відзначити, що алгоритм 1R, незважаючи на свою простоту, в багатьох випадках на практиці виявляється досить ефективним. Це пояснюється тим, що багато об'єктів дійсно можна класифікувати лише по одному атрибуту. Крім того, нечисленність сформованих правил дозволяє легко зрозуміти і використати отримані результати.

В процесі реалізації алгоритму було створено 2 класи для яких було сформовано правила класифікації. У випадку наших правил ми можемо сказати що алгоритм підходить для вирішення нашого завдання.

Метод Наївного Байеса. Метод класифікації, що ґрунтується на застосуванні теореми Байеса з сильним припущенням, що всі предиктори незалежні один від одного. Простіше кажучи, передбачається, що наявність функції у класі не залежить від присутності будь-якої іншої функції у тому ж класі.

НУБІП України

В ході аналізу даного алгоритму було визначено два класи за якими ми будемо класифікувати дані:

- **Висока продуктивність (H)** – коли рівень продаж є більшим за вказаний показник.

НУБІП України

- **Низька продуктивність (L)** – коли рівень продаж є нижчим за вказаний показник.

Двома незалежними змінними є Продавець та Продукт. Результати реалізації методу зображені на рис. 4.2.4 – 4.2.6.

Метод Naive Bayes

Є два класи за якими проводиться класифікація: H – Високий рівень продажу, L – Низький рівень продажу.

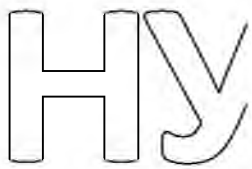
Двома незалежними змінними є Продавець та Продукт

#	Продавець	Продукт	Рівень продажу
1	Kozachok	Samsung A30	L
2	Prom	iPhone 12	H
3	Kozachok	Notebook lenovo 500	L
4	Prom	Notebook lenovo 500	H
5	Rozetka	Chair 90 year 2021	H
6	SportLike	Notebook lenovo 500	L
7	SportLike	Samsung A30	L
8	Kozachok	Notebook lenovo 500	L
9	Kozachok	Chair 90 year 2021	L
10	Kozachok	Samsung A30	L
11	Rozetka	iPhone 12	H
12	Prom	Notebook lenovo 500	H
13	Prom	Samsung A30	H
14	SportLike	iPhone 12	L
15	Prom	Notebook lenovo 500	H

Рівень продажу	Кількість	Ймовірність
H	10	0.67
L	5	0.33

Рис. 4.2.4 Метод Найвого Байєса

НУБІП України



Метод Naive Bayes

Є два класи за якими проводиться класифікація. **H** - Високий рівень продажу. **L** Низький рівень продажу.

Двома незалежними змінними є **Продавець** та **Продукт**

Виміри Данні Розрахунок ймовірності

Продавець

Rozetka

Prom

SportLike

Kozachok

Продукт

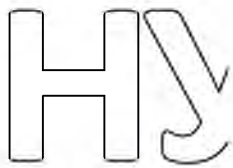
Notebook lenovo 500

iPhone 12

Samsung A30

Chair 90 year 2021

Рис. 4.2.5 Метод Наївного Байєса



Метод Naive Bayes

Є два класи за якими проводиться класифікація. **H** - Високий рівень продажу. **L** Низький рівень продажу.

Двома незалежними змінними є **Продавець** та **Продукт**

Виміри Данні Розрахунок ймовірності

	Рівень продажу		Продукт	Ймовірність	
	H	L		H	L
Rozetka	2	0	Notebook	0.5	0.0
Prom	4	1	iPhone	0.4	0.0
SportLike	7	1	Samsung	0.7	0.4
Kozachok	1	2	Chair	0.3	0.8

	Рівень продажу		Продукт	Ймовірність	
	H	L		H	L
Notebook lenovo 500	0	1	Notebook lenovo 500	0.1	0.0
iPhone 12	1	1	iPhone 12	0.7	0.1
Samsung A30	0	1	Samsung A30	0.3	0.7
Chair 90 year 2021	0	0	Chair 90 year 2021	0.3	0.0

Отже

H Prom Samsung A30 0.00 0.85

L Prom Samsung A30 0.01 0.14

Sum = 1.0

H SportLike Samsung A30 0.0201 0.43

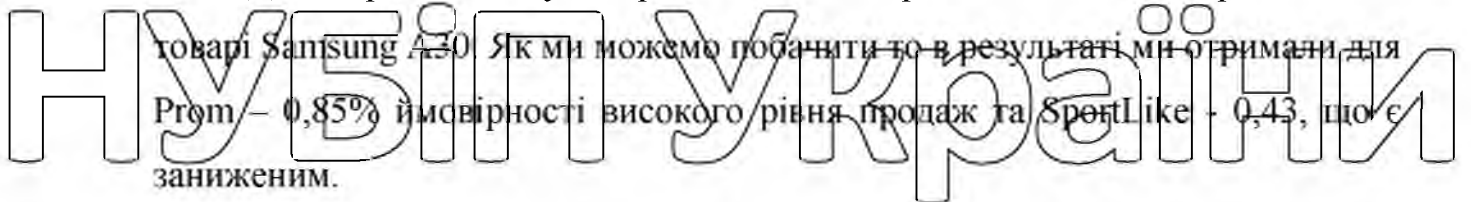
L SportLike Samsung A30 0.026 0.56

Sum = 1.0

Рис. 4.2.6 Метод Наївного Байєса

Метод Наївного Байєса дає можливість більш детально класифікувати ніж алгоритм IR. На практиці цей метод виявляється досить ефективним. Це пояснюється тим, що багато об'єктів дійсно можна класифікувати за декількома незалежними змінними. В ході використання методу було створено 2 класи для яких було сформовано правила класифікації.

Для порівняння було проаналізовано продавця Prom та SportLike на товарі Samsung A30. Як ми можемо побачити то в результаті ми отримали для Prom – 0,85% ймовірності високого рівня продаж та SportLike – 0,43, що є заниженим.



Пошук асоціативних правил. Асоціативні правила є механізмом знаходження логічних закономірностей між пов'язаними елементами (подіями або об'єктами). Відмінність асоціації від двох попередніх задач Data Mining: пошук закономірностей здійснюється не на основі властивостей аналізованого об'єкта, а між декількома подіями, які відбуваються одночасно.

За допомогою майстра інтелектуального аналізу було створено модель для аналізу даних.

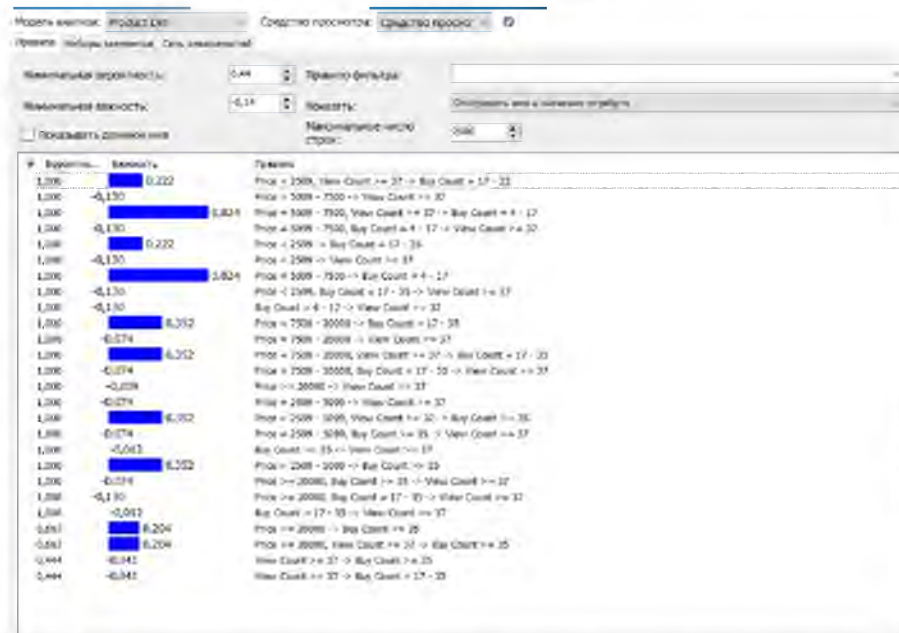


Рис. 4.2.7 Метод Найвого Баїсса

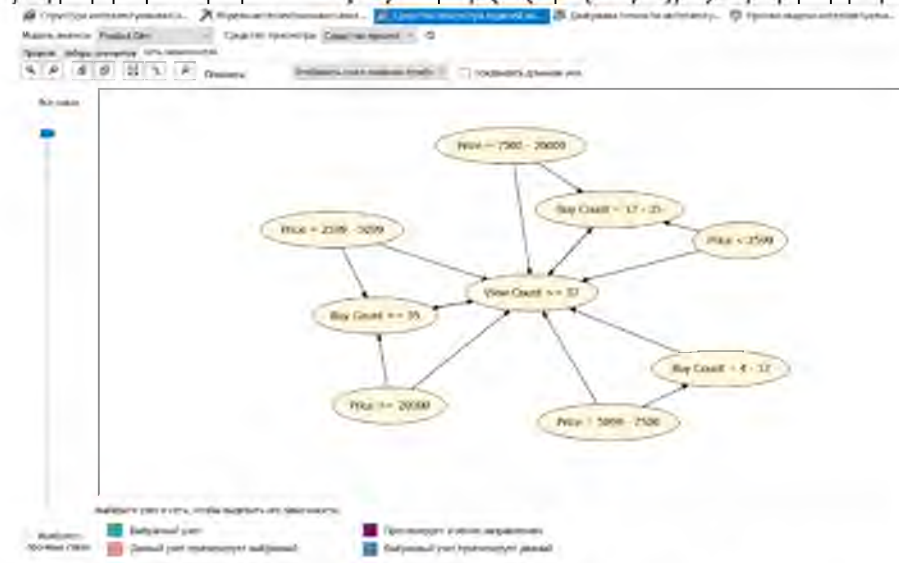


Рис. 4.2.8 Метод Найвого Баїсса

Була побудована модель інтелектуального аналізу даних по алгоритму асоціативних правил. В результаті її рішення визначаються набори товарів, за кількістю переглядів та ціною, що набувають покупцями. Знайдені правила можуть бути використані для вирішення різних завдань, зокрема для розміщення товарів, надання знижок на пари товарів для підвищення об'єму продажів. На рис. 4.2.8 отримано графічне відображення зв'язків між наборами елементів.

Для роботи з моделлю було використане раніше створено інформаційне сховище, та процедури перевантаження даних з транзакційної бази даних в інформаційне сховище.

Візуалізація моделі здійснюється в Visual Studio, представляючи аналітичний і графічний вигляд моделі. Були використанні такі служби MSSQLServer 2014: AnalysisServices, IntegrationServices.

4.3 Розрахунок KPI

У службах SSAS ключовий індикатор продуктивності представляє собою колекцію обчислень, пов'язаних з групою мір куба і використовуються для оцінки успішності бізнесу. Ключовий індикатор продуктивності застосовується для обробки інформації про задану мету, записаної в кубі фактичної формули продуктивності і результати вимірювань, які показують тренд і стан продуктивності.

Для даного дослідження було створено два показники KPI на прикладі середнього значення кількості перегляду та купівлі товару, а саме:

- Середнє значення перегляду товарів по продавцю 20
- Середнє значення кількості проданих товарів по продавцю 10.

На рис. 4.3.1 відображені обраховані KPI, на якому відображено, що середні показники продавців не скрізь задовольняють поставлені KPI, а особливо це стосується середній показник продаж.

Назва сторінки	Кількість переглядів	Середній показник переглядів	Ціль переглядів	Status	Кількість покупок	Середній показник покупок	Ціль покупок	Status
ATB	50	16,67	20	●	13	4,33	10	●
DanceShop	50	16,67	20	●	13	4,33	10	●
EKO	88	17,60	20	●	30	6,00	10	●
Epicentr	113	18,83	20	●	43	7,17	10	●
Forage	20	20,00	20	●	5	5,00	10	●
Kozachok	111	13,88	20	●	55	6,88	10	●
METRO	136	27,20	20	●	49	9,80	10	●
Natali	91	30,33	20	●	49	16,33	10	●
Omega	62	20,67	20	●	23	7,67	10	●
Prom	168	16,80	20	●	73	7,30	10	●
romashka			20				10	
Rozetka	16	4,00	20	●	19	4,75	10	●
SportLike	64	21,13	20	●	24	8,00	10	●
Об'єднаний стор	969	17,94	20	●	396	7,33	10	●

Рис. 4.3.1 Побудовані KPI

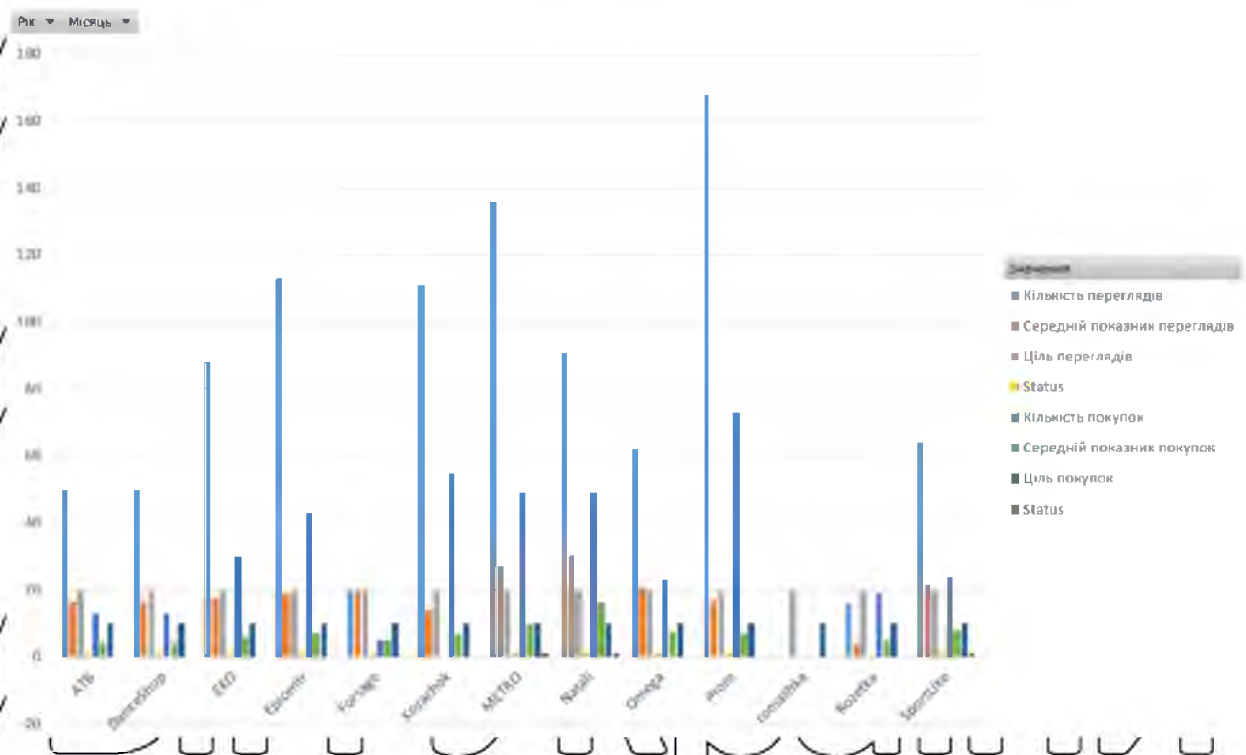


Рис. 4.3.2 Побудовані звіти

KPI дійсно можуть бути корисні не тільки керівництву підприємства, а на них зручно орієнтуватися також і рядовим співробітникам, особливо тим, чий дохід безпосередньо залежить від результатів діяльності.

З отриманих нами результатів можна зробити такий висновок, що потрібно проводити детальний перегляд по продавцю і можливо потрібно попрацювати над кращим наповненням інформації по товарах або рекомендувати покращувати свої послуги по ставленню до клієнта тим самим підвищувати свій рейтинг та довіру.

НУБІП України

НУБІП України

НУБІП України

НУБІП України

НУБІП України

НУБІП України

ВИСНОВОК

В ході виконання наукової роботи було проведено аналіз предметної області та системний аналіз OLAP-технологій. Було побудовано БД об'єкту дослідження та архітектуру систему з всіма необхідними вузлами. Спроектовано і створено сховище даних на основі реляційної бази даних для аналізу перегляду та продажу товарів.

НУБІП України

Для роботи було використано такі інструменти:

- MS SQL Server 2014 (для створення бази та сховища даних)
- SQL Server Data Tools (BI) (служби SSAS та SSIS)

Всі дані мігрували з бази в новостворене сховище даних у СУБД MS SQL Server. Було визначено структуру сховища даних, розгорнуто куб, створено потоки та заповнено куб даними за допомогою Data Flow в середовищі SQL Server BI. Після цього було обчислено KPI та створено звіти в середовищі Microsoft Excel.

Створені звіти відповідають на всі визначені у постановці завдання питання, а обчислене KPI показав яким з продавців клієнти довіряють більше.

Результати дослідження дозволять зробити висновок про доцільність використання технологій OLAP та Data Mining для побудови експертної системи на основі SYMS Marketplace.

Дану експертну систему можна використовувати для аналізу даних в подібних системах в допомогу для прийняття подальших рішень. Результат моєї роботи пришвидшує роботу аналітиків і дає можливість робити аналіз в будь який момент часу.

Планується продовження проведення аналізу OLAP-технологій не тільки для покращення роботи системи, а й можливості реалізації автоматичної системи прийняття рішень.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Branson T. 8 Major Advantages of Using MySQL [Електронний ресурс] / Tony Branson // Datamation Режим доступу до ресурсу: <http://www.datamation.com/storage/8-major-advantages-of-using-mysql.html>.

2. James Rumbaugh, Ivar Jacobson, Grady Booch (1999). The unified modeling language reference manual (англ.). Addison Wesley Longman Inc. ISBN 0-201-30998-X.

3. Silberschatz, Abraham; Korth, Henry F.; Sudarshan, S. (2011). Database system concepts (вид. 6). New York: McGraw-Hill. ISBN 978-0-07-352332-3. OCLC 436031093. Режим доступу -

<https://www.worldcat.org/oclc/436031093>

4. SQL Server Business intelligence [Електронний ресурс]. – Режим доступу: <https://www.microsoft.com/ru-ru/sql-server/sql-business-intelligence>

5. Барсетян А. А., Куприянов М. С., Степаненко В. В., Холод И. И. Методы и модели анализа данных: OLAP и Data Mining. — СПб.: БХВ-Петербург, 2004. — 336 с.

6. Визуальные элементы ключевого показателя эффективности (КПО). Режим доступу - <https://docs.microsoft.com/ru-ru/power-bi/visuals/power-bi-visualization-kpi>

7. Голуб Б.Л., Ящук Д.Ю. Методичні вказівки до виконання курсового проекту з дисципліни «Організація сховища даних» Київ: Національний університет біоресурсів і природокористування України, 2018. 20 с.

8. Мюллер Р.Дж. Базы данных и UML. Проектирование [Текст] / Р.Дж. Мюллер - М.: ЛОРИ, - 2002. — 420с.

9. ОЄ Коваленко Стандартизація формального опису системної архітектури ситуаційних центрів: Інститут проблем математичних машин і систем НАН України, м. Київ, 2015. 4 с. Режим доступу - http://conf.atsuki.org.ua/conf/files/conf_dir_24/Kovalenko_sppr2016.pdf

10. Створення сховищ даних. Технології OLAP та Data Mining [Електронний ресурс]. – Режим доступу: https://pidruchniki.com/16120414/informatika/stvorennva_shovisch_daniv_tehnologi_olap_data_mining

11. Сховища та простори даних : монографія / Н. Б. Шаховська, Е. В. Пасічник ; М-во освіти і науки України, Нац. ун-т «Львів. політехніка». – Д. :

Вид-во Нац. ун-ту «Львів. політехніка», 2009. – 240 с. – Бібліогр.: с. 230–240
(207 назв). – ISBN 978-966-553-796-0.

12. Сховища та простори даних : монографія / Н. Б. Шаховська, В. В.

Пасічник ; М-во освіти і науки України, Нац. ун-т «Львів. політехніка». – Л. :

Вид-во Нац. ун-ту «Львів. політехніка», 2009. – 240 с. – Бібліогр.: с. 230–240
(207 назв). – ISBN 978-966-553-796-0.

13. Codertlessons Навчальний посібник з SSAS [Електронний ресурс] /

Режим доступу до ресурсу: [http://www.datamation.com/storage/8-major-](http://www.datamation.com/storage/8-major-advantages-of-using-mysql.html)

[advantages-of-using-mysql.html](http://www.datamation.com/storage/8-major-advantages-of-using-mysql.html)

14. MySQL Products Офіційна документація [Електронний ресурс] /

Режим доступу до ресурсу: <https://www.mysql.com/products/>

НУБІП України

НУБІП України

НУБІП України

НУБІП України