

Вінницький національний технічний університет
Факультет інформаційних технологій та комп'ютерної інженерії
Кафедра комп'ютерних наук

Пояснювальна записка
до магістерської кваліфікаційної роботи
на тему «Інформаційна технологія підбору покупок»

Виконав: студент 2 курсу,
групи 1КН-18 м
спеціальності 122 «Комп'ютерні науки»
Замковий О. Д.
Керівник: к.т.н., доц. Арсенюк І. Р.
Рецензент: к. т. н., доц кафедри ПЗ
Кательников Д. І.

Вінниця - 2019 року

ЗАТВЕРДЖУЮ
Завідувач кафедри _____ КН _____
д.т.н., проф.. Яровий А.А.

(підпис)
“ _____ ” _____ 2019 року

ЗАВДАННЯ

на магістерську кваліфікаційну роботу на здобуття кваліфікації магістра зі спеціальності: 122 – «Комп'ютерні науки»

08-22.МКР.005.18.000.ПЗ

Магістранта групи 1КН-18м Замкового Олександра Дмитровича

Тема магістерської кваліфікаційної роботи: «Інформаційна технологія підбору покупок»

Вхідні дані: мінімальна кількість товарів – 200; максимальна кількість товарів у списку покупок – 50; можливість гостьового входу та входу за допомогою авторизації: кросплатформеність.

Короткий зміст частин магістерської кваліфікаційної роботи:

1. Графічна: Структурна схема алгоритму функціонування програми; структура інформаційної технології підбору покупок на основі дерева рішень; модель роботи інформаційної технології; схема загального алгоритму функціонування системи; загальна UML-діаграма класів; приклади роботи програми.

2. Текстова (пояснювальна записка): вступ, аналіз предметної області інформаційної технології підбору покупок, розробка інформаційної технології підбору покупок, програмна реалізація інформаційної технології підбору покупок, економічна частина, висновки, перелік використаних джерел, додатки.

КАЛЕНДАРНИЙ ПЛАН ВИКОНАННЯ МКР

№ етапу	Назва етапу	Термін виконання		Очікувані результати
		початок	кінець	
1	Аналіз сучасного рівня розвитку інформаційних технологій підбору покупок. Постановка задач дослідження			Аналітичний огляд літературних джерел, задачі досліджень, розділ 1 ПЗ
2	Розробка методу та інформаційної технології підбору покупок.			Метод, інформаційна технологія, розділ 2
3	Програмна реалізація розробленої інформаційної технології, тестування та оцінка параметрів			Програмне забезпечення, розділ 3
4	Підготовка економічної частини			розділ 4
5	Апробація та/або впровадження результатів дослідження			тези доповідей/акт впровадження
6	Оформлення пояснювальної записки, графічного матеріалу та презентації			Пояснювальна записка, графічний матеріал, презентація

Консультанти з окремих розділів магістерської кваліфікаційної роботи

1. Науковий керівник _____ (підпис)
 “ ____ ” _____ 20__ р.
к. т. н, доц., доц. кафедри КН
 наук. ступінь, вчене звання (посада)
Арсенюк І. Р.
 ініціали та прізвище

2. Економічна частина _____ (підпис)
 “ ____ ” _____ 20__ р.
к. е. н, доц., доц. кафедри ЕПВМ
 наук. ступінь, вчене звання (посада)
Бальзан М. В.
 ініціали та прізвище

Дата попереднього захисту роботи “ ____ ” _____ 20__ р.

Рецензент _____ (підпис)
к. т. н, доц., доц. кафедри ПЗ
 наук. ступінь, вчене звання (посада)
Кательников Д. І.
 ініціали та прізвище

Завдання видав науковий керівник _____ (підпис)
к. т. н, доц., доц. кафедри КН
 наук. ступінь, вчене звання (посада)
Арсенюк І. Р.
 ініціали та прізвище
 “ ____ ” _____ 20__ р.

Завдання отримав магістрант _____ (підпис)
О. Д. Замковий
 ініціали та прізвище
 “ ____ ” _____ 20__ р.

АНОТАЦІЯ

У магістерській кваліфікаційній роботі розроблено та реалізовано інформаційну технологію підбору покупок. Виконано аналіз сучасних програм-аналогів, які використовуються для підбору покупок, в тому числі і продуктів харчування та наведено коротку порівняльну характеристику знайдених програм-аналогів підбору покупок, досліджено методи, які можуть бути використані для реалізації поставленої задачі.

Проведено обґрунтування моделі підбору покупок. Запропоновано використати поєднання алгоритму Баєса та дерева рішень.

Розроблено алгоритм підбору покупок та відповідне програмне забезпечення на мові програмування Java у середовищі Eclipse. Аналіз роботи програмного забезпечення показав достатньо високу швидкість процесу підбору покупок.

ABSTRACT

In this master's qualification work is implemented information technology of selection of purchases. The analysis of modern programs that are used for the selection of purchases, including food, is carried out, and a brief comparative description of the found programs of the analogues of the purchase, the methods that can be used to accomplish the task are investigated.

It is suggested to use a combination of the Bayesian algorithm and the genetic algorithm to increase the speed of the purchasing process.

The result of the design is the program. The application is implemented in Java in Eclipse environment.

ЗМІСТ

1	АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПІДБОРУ ПОКУПОК	11
1.1	Постановка задачі	11
1.2	Поняття споживчого вибору та його класифікація	12
1.3	Характеристика моделі поведінки користувача	14
1.4	Аналіз основних методів вирішення поставленої задачі.....	16
1.4	Аналіз програм-аналогів	22
1.5	Висновок.....	29
2	РОЗРОБКА МОДЕЛІ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПІДБОРУ ПОКУПОК	30
2.1	Дослідження можливостей методу підбору покупок	30
2.2	Обґрунтування вибору методу наївного байєсівського класифікатора для інформаційної технології підбору покупок	31
2.2	Обґрунтування вибору методу дерева рішень для інформаційної технології підбору покупок	35
2.4	Розробка математичної моделі підбору покупок	38
2.5	Проектування структури інформаційної технології підбору покупок.....	42
2.6	Розробка ER-моделі	45
2.7	Висновок	46
3	ПРОГРАМНА РЕАЛІЗАЦІЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПІДБОРУ ПОКУПОК ТА ЇЇ ПРОГРАМНА РЕАЛІЗАЦІЯ.....	47
3.1	Обґрунтування вибору мови програмування.....	47
3.2	Програмна реалізація інформаційної технології підбору покупок	50
3.3	Тестування інформаційної технології інтелектуального підбору покупок та аналіз результатів її роботи.....	52
3.4	Висновок.....	55
4	ЕКОНОМІЧНА ЧАСТИНА.....	56
4.1	Оцінювання комерційного потенціалу розробки	56

4.2 Прогнозування витрат на виконання науково-дослідної роботи та конструкторсько–технологічної роботи.....	59
4.3 Прогнозування комерційних ефектів від реалізації результатів розробки	64
4.4 Розрахунок ефективності вкладених інвестицій та період їх окупності ...	67
4.5 Висновок	70
ВИСНОВКИ.....	72
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ	74
ДОДАТКИ.....	Ошибка! Закладка не определена.
ДОДАТОК А ІНСТРУКЦІЯ КОРИСТУВАЧА	Ошибка! Закладка не определена.
ДОДАТОК Б ЛІСТИНГ ПРОГРАМИ	Ошибка! Закладка не определена.
ДОДАТОК В ГРАФІЧНА ЧАСТИНА	Ошибка! Закладка не определена.

ВСТУП

Актуальність теми дослідження. Споживчий ринок - це структура, що швидко розвивається. За наявності великої кількості потенційних покупців її поведінку важко моделювати.

Бізнес досягає успіху, якщо він пропонує споживачеві вибір або реальну вигоду. Розуміння та постійне пристосування до поведінки клієнтів є однією з передумов конкурентоспроможності компанії.

Вивчення мотивації споживачів та поведінки у процесі вибору та купівлі товарів проводиться за допомогою моделювання. Наскільки різні покупці, настільки різна їх поведінка на ринку, але є одне спільне: споживачі можуть однаково реагувати на різні характеристики товару.

Сьогодні інтелектуальні помічники все частіше використовують для вибору різних типів ринків. Ці системи дозволяють користувачам здійснювати всі необхідні покупки для свого пристрою чи чогось іншого, і ці заходи дають виробнику чи продавцю додатковий дохід, який становить значний відсоток від загальної суми.

В наш час існує багато супермаркетів, які купують продукцію від виробника великими партіями. З одного боку, якщо партія має більший обсяг, то ціна товару нижча і прибуток вищий, але з іншого боку, товар може з часом погіршуватися, і це призведе до величезних втрат. Одне з найактуальніших завдань - забезпечити закупівлю певного товару певним споживачем. З цією метою рекомендується використовувати розумну систему покупок, яка зможе проаналізувати покупки конкретного клієнта та вибрати їх покупки на майбутнє, щоб ви могли зробити конкретний прогноз щодо обсягу продажів. Це дозволить точніше визначити обсяг поставок, що може призвести до зниження витрат та збільшення прибутку. Практично всі супермаркети вже використовують посвідчення особи споживачів. Це дозволяє створити надійну базу даних споживачів та продуктів, які вони купують.

Оскільки сучасний темп життя унеможлиблює виконання конкретних завдань протягом тривалого часу методика планування набирає популярності. При чому планують усе: від відрядження, походу до лікаря, бюджету чи списку покупок.

Планування останнього є досить поширеним явищем на всіх рівнях фінансової піраміди, оскільки це допомагає уникнути зайвих покупок, заощадити кошти, та не забувати, що насправді потрібно.

Майже щодня в продуктовому магазині покупці забувають придбати часто вживаний продукт, який зараз саме закінчився. У цьому випадку важливо скласти список покупок на певний день або тиждень. У нас не завжди папір і ручка в руці, але 90% матиме смартфон. Щоб спростити процес списку покупок, ви можете використовувати його як нотатки, що легко виносяться на робочий простір чи спеціалізованими програмними додатками, які мають багато функцій для полегшення процесу складання списку покупок.

Зв'язок роботи з науковими програмами, планами, темами. Магістерська робота виконана відповідно до напрямку наукових досліджень кафедри комп'ютерних наук Вінницького національного технічного університету 22 К1 «Моделі, методи, технології та пристрої інтелектуальних інформаційних систем управління, економіки, навчання та комунікацій» та плану наукової та навчально-методичної роботи кафедри.

Мета та завдання дослідження. Метою дослідження магістерської кваліфікаційної роботи є підвищення швидкості підбору покупок.

Для досягнення наведеної мети були поставлені та вирішені наступні задачі:

- 1) розглянути та проаналізувати існуючі методи та технології розв'язання задачі підбору покупок;
- 2) дослідити перелік необхідних функцій, які повинна містити інформаційна технологія;
- 3) запропонувати математичну модель для інформаційної технології підбору покупок;

- 4) виконати програмну реалізацію запропонованої інформаційної технології підбору покупок;
- 5) провести тестування програмного продукту та виконати аналіз отриманих результатів.

Об'єкт дослідження - є процес підбору покупок.

Предмет дослідження - інформаційна технологія підбору покупок.

Методи дослідження. У роботі використані наступні методи наукових досліджень: системного аналізу для аналізу структури інформаційної системи, метод наївного Баєсівського класифікатора, метод дерева рішень для задачі підбору покупок, об'єктно-орієнтованого програмування для автоматизації розрахунків.

Наукова новизна одержаних результатів полягає в наступному:

– вдосконалено технологію підбору покупок в тому числі для розумного дому, що відрізняється від відомих застосуванням поєднання методу наївного Баєсівського класифікатора та методу дерева рішень для підбору покупок та забезпечує ефект підвищення швидкості підбору покупок.

– вдосконалено модель підбору покупок шляхом сумісного застосування Баєсівського класифікатора та дерева рішень, що забезпечує підвищення швидкості отримання вихідних даних.

Практичне значення одержаних результатів полягає у наступному:

1. Розроблено алгоритм що реалізує підбір покупок, який використовує поєднання методу наївного Баєсівського класифікатора та методу дерева рішень.
2. Розроблено програмне забезпечення для підбору покупок.

Отриманий досвід та результати можуть потенційно застосуватися для вирішення задач підбору покупок для користувачів будь-яких гаджетів, а також розумних холодильників.

Розроблені алгоритми можуть бути впроваджені в начальний процес як основа лекції на тему «Розв'язання задачі пошуку рішень» дисциплін «Методи та засоби штучного інтелекту» та «Теорія прийняття рішень».

Достовірність теоретичних положень магістерської кваліфікаційної роботи підтверджується строгістю постановки задач, коректним застосуванням математичних методів під час доведення наукових положень, строгим виведенням аналітичних співвідношень, порівнянням результатів з відомими, та збіжністю результатів математичного моделювання з результатами, що отримані під час впровадження розроблених програмних засобів.

Особистий внесок магістранта. Усі результати, наведені у магістерській кваліфікаційній роботі, отримані самостійно.

Апробація результатів роботи. Результати роботи були апробовані на всеукраїнській науково-практичній інтернет-конференції студентів, аспірантів та молодих науковців «Молодь в науці: дослідження, проблеми, перспективи (МН 2019)» (м. Вінниця, Україна, 2019 р.) [1]; XLVII науково-технічній конференції ВНТУ [2]; XI міжнародній науково-практичній конференції «ІОН-2018» [3].

Публікації. За результатами магістерської кваліфікаційної роботи опубліковано троє тез доповідей міжнародних науково-практичних конференцій [1], [2], [3].

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПІДБОРУ ПОКУПОК

1.1 Постановка задачі

У цій роботі необхідно реалізувати інформаційну технологію підбору покупок.

Споживач - фізична особа, яка купує, замовляє, використовує або має намір придбати або замовити товари для особистих потреб, які безпосередньо не пов'язані з господарською діяльністю або виконують зобов'язання працівника.

Розумний шопінг - це нетривіальне завдання для оптимального рішення, яке не існує. Тому методи часто використовуються для того, щоб забезпечити оптимальне поєднання швидкості алгоритму та критерії вибору продукту, оскільки швидкість алгоритму зменшується через велику кількість обчислень.

Отже, потрібно розробити інформаційну технологію, яка реалізує інтелектуальний алгоритм вибору покупок, який працює з прийнятною швидкістю.

У цій роботі вихідні дані - це список продуктів, складених користувачем за допомогою інтелектуального алгоритму покупки.

Розробляючи програму, необхідно звернути увагу на функції, які вона повинна виконувати, а саме: якість та швидкість підбору.

Інформаційна технологія підбору покупок повинна мати наступні вихідні дані: список покупок, складених користувачем за допомогою інтелектуального алгоритму покупки.

Це має полегшити складання списків покупок для людей, які постійно зайняті або часто щось забувають купити, тобто для споживачів. Інформаційна технологія - це інструмент, який забезпечить зручний та ефективний підбір продуктів в кошику споживача на відміну від неякісних або важко зрозумілих аналогів.

Проблема полягає у відсутності простого, безкоштовного і водночас багатofункціонального та продуктивного інструменту для підбору покупок. Вирішення проблеми підвищить популярність створення списків покупок за допомогою мобільних додатків, створить ефективний та зручний інструмент для підбору покупок.

Тому для вирішення проблеми підбору покупок необхідно вирішити такі основні завдання:

1. Обґрунтувати вибір методу підбору покупок.
2. Розробити математичну модель.
4. Спроекувати структуру інформаційної технології.
5. Розробити базу даних покупок.
6. Програмно реалізувати інформаційну технологію.
7. Провести тестування інформаційної технології.

Отже, потрібно розробити інформаційну технологію, в якій буде реалізовано алгоритм підбору покупок, які зберігаються в базі даних, а також виводити список з підібраними продуктами.

1.2 Поняття споживчого вибору та його класифікація

Вибір товарів та послуг для споживання, тобто вибір споживача, залежить насамперед від наших потреб та уподобань, звичок, традицій, тобто наших переваг.

Вибір користувача залежить від уподобань клієнта. Цей вибір вважається найкращим поєднанням благ (або груп користувачів) усіх можливих комбінацій. Кращу в тому сенсі, що цей споживчий набір приносить покупцеві найбільше переваг.

Поведінка споживачів - це процес формування попиту на ринку, який приймає рішення, виходячи з існуючих цін.

Вибір споживача - це сукупність переваг, які приносять максимальну загальну вигоду споживачеві під бюджетні обмеження [5].

Окрім загальних принципів вибору раціонального споживача, є характеристики, які визначають вплив на смак та уподобання (рисунок 1.1).

Функціональний попит - це та частина попиту, яка керується характеристиками споживачів, притаманними економічному благу (товар чи послуга).

Нефункціональний попит є частиною попиту через чинників, які безпосередньо не пов'язані з притаманними їм економічними вигодами.

Соціальні, спекулятивні та ірраціональні фактори можна відрізнити від нефункціонального попиту з деякою пропорційністю.

Перший пов'язаний із ставленням покупців до товару. Деякі прагнуть підтримувати загальний стиль і купувати те, що купують ті, на кого вони рівняються [3].

Ефект приєднання до більшості відноситься до ефекту збільшення попиту споживачів, через те, що споживач, дотримуючись загальноприйнятих норм, купує ті ж товари, що й інші.

У випадку снобічного ефекту у споживача переважає бажання виділитися з натовпу [4].

Спекулятивний попит виникає в суспільстві з високими інфляційними очікуваннями, коли ризик зростання цін у майбутньому стимулює додаткове споживання (купівлю) товарів у сьогодні.

Ірраціональний пошук - це незапланований пошук, який виник під впливом миттєвого бажання, раптової зміни настрою, примхи чи примхи, попиту, що порушує передумови раціональної поведінки споживачів. Слід зазначити, однак, що багато людей більш-менш схильні до поривів нерегулярного попиту і часто роблять покупки, про які вони в майбутньому часто шкодують [5].

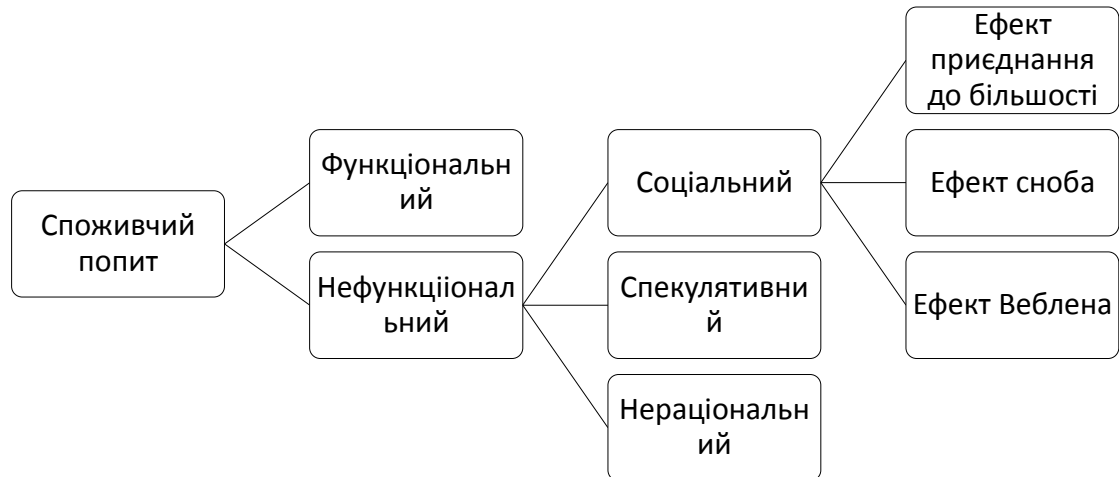


Рисунок 1.1 – Класифікація споживчого вибору

Ефект приєднання до більшості відноситься до ефекту збільшення попиту споживачів, через те, що споживач, дотримуючись загальноприйнятих норм, купує ті ж товари, що й інші.

У випадку снобічного ефекту у споживача переважає бажання виділитися з натовпу [4].

Спекулятивний попит виникає в суспільстві з високими інфляційними очікуваннями, коли ризик зростання цін у майбутньому стимулює додаткове споживання (купівлю) товарів у сьогодні.

Ірраціональний пошук - це незапланований пошук, який виник під впливом миттєвого бажання, раптової зміни настрою, примхи чи примхи, попиту, що порушує передумови раціональної поведінки споживачів. Слід зазначити, однак, що багато людей більш-менш схильні до поривів нерегулярного попиту і часто роблять покупки, про які вони в майбутньому часто шкодують [6].

1.3 Характеристика моделі поведінки користувача

Основне завдання теорії поведінки споживачів - дослідити проблему реалізації доходу покупців та отримати найбільшу користь для себе. Теорія поведінки споживачів визначає умови та ступінь впливу на вибір переваг

закупівлі на ціни товарів, величину доходу, а також способи максимальної вигоди покупців від придбання певних товарів і послуг.

Категорії споживачів та корисні послуги лежать в основі теорії поведінки споживачів [5].

Передумови поведінки споживачів можна сформулювати так:

- споживачі дуже добре знають, яка продукція найкраща для них;
- споживачі діють раціонально;
- споживачі точно оцінюють рівень своїх доходів та знають ціни на товари;
- вибір споживачів обмежений їх доходами та часом.

Поведінка споживача - це процес формування ринкового попиту покупців, які здійснюють вибір благ з урахуванням існуючих цін [6].

Наш вибір товарів і послуг для споживання, тобто вибір споживача, залежить, перш за все, від наших потреб і смаків, звичок, традицій, тобто від наших переваг [7].

Вподобання покупця суб'єктивні. Корисність кожного обраного елемента також є суб'єктивною. Однак вибір споживача визначається не тільки його вигодами, але й ціною обраної продукції та її доходом [8]. Як і в економіці, індивідуальні споживчі ресурси обмежені. Практично необмежені потреби споживачів та обмежені ресурси обумовлюють необхідність вибору з різних поєднань переваг, тобто. потреба у виборі споживачів [9].

Корисність блага - це задоволення, яке відчуває людина, споживаючи добро; Корисність заснована на різних фізичних, хімічних, біологічних та інших властивостях блага.

Передбачається, що споживач товару якимось чином визначає ступінь корисності споживання товару і, знаючи корисність різних переваг, він може зробити вибір. Цей вибір товару повинен бути найкращим на його думку, тобто приносити йому найбільшу корисність, найбільшу ступінь задоволеності [10].

Можна узагальнити деякі принципи поведінки споживачів на ринку, тобто закономірність його поведінки:

- вибираючи товар для споживання, покупець орієнтується на його уподобання;
- поведінка споживачів є раціональною, зокрема вона ставить конкретні цілі і керується власним інтересом, тобто діє в рамках розумного егоїзму;
- користувач прагне до максимальної корисності, іншими словами, прагне вибрати набір переваг, які приносять йому найбільше загальне значення корисності;
- при виборі товару можливості споживача обмежуються цінами товару та його доходами; ця межа називається бюджетним обмеженням [11].

Модель поведінки споживачів пов'язана із загальними принципами поведінки споживачів на ринку, які включають, перш за все, максимальне використання сукупної корисності, граничне законодавство про корисні послуги та обмеження бюджету [12].

1.4 Аналіз основних методів вирішення поставленої задачі

Дерева рішень

За більшістю загальних визначень дерево рішень є інструментом підтримки прийняття рішень при прогнозуванні, який широко використовується в статистиці та аналізі даних.

Дерево рішень, подібне до свого "прототипу" живої природи, складається з "гілок" та "листя" (мал. 2.1). В гілках (ребрах графіка) зберігаються значення атрибутів, від яких залежить цільова функція; значення цільової функції записується на аркуші. Існують також інші вузли - батьківський і потомський, - за якими відбувається розгалуження, і випадки можна виділити [18].

Метою всього процесу побудови дерева рішень є створення моделі, за якою класифікувати випадки та вирішувати, які значення може взяти цільова функція, маючи на вході кілька змінних [19].

Дерева рішень можуть бути ефективно застосовані до даних із відсутніми значеннями, що дуже корисно для вирішення практичних завдань, коли пропущені значення - це правило, а не виняток [20].

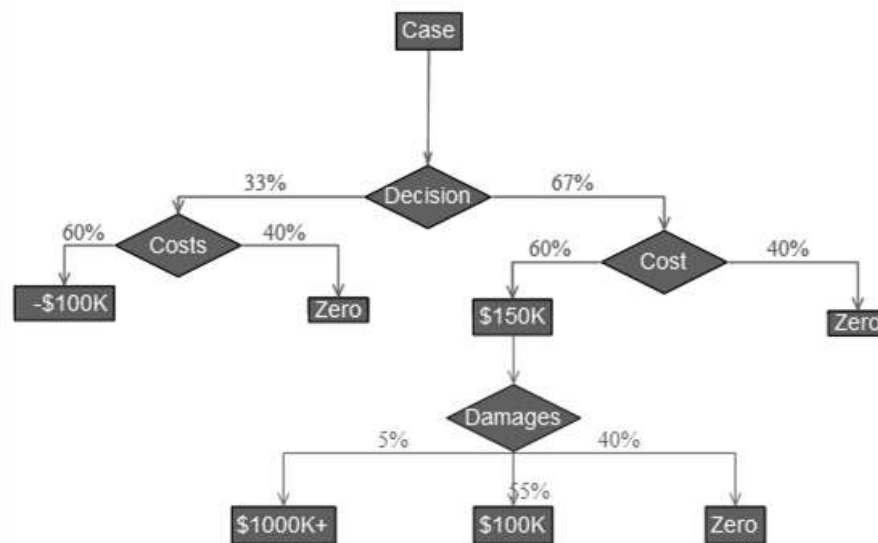


Рисунок 1.2 – Приклад дерева рішень

Баєсові методи були розроблені в результаті численних спроб вчених визначити проблеми статистичного аналізу поведінки різних процесів та знайти їх рішення, застосовуючи основу байєсівської методології - теорему Байєса. Використання цієї теореми має ряд передумов, основна з яких - наявність певних зв'язків між ймовірністю явищ різної природи та специфікаціями будь-якого явища на бажаному рівні [26].

Байєсівська методологія відрізняється від інших підходів тим, що ще до отримання даних дослідник визначає рівень своєї впевненості у можливих моделях і згодом подає їх у вигляді певних ймовірностей [27].

Детальніше розгляньте переваги та недоліки класифікатора Байєса:

Позитивні сторони:

Класифікація, включаючи багатокласну, проста та швидка. Це дозволяє передбачити ймовірність для кількох значень цільової змінної.

Коли припущення щодо незалежності виконуються, НБК перевершує інші алгоритми, такі як логістична регресія, що вимагає менших даних про навчання

НБК працює краще за категоричними ознаками, ніж тверді. Для безперервних ознак передбачається нормальний розподіл, що є досить сильним припущенням.

Класифікація в режимі реального часу. НБК вчиться дуже швидко, саме тому його можна використовувати для обробки даних у режимі реального часу.

Недолік:

Якщо тестовий набір даних має певне значення атрибута, він не відображається в навчальному наборі даних, тоді встановить модель на нуль, і це значення навряд чи зможе зробити передбачення. Це явище відоме

називається "нульовою частотою". Цю проблему можна вирішити згладжуванням. Ще одне обмеження НБК - прийняття функції незалежності. Насправді набори абсолютно незалежних функцій надзвичайно рідкісні.

Лінійна регресія

Алгоритм лінійної регресії - це свого роду алгоритм дерева рішень, який допомагає обчислити лінійну залежність між залежною та незалежною змінною, а потім використовувати цей взаємозв'язок при прогнозуванні або виборі покупок [23].

Вибираючи алгоритм лінійної регресії, називається спеціальний варіант алгоритму дерева рішень з параметрами, що обмежують поведінку алгоритму та вимагають використання певних типів вхідних даних. Більше того, у лінійній регресійній моделі весь набір даних використовується для обчислення з'єднань при початковому проході; тоді як у стандартній моделі дерева рішень дані неодноразово розбиваються на менші підмножини або дерева [24].

Цільова функція лінійної регресійної моделі

$(y = f(x, b) + \varepsilon, E(\varepsilon) = 0$, де b — параметри моделі, ε — випадкова похибка моделі) $f(x, b)$ має вигляд:

$f(x, b) = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k$, де b_j — параметри (коефіцієнти) регресії, x_j — регресори (фактори моделі), k — кількість факторів моделі [25].

Генетичні алгоритми - це нова тенденція в алгоритмах, вони не тільки здатні зменшити бюст у складних завданнях, але і легко адаптуються.

Генетичний алгоритм використовується для вирішення задач оптимізації за допомогою еволюційного методу, тобто шляхом вибору з набору потенційно можливих рішень у певному сенсі найкращих. У типовій проблемі оптимізації вони ідентифікують набір змінних, які впливають на процес, та будують відповідну формулу або / та алгоритм, який використовує ці змінні для побудови моделі процесу. При цьому проблема полягає у знаходженні значень змінних, що оптимізують рішення задачі (задачі). Зокрема, коли модель є формулою, вона зазвичай шукає максимум або мінімум функції, яку описує формула.

Сьогодні генетичні алгоритми використовуються для вирішення задач у таких напрямках: екстремальні проблеми (знаходження точок мінімуму та максимуму), проблеми оптимального розподілу інвестиційних фондів, проблеми найкоротшого шляху (завдання продавця), задачі планування, планування, наближення функцій, вибір) введення, штучні параметри нейронної мережі, біоінформатика, ігрові стратегії, нелінійна фільтрація тощо.

Генетичний алгоритм є більш ефективним інструментом пошуку, ніж класичні методи оптимізації в тих випадках, коли область можливих рішень може бути достатньо великою і не гладкою (можуть бути точки розриву, тобто дискретні); може бути кілька локальних оптималів; завдання не потребує пошуку надзвичайно точного глобального оптимуму. На відміну від класичного підходу до багатокритеріальної оптимізації, генетичні алгоритми належать до класу багатоточкових методів пошуку, а тому проблему оптимізації з їх допомогою можна вирішити навіть у випадку полімодальності природи цільової функції, тобто у випадку, коли цільова функція має кілька максимумів (або мінімумів).

На відміну від традиційних методів оптимізації, генетичні алгоритми обробляють не значення параметрів самої задачі, а їх закодовану форму; пошук рішень не з однієї точки, а з певної їх підмножини; використовувати лише цільову функцію, а не її похідні або іншу додаткову інформацію; застосовувати ймовірнісні, а не детерміновані правила відбору.

До переваг використання генетичних алгоритмів можна віднести:

- факультативні знання конкретних знань про проблему;

- концептуальна простота та прозорість реалізації;
- можливість паралелізації (одночасне використання декількох точок простору можливих рішень);
- простота кодування вхідної та вихідної інформації;
- можливість застосовувати до широкого кола завдань без внесення істотних змін у внутрішню структуру методу;
- здатність адаптувати параметри генетичного пошуку до особливостей проблеми;
- менша ймовірність потрапляння та петлі в локальному оптимумі, що досягається використанням підходу населення;
- можливість використання інших ефективних процедур пошуку в методі.

До недоліків генетичного алгоритму можна віднести:

- висока ітерація;
- сильна залежність ефективності генетичного алгоритму від його параметрів (розмір популяції, початкова точка пошуку, імовірнісна характеристика генетичних операторів тощо);
- епістаз - взаємодія генів (змінних), при якій на активність одного гена впливають зміни інших генів. Існування цього явища призводить до появи низько адаптованого потомства. Рішення проблеми епістазу полягає в зберіганні взаємозалежних генів у хромосомі, розміщенні їх близько один до одного. Групування залежних генів значно зменшує ймовірність того, що вони будуть знищені схрещуванням;
- передчасна конвергенція, яка в більшості випадків пов'язана з відсутністю різноманітності індивідів у популяції. Найпоширенішою причиною передчасної конвергенції є відсутність чисельності населення.

Алгоритми визначення асоціацій

Алгоритми визначення асоціацій знаходять правила про окремі предмети, що з'являються разом в одній транзакції, наприклад, в одній покупці.

Послідовність - це теж асоціація, але залежна від часу. Асоціація записується як $A > B$, де A називається передумовою, B - наслідком. Частота появи

кожного окремого предмета або групи предметів, визначається в такий спосіб - підраховується кількість появ цього предмета у всіх подіях (покупках) і ділиться на загальну кількість подій. Ця величина вимірюється у відсотках і зветься "поширеність". Низький рівень поширеності (менш одного тисячної відсотка) говорить про неістотність асоціації.

Для визначення важливості кожного отриманого асоціативного правила необхідно одержати величину, що зветься "довірчість А до Б" (взаємозв'язок А і Б). Ця величина показує, як часто з появою А з'являється Б і розраховується як відношення частоти появи (поширеності) А і Б разом до поширеності А. Тобто, якщо довірчість А до Б дорівнює 20%, то це означає, що при покупці товару А в кожному п'ятому випадку купують і товар Б. Якщо поширеність А не дорівнює поширеності Б, то і довірчість А до Б не дорівнює довірчості Б до А. Справді, покупка комп'ютера частіше веде до покупки дискет, чим покупка дискети до покупки комп'ютера.

Ще одною важливою характеристикою асоціації є потужність асоціації. Чим більше потужність, тим сильніший вплив, що поява А робить на появу Б. Потужність розраховується по формулі: (довірчість А до Б) / (поширеність Б).

Деякі алгоритми пошуку асоціацій спочатку сортують дані і тільки після цього визначають взаємозв'язок і поширеність. Єдиною розбіжністю таких алгоритмів є швидкість або ефективність перебування асоціацій. Це важливо, у зв'язку з величезною кількістю комбінацій, яку необхідно перебрати для знаходження більш значимих правил. Алгоритми пошуку асоціацій можуть створювати свої бази даних поширеності, довірчості і потужності, до яких можна звертатися при запиті. Наприклад: "Знайти всі асоціації, у яких для товару Х довірчість більш 50% і поширеність не менш 2,5%". При знаходженні послідовностей додається змінна часу, що дозволяє працювати 24 із серією подій для знаходження послідовних асоціацій протягом деякого періоду часу.

Підводячи підсумки цьому методу аналізу, необхідно сказати, що може виникнути така ситуація, коли товари в супермаркеті будуть згруповані за допомогою знайдених моделей, але це, замість очікуваного прибутку, дасть

зворотний ефект. Це може відбутися через те, що клієнт не буде довго ходити по магазину в пошуках бажаного товару, купуючи при цьому ще щось, що потрапляється на очі, і те, що він ніколи не планував купувати.

1.4 Аналіз програм-аналогів

Сьогодні існує безліч професійних інструментів для покупок з великою кількістю персоналізації.

Для порівняльного аналізу характеристик аналогових програм дивіться наступні програми, які представлені в таблиці 1.1.

Таблиця 1.1 – Представлення порівняльних характеристик існуючих програм-аналогів

Програмна система	Характеристика системи
«Купи батон!»	Дозволяє скласти докладні списки покупок, що призводить до безпроблемних і швидких покупок для придбання продуктів харчування та інших товарів.
Remember the Milk	Перевірений часом ретроградний планувальник, який не хоче нічого спільного з 3D-інтерфейсами, голосовими командами та іншими перевагами майбутнього.
Clear	Гарний та функціональний планувальник, ідеально підходить не для особистих потреб, а для важкої офісної роботи з дуже швидко мінливими завданнями.
За покупками: Listick	Дозволяє створювати необмежені списки та шаблони. У програмі вже є вбудована база продуктів і, ввівши перші літери, вона запропонує варіанти імен, додасть їх до списку та автоматично визначить категорію.
Купи мене	Список покупок легко скласти та редагувати. У програмі вже є власна база даних для всіх існуючих продуктів, і коли ви додаєте елементи, вона автоматично замінює правильні слова та завантажує вже доступні зображення продуктів, які допомагають ідентифікувати предмет з першого погляду.

"Купи батон!" Це зручний та елегантний список покупок, розроблений для розумної організації та спрощення ваших покупок.

Ви можете створити стільки списків покупок, скільки вам потрібно купити зараз! Наприклад, ви можете скласти окремий перелік товарів, окремий список канцтоварів, інший список побутової хімії тощо.

Кількість елементів у кожному списку необмежена. Додавання нових продуктів (це взагалі речі / предмети) дуже легко. Як тільки ви почнете писати назву, додаток Buy Batman! Це відразу запропонує вам кілька варіантів, тому вам не доведеться вводити повне ім'я повністю і не витратити час. Однак трапляється, що база даних ще не має того, що вам потрібно - тоді просто введіть ім'я один раз, оскільки програма запам'ятає це назавжди. Для всіх продуктів ви можете вказати необхідну кількість. Які заходи ви вирішите - літри, шматки, пляшки, упаковка тощо. - що завгодно. Ви також можете сортувати список за категоріями: кожен товар можна переглянути у певному кольорі.

Переваги:

- перелік покупок може бути сформований за допомогою SMS;
- різні групи продуктів поділяються за кольором.

Недоліки:

- ви не можете видалити весь список продуктів одним пальцем;
- щоб створити кілька списків покупок, вам доведеться заплатити 66 рублів за повну версію;
- немає можливості додавати продукти до списку шляхом сканування штрих-коду [29].



Рисунок 1.3 – Приклад роботи програми

«Remember the Milk» - це послуга зі списку справ.

Дозволяє легко створювати, сортувати завдання, надсилати та отримувати завдання від співробітників. Він інтегрується з Google Calendar і Gmail, вам працювати в автономному режимі з Google Gears. Підтримує Email, IM, СМС.

Онлайнний органайзер Remember The Milk отримав підтримку від офлайн-платформи Google Gears. Практично всі функції тепер доступні в режимі офлайн: ви можете переглядати списки, додавати нові завдання та нотатки, редагувати наявні, використовувати власний пошук, створювати нові "розумні списки". Все в режимі офлайн синхронізується, як тільки доступ до Інтернету буде доступний. Ви можете легко переходити між режимами онлайн та офлайн: Remember The Milk автоматично визначить відсутність підключення.



Рисунок 1.4 – Приклад роботи програми

Переваги

- кожному завданню може бути присвоєний рівень важливості, простору, простору тощо;
- Не найсучасніший, але досить функціональний дизайн.

Недоліки

- рахунок, який суттєво не розширює можливості для застосування, коштуватиме 350 грн на місяць;
- Ви можете синхронізувати безкоштовно лише один раз на день [30].

Clear - це невеликий, але універсальний менеджер із завдань. Це відрізняється від багатьох подібних додатків у вашому контролі: вам не потрібно натискати кнопки, складати палі чи перетягувати повзунки. Усі дії в цій програмі виконуються за допомогою регулярних жестів, що стосуються кількох дотиків.

Інтерфейс Clear - це простий екран з плитками з іменами створення завдань. За допомогою щіпки ви можете зменшити або розширити конкретну задачу і помістити нове завдання в середину двох завдань. Проведення плиткою вліво видалить завдання, ілюструючи відповідну анімацію.

Проведіть праворуч - виділяє виконане завдання і автоматично видаляє його зі списку всіх запланованих завдань. Ось функція перетягування, яка вже

відома всім користувачам, що дозволяє створювати нові завдання або видаляти всі існуючі завдання одразу.

Переваги:

- завдання, залежно від їх важливості, виділяються різними відтінками червоного;
- кожне завдання можна перемістити до будь-якої частини списку, щоб змінити його пріоритет;
- завдяки синхронізації iCloud завдання будуть доступні для всіх ваших пристроїв;
- мінімалістичний дизайн без захащення;
- Вам навіть не потрібно підписуватися, щоб розпочати роботу.

Недоліки:

- У додатку взагалі немає налаштувань, тому вам доведеться працювати з кольорами та шрифтом за замовчуванням [31].



Рисунок 1.5 – Приклад роботи програми

«За покупками: Listick». Окрім створення списку покупки через додаток, ви можете відстежувати свої витрати за датою та категорією продукту. Створені

списки синхронізуються в режимі реального часу в хмарі і доступні на всіх пристроях та на сайті програми.

Ви можете додати віджет на робочий стіл свого смартфона для швидкого доступу до програми. Крім того, у програмі є кілька типів груп, що виділяють найважливіші елементи, голосовий ввід, кілька типів дизайну, а також видаляють придбані товари за один раз, створює список вхідних SMS та надсилає готові списки електронною поштою та sms.

Переваги:

- можливе спілкування електронною поштою;
- Програма займає мало оперативної пам'яті.

Недоліки:

- незручні налаштування програми;
- не можна створити більше двох списків покупок;
- тільки для користувачів девайсів Apple [32].

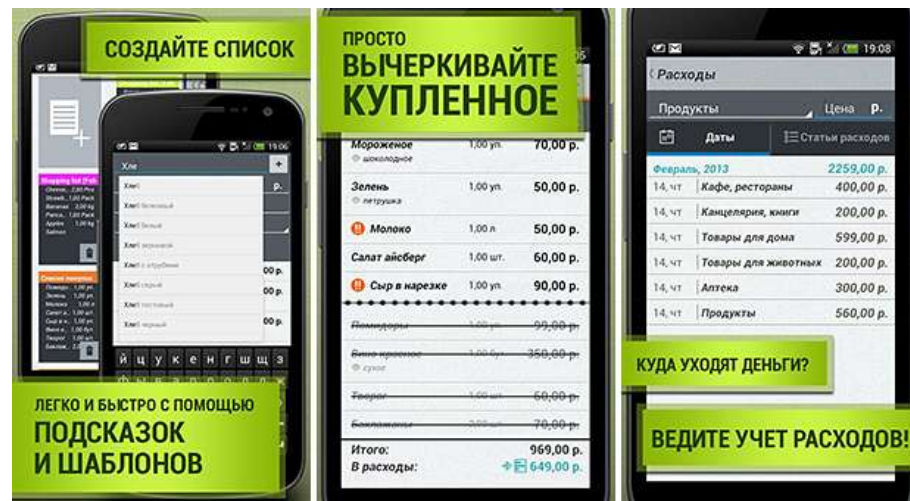


Рисунок 1.6 – Приклад роботи програми

«Купи меня» - це простий і зручний додаток, яке дозволяє легко і швидко створити список необхідних покупок. Особливо важливі покупки можуть бути виділені загальним списком тегів та перехресними перехресними покупками одним дотиком. Програма має вбудовану функцію для голосового набору та

зчитування штрих-коду (крім цього повинен бути встановлений сканер штрих-коду).

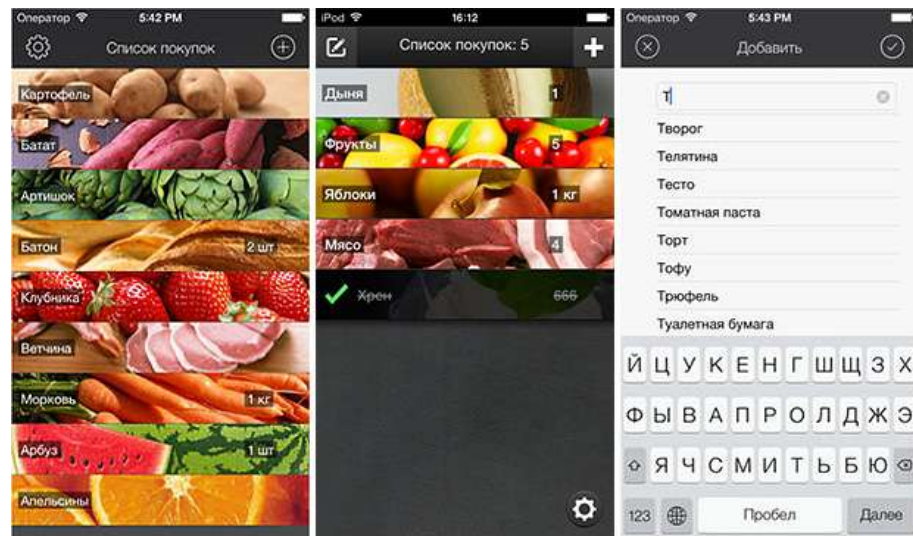


Рисунок 1.7 – Приклад роботи програми

Ви можете створювати окремі списки покупок для різних магазинів, переносити списки з іншими телефонами та створювати резервні копії. Додаток зовсім не заряджає телефон, не сповільнюється і швидко заряджається. У платній версії ви можете встановити віджет на робочий стіл.

Переваги:

- для використання програми не потрібно входити як користувач;
- зручний інтерфейс;

Недоліки:

- голосові команди не завжди розпізнаються правильно;
- ви не можете створити список текстових повідомлень [33].

Розглянувши функціональні можливості програм-аналогів таких як «Купи батон!», «Remember the Milk», «Clear», «За покупками: Listick», «Купи меня». можна зробити висновок, що ці програми мають деякі недоліки. Перш за все, більшість з них не підтримує сканування. Крім того, у багатьох розглянутих програмах налаштування дуже незручні для користувача. Також інтерфейс

більшості програм може бути важким у використанні без використання дуже яскравої палітри, що може спричинити втому очей.

Серед проаналізованих програмних систем «Купи батон!» найбільш близьким за характеристиками до розробленої системи покупок, яка полегшує процес вибору продуктів. Ось чому в майбутньому ми вирішимо прототипувати цю систему.

1.5 Висновок

Проаналізовано предметну область підбору покупок, що показало актуальність та доцільність розробки, яка полягає у потребі багатофункціонального інтелектуального інструменту для підбору покупок.

Розкрито поняття споживчого вибору та наведено його класифікаційні ознаки. Проаналізовано модель поведінки споживача товарів і визначено особливості його поведінки на ринку.

У результаті дослідження методів, що застосовують для вирішення подібних задач, зокрема метод дерева рішень, метод Байєса, алгоритм лінійної регресії, генетичні алгоритми, алгоритми визначення асоціацій, обрано методи Байєса та дерева рішень, що забезпечує високу швидкість роботи, підтримку поступового навчання та відносно просту програмну реалізацію алгоритму, що робить доцільним його використання як основа аналізу.

Проведено аналіз сучасних програм-аналогів, які використовуються для автоматизації процесу підбору покупок. Коротко охарактеризовано основні функції, які виконують дані програми. Вибрано аналог (мобільний додаток «Купи батон!»).

Аналіз існуючих систем показав, що на даний час подібні програми для підбору покупок не забезпечують потрібні швидкість та достовірність підбору покупок.

2 РОЗРОБКА МОДЕЛІ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПІДБОРУ ПОКУПОК

2.1 Дослідження можливостей методу підбору покупок

Моделювання - не єдиний метод вивчення зовнішнього світу [3]. Існує ціла область знань - методологія, яка спеціально стосується вивчення методів пізнання.

Моделювання - це метод пізнання зовнішнього світу, який можна віднести до загальнонаукових методів, які застосовуються як на емпіричному, так і на теоретичному рівнях пізнання. У побудові та вивченні моделі можуть бути використані майже всі інші методи пізнання [4].

Модель посилається на такий матеріальний або образно представлений об'єкт, який у процесі пізнання (дослідження) замінює об'єкт-оригінал, зберігаючи деякі важливі для цього дослідження типові його ознаки.

Для отримання аналітичних залежностей, що описують великі масиви даних, вони використовують методи наближення, засновані на тому, що масив замінюється простою функцією (лінійною, квадратичною, кубічною чи іншою), яка не обов'язково проходить через усі експериментальні точки, але описує зміни цих даних і надає мінімум квадратів експериментальних відхилень даних від цієї функції.

Апроксимація (від лат. *Approimo* - наближається) - заміна одних математичних об'єктів іншими, певним чином близькими до вихідних [5]. Апроксимація дозволяє досліджувати числові характеристики та якісні властивості об'єкта, зменшуючи завдання вивчення більш простих чи зручних об'єктів.

Є одна проста істина в прогнозуванні, що синоптики давно прийшли: висока точність наближення даних не гарантує високої точності прогнозів. Це означає, що модель, яка створює найменші помилки у навчальному зразку, не обов'язково дасть однакові найменші помилки в тестовій вибірці. Протилежне,

як правило, можливе лише для оборотних процесів, в яких якісних змін не відбувається. Але, як ми знаємо, у світі немає нічого більш постійного, ніж зміни.

Однак у цієї простої істини є ще одна правда: модель, яка навмисно погано наближається до набору даних, навряд чи дасть точний прогноз. Це, в свою чергу, означає, що для отримання більш точних прогнозів модель повинна точно описувати рядок шляхом внутрішнього вибірки, бути адекватною задачі, що виконується, і включати всі значущі змінні. Тому що, якщо модель - це дуже приблизний та умовний опис об'єкта, то він може зробити адекватні висновки щодо об'єкта дослідження.

Об'єктом дослідження у прикладній статистиці є статистика, отримана в результаті спостережень або експериментів [6]. Статистика - це сукупність об'єктів (спостережень, випадків) та ознак (змінних), що характеризують їх.

Наприклад, є завдання прогнозувати попит. Кожен товар повинен бути придбаний у певній кількості. Якщо ви купуєте більше, ніж потрібно, то, якщо товар швидко псується, він зіпсується, інакше він займе місце на складі, за що також потрібно заплатити. Якщо ви купуєте занадто мало товару, на сайті завжди буде показано, що він недоступний, і користувачі будуть ходити по магазинах з нашими конкурентами. Тобто необхідно передбачити кількість товарів, які будуть користуватися попитом. Тут виникає питання про раціональну закупівлю товарів.

2.2 Обґрунтування вибору методу найвнього байєсівського класифікатора для інформаційної технології підбору покупок

Наївний класифікатор Байєса інтегрує модель з правилом рішення. Одне загальне правило - вибір найбільш вірогідної гіпотези. Це правило відоме як апостеріорне правило прийняття рішень (MAP). Відповідний класифікатор - це класифікаційна функція, яка реалізується наступним чином:

$$\text{classify}(f_1, \dots, f_n) = \arg \max p(C = c) \prod_{i=1}^n p(F_i = f_i | C = c) \quad (2.1)$$

Імовірнісна модель для класифікатора є умовною ймовірністю:

$p(Y | F_1, F_2, \dots, F_n)$ над залежною змінною класу Y з невеликою кількістю результатів або класів, яка залежить від кількох змінних F_1, F_2, \dots, F_n . Проблема полягає в тому, що якщо кількість символів n велика або коли символи можуть приймати багато значень, то побудувати таку модель на таблицях ймовірностей неможливо, тому ми переформулюємо модель.

За допомогою теореми Баєса ми можемо записати [8]:

$$p(Y | F_1, F_2, \dots, F_n) = \frac{p(Y)p(F_1, F_2, \dots, F_n | Y)}{p(F_1, F_2, \dots, F_n)} \quad (2.2)$$

На практиці нас цікавить тільки чисельник цього дробу, оскільки знаменник не залежить від Y , так що знаменник є константою. Чисельник еквівалентний ймовірності спільної появи ознак F_1, F_2, \dots, F_n і класу Y - $p(Y, F_1, F_2, \dots, F_n)$, ця ймовірність може бути переписана наступним чином, використовуючи повторні застосування визначення умовної ймовірності:

$$\begin{aligned} p(Y, F_1, F_2, \dots, F_n) &= p(Y) p(F_1, F_2, \dots, F_n | Y) = \\ &= p(Y) p(F_1 | Y) p(F_2, \dots, F_n | Y, F_1) = \\ &= p(Y) p(F_1 | Y) p(F_2 | Y, F_1) p(F_3, \dots, F_n | Y, F_1, F_2) = \\ &= p(Y) p(F_1 | Y) p(F_2 | Y, F_1) \cdot \dots \cdot p(F_n | Y, F_1, F_2, F_3, \dots, F_{n-1}) \end{aligned} \quad (2.3)$$

Тепер "наївні" припущення про умовні обмеження вступають у гру: припустимо, що кожна ознака F_i умовно незалежна від будь-якої іншої ознаки F_j при $i \neq j$. Це означає, що

$$p(F_i | Y) = p(F_i | Y, F_j) \text{ при } i \neq j \quad (2.4)$$

і тому спільна модель може бути виражена як:

$$\begin{aligned} p(Y, F_1, F_2, \dots, F_n) &= p(Y) p(F_1 | Y) p(F_2 | Y) p(F_3 | Y) \cdot \dots \cdot p(F_n | Y) = \\ &= p(Y) \prod_{i=1}^n p(F_i | Y) \end{aligned} \quad (2.5)$$

Це означає, що за наведеними вище припущеннями про незалежність, умовний розподіл над змінною класу Y може бути виражений так:

$$p(Y | F_1, F_2, \dots, F_n) = \frac{1}{A} p(Y) \prod_{i=1}^n p(F_i | Y) \quad (2.6)$$

де A - коефіцієнт масштабування, що залежить тільки від ознак F_1, F_n , тобто, константа, якщо значення змінних ознак відомі.

Моделі цієї форми більш керовані, оскільки вони враховують так званий клас апріорних $p(Y)$ і незалежних розподілів ймовірностей $p(F_i | Y)$. Якщо є k класів і якщо модель для кожного з них $p(F_i | Y = y)$ може бути виражена через r параметрів, то відповідна наївна модель Баєса має $(k - 1) + n r$ параметрів. На практиці часто $k = 2$ (двійкова класифікація) і $r = 1$ є загальними, і тому загальна кількість параметрів наївної моделі Баєса є $2n+1$, де n - число двійкових ознак, що використовуються для класифікації і прогнозування.

Всі параметри моделі (тобто попередні класи та розподіл ймовірностей функцій) можна наблизити до відносних частот із навчального набору даних. Це оцінки максимальної ймовірності ймовірностей. Априорні ймовірності класів можна обчислити, припустивши, що класи рівні: $p = \frac{1}{k}$, де k - кількість класів; або шляхом обчислення оцінки ймовірності класу з навчального набору p_i (для даного класу) $= \frac{m_i}{m}$, де m_i - кількість зразків у класі i ; m - загальна кількість зразків.

Щоб оцінити параметри розподілу атрибутів F_i , слід зробити припущення про розподіл або створити непараметричні моделі для ознак з навчального набору даних. При роботі з безперервними даними типовим припущенням є те,

що безперервні значення, пов'язані з кожним класом, розподіляються відповідно до розподілу Гаусса.

Наприклад, припустимо, що навчальні дані містять неперервну ознаку x . Спочатку сегментуємо дані за класом, а потім обчислюємо середнє і дисперсію x для кожного класу. Нехай μ_y буде середнім значенням x , яке пов'язане з класом y , і нехай σ_y^2 буде дисперсією значень x , пов'язана з класом y . Тоді ймовірність деякого значення, заданого класом, $P(x = v | y)$, може бути обчислена шляхом включення v до рівняння нормального розподілу, з параметрами μ_y і σ_y^2 .

$$P(x = v | y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{(v-\mu_y)^2}{2\sigma_y^2}} \quad (2.7)$$

Ще один поширений спосіб обробки безперервних значень - використання квантування для вибіркового значень. Загалом, метод розподілу є найкращим вибором, якщо є невеликий обсяг даних про навчання або якщо відомий точний розподіл даних.

Серед переваг:

- Можливість вчитися на невеликих тренувальних наборах, не втрачаючи точності.
- Інкрементальна здатність до навчання. Інкрементальне навчання має на увазі наступне: Кожен новий документ, який отримує класифікатор, може оновити свою навчальну базу. Ця особливість пов'язана з тим, що підготовка байєсівського класифікатора є ітеративною.
 - Висока швидкість роботи.
 - Масштабована здатність.
 - Просте тлумачення моделей.
 - Простота впровадження програмного забезпечення.
 - Відносно низькі вимоги ОЗУ.
 - Класифікація як двійкової, так і багатокласової.

- Можливість багатозадачності класифікації.

2.2 Обґрунтування вибору методу дерева рішень для інформаційної технології підбору покупок

Дерево рішень - це ієрархічна структура дерева, яка використовується для класифікації класів на основі ряду питань (або правил) щодо атрибутів (атрибутів) класу. Атрибути класу можуть бути будь-яким типом змінної (двійкових, номінальних, порядкових та кількісних значень, тоді як класи повинні бути якісного типу (категоріальні чи двійкові чи порядкові). Коротше кажучи, з огляду на атрибути навчальних даних разом із класами, дерево рішень створює послідовність правил (або серії запитань), які можна використовувати для ідентифікації класу.

Для випадкової величини x з k дискретними значеннями, розподілених за $P = (p_1, p_2, \dots, p_k)$, домішкою є функція $\phi: [0, 1]^k \rightarrow \mathbb{R}$, яка задовольняє наступні умови:

$$\phi(P) \geq 0$$

$\phi(P)$ мінімальна, якщо $\exists i$ такий, що компонент $p_i = 1$.

$\Phi(P)$ максимальна, якщо $\forall i, 1 \leq i \leq k, p_i = \frac{1}{k}$.

$\Phi(P)$ є гладкою (диференційована всюди) в своєму діапазоні.

Зауважте, що якщо вектор ймовірності має компонент 1 (змінна x отримує лише одне значення), то змінна однозначно визначається (чисто). З іншого боку, якщо всі компоненти рівні, рівень домішок досягає максимуму.

Процедура розподілу вибирає найкращу змінну з простору знаків x та найкращий поріг для отримання найбільшого зменшення домішки цільової змінної y :

$$\Delta\phi(\tau) = \phi(\tau) - \sum_{b=1}^B p(\tau_b|\tau) \phi(\tau) \quad (2.8)$$

де τ_b позначає b-й дочірній вузол;

$p(\tau_b|\tau)$ – частка спостережень у вузлі τ , що стосуються вузла τ_b ;

B – кількість гілок після поділу вузла τ

Критерії зменшення домішок, доступні для дерев класифікації, ґрунтуються на різних домішках $\phi(\tau)$.

Інформаційний виграш [13]:

$$I(A_i) = \sum_{j=1}^{M_i} p(A_{ij}) H(C|A_{ij}) \quad (2.9)$$

де A_i – атрибут, який використовується для розбиття вузла дерева;

M_i – число різних значень для A_i ;

$p(A_{ij})$ – ймовірність, що атрибут A_i приймає j-те значення;

$H(C | A_{ij})$ – класова ентропія множини випадків, коли A_i приймає j-те значення.

Ця ентропія класифікації визначається як:

$$H(C|A_{ij}) = - \sum_{k=1}^J p(C_k|A_{ij}) \log_2 p(C_k|A_{ij}) \quad (2.10)$$

де J – число класів у нашій задачі класифікації;

$p(C_k | A_{ij})$ – ймовірність того, що приклад належить до класу C_k , враховуючи, що його атрибут A_i приймає j-те значення

Індекс Джині намагається звести до мінімуму домішки, що містяться в підмножинах навчання, створених після розбиття вузла дерева рішень. Він використовує наступну функцію:

$$G(A_i) = \sum_{j=1}^{M_i} p(A_{ij}) G(C|A_{ij})$$

$$G(C|A_{ij}) = - \sum_{k=1}^J p(C_k|A_{ij}) p(\neg C_k|A_{ij}) = 1 - \sum_{k=1}^J p^2(C_k|A_{ij}) \quad (2.11)$$

де $p(\neg C_k | A_{ij}) = 1 - p(C_k | A_{ij})$.

Як ми можемо побачити з наведених вище виразів, ці формули досить схожі на ті, які ми використовували для визначення інформаційного виграшу: ми просто замінили $p(-C_k | A_{ij})$ на $-\log_2 p(C_k | A_{ij})$. Насправді вони є як альтернативними домішками, так і мають загальні властивості. Однак індекс Джині надає перевагу розділам, які розміщують найбільший клас в одному мережевому вузлі, а всі інші - в іншому, а критерій ентропії, наприклад, робить акцент на врівноваженні розмірів двох дочірніх вузлів.

Критерії поділу на основі домішок (наприклад, максимізація приросту інформації) мінімізують ентропію. Імовірність, що відповідає найпоширенішому класу, - це вираз, який найменше збільшує загальну ентропію. Таким чином, ця ймовірність є потенційним критерієм максимізації для вимірювання якості даного розділу. Тому ми схильні використовувати таку функцію:

$$K(A_i) = \sum_{j=1}^{M_i} p(A_{ij}) K(C|A_{ij})$$

$$K(C|A_{ij}) = \max_k p(C_k | A_{ij}) \quad (2.12)$$

де A_i – атрибут, який використовується для розбиття вузла дерева;

M_i – число різних значень для A_i ;

$p(A_{ij})$ – ймовірність, що атрибут A_i приймає j -те значення;

$K(C | A_{ij})$ відповідає оцінці кожного дочірнього вузла.

Ця функція оцінювання - це просто максимальна ймовірність класу $p(C_k | A_{ij})$ у дочірньому вузлі (тобто максимальна ймовірність класу в навчальному наборі даних, враховуючи, що атрибут A_i приймає j -е значення).

Критерії зупинки

Фаза росту дерева триває, поки не працює критерій зупинки. Наступні умови є загальними правилами зупинки:

1. Усі екземпляри навчального набору належать до одного значення.
2. Досягнута максимальна глибина дерева.

3. Кількість прикладів у кінцевому вузлі менше мінімальної кількості прикладів для батьківських вузлів.

4. Якби вузол був розділений, кількість прикладів в одному або декількох дочірніх вузлах була б меншою, ніж мінімальна кількість прикладів для дочірніх вузлів.

2.4 Розробка математичної моделі підбору покупок

Абстрактно імовірнісна модель для класифікатора — це умовна модель (формула 2.13).

$$p(C | F_1, \dots, F_n) \quad (2.13)$$

над залежною змінною класу C з невеликою кількістю результатів або класів, залежна від кількох змінних F_1, \dots, F_n [7]. Проблема полягає в тому, що коли кількість властивостей n дуже велика або коли властивість може приймати багато значень, тоді неможливо побудувати таку модель на таблицях ймовірностей. Тому ми переформулюємо модель, щоб зробити її легшою в обробці.

Використовуючи теорему Баєса, запишемо формулу 2.14.

$$p(C | F_1, \dots, F_n) = \frac{p(C)p(F_1, \dots, F_n|C)}{p(F_1, \dots, F_n)} \quad (2.14)$$

На практиці актуальний лише чисельник цього дроби, так як знаменник не залежить від C і значення властивостей даних, так що знаменник - константа [8].

Чисельник еквівалентний спільній ймовірності моделі $p(C | F_1, \dots, F_n)$, яка може бути переписана таким чином, використовуючи повторні додатки визначень умовної ймовірності (формула 2.15).

$$p(C, F_1, \dots, F_n) = p(C)p(F_1, \dots, F_n|C) = p(C)p(F_1|C)p(F_2, \dots, F_n|C, F_1) =$$

$$\begin{aligned}
&= p(C)p(F_1|C)p(F_2|C, F_1)p(F_3, \dots, F_n|C, F_1, F_2) = \\
&= p(C)p(F_1|C)p(F_2|C, F_1) * \dots * p(F_n|C, F_1, F_2, F_3, \dots, F_{n-1}) \quad (2.15)
\end{aligned}$$

Тепер можна використовувати «наївні» припущення умовної незалежності: припустимо, що кожна властивість умовно незалежна від будь-якої іншої властивості F_j при $j \neq i$.

Це означає (формула 2.16):

$$p(F_i|C, F_j) = p(F_j|C) \quad (2.16)$$

Таким чином, спільна модель може бути виражена як формула 2.17.

$$\begin{aligned}
p(C | F_1, \dots, F_n) &= p(C) = p(F_1|C)p(F_2|C)p(F_3|C) * \dots * p(F_n|C) = \\
&= p(C) \sum_{i=1}^n p(F_i|C) \quad (2.17)
\end{aligned}$$

Це означає, що з припущення про незалежність, умовний розподіл за класової змінної C може бути виражене формулою 2.18.

$$p(C | F_1, \dots, F_n) = \frac{1}{Z} p(C) \sum_{i=2}^n p(F_i|C) \quad (2.18)$$

де Z — це масштабний множник, що залежить тільки від F_1, \dots, F_n , тобто константа, якщо значення змінних відомі.

Оцінка параметрів

Всі параметри моделі можна наблизити за відносними частотами навчального набору даних [9]. Це оцінки максимальної ймовірності ймовірностей. Постійні властивості зазвичай оцінюються за нормальним розподілом. В якості математичного очікування та дисперсії обчислюються статистичні дані - середнє та стандартне відхилення відповідно.

Якщо задані значення класу та властивостей ніколи не зустрічаються разом у навчальному наборі, то оцінка, заснована на ймовірності, буде нульовою. Це проблема, оскільки множення нульової оцінки призведе до втрати інформації про інші ймовірності. Тому доцільно вносити невеликі корективи у всі оцінки ймовірностей, щоб жодна ймовірність не була рівно нульовою.

Побудова класифікатора за ймовірнісної моделі

Наївний Баєсівський класифікатор об'єднує модель з правилом рішення. Одне загальне правило має обрати найбільш ймовірну гіпотезу; воно відоме як апостеріорне правило прийняття рішення (MAP). Відповідний класифікатор — це функція *classify*, визначена за формулою 2.19.

$$\text{classify}(f_1, \dots, f_n) = \operatorname{argmax} p(C = c) \sum_{i=1}^n p(F_i = f_i | C = c) \quad (2.19)$$

Після аналізу цього методу був розроблений алгоритм, який є основою для подальшого розвитку. Основний алгоритм програмного модуля включає весь процес вибору покупок за допомогою наївного байєсівського класифікатора.

Введення дерева прийняття рішення

Дерево рішень використовується в статистиці та аналітиці даних для прогнозних моделей. Структура дерева містить такі елементи: "листя" та "гілки". На краях ("гілках") дерева рішень записуються атрибути, від яких залежить цільова функція, в "букві" записуються значення цільової функції, а в інших вузлах - атрибути, за якими випадки відрізняються. Щоб класифікувати новий випадок, потрібно спуститись по дереву до листа і надати відповідне значення.

Подібні дерева рішень широко використовуються при обробці даних для вирішення завдань покупок.

Метою даного методу є створити модель, яка прогнозує значення цільової змінної на основі декількох вхідних змінних. Кожна літера представляє значення цільової змінної, оскільки вона рухається від кореня до листа. Кожному внутрішньому вузлу відповідає одна з вхідних змінних. Дерево також можна

"дослідити", розділивши оригінальні набори змінних на підмножини на основі значень атрибутів тестування. Це процес, який повторюється для кожної з отриманих підмножин [42].

Рекурсія завершується тоді, коли підмножина в вузлі має ті ж значення цільової змінної, таким чином, вона не додає цінності для прогнозування. Процес, що йде «згори донизу», індукція дерев рішень (TDIDT), є прикладом поглинаючого «жадібного» алгоритму, і на сьогодні є найбільш поширеною стратегією дерев рішень для даних, але це не єдина можлива стратегія. В інтелектуальному аналізі даних, для розв'язання задачі підбору покупок, дерева рішень можуть бути використані як математичні та обчислювальні методи, щоб допомогти описати, класифікувати і узагальнити набір даних, які можуть бути записані таким чином (формула 2.20):

$$(x, Y) = (x_1, x_2, x_3 \dots x_k, Y) \quad (2.20)$$

Залежна змінна Y є цільовою змінною, яку необхідно проаналізувати, класифікувати й узагальнити. Вектор x складається з вхідних змінних x_1, x_2, x_3 тощо, які використовуються для виконання цього завдання [43].

Отже, при поєднанні методу дерева рішень та використанні найвного байєсівського класифікатора було розроблено модель майбутньої розумної системи закупівель. Впровадивши метод максимальної ймовірності найвного байєсівського класифікатора, можна буде досягти найбільш точного відбору товарів конкретного споживача.

Тому при поєднанні алгоритму дерева рішень з найвним методом класифікатора Баєса було визначено модель майбутньої інформаційної технології закупівлі. Впровадивши метод максимальної ймовірності найвного байєсівського класифікатора, можна буде досягти найбільш точного відбору товарів конкретного споживача.

2.5 Проектування структури інформаційної технології підбору покупок

Система матиме модуль інтерфейсу, модуль авторизації, модуль бази даних, модуль вибору покупок та модуль списку покупок.

Спочатку модуль інтерфейсу асоціюється з модулем авторизації та модулем списку покупок. Він несе відповідальність за правильне відображення списків покупок, а також за внесення та відображення відповідної інформації.

Модуль авторизації може викликати модуль інтерфейсу за умови, що користувач бажає видалити певний список покупок із бази даних. Цей модуль відповідає за авторизацію користувача; логін та пароль потрібні для правильної авторизації. Він пов'язаний з модулем інтерфейсу та модулем списку покупок.

Модуль списку покупок пов'язаний з інтерфейсними модулями, авторизацією, керуванням базами даних.

Модуль вибору придбання пов'язаний з модулем бази даних. Цей модуль є основним модулем програми і містить інтелектуальний алгоритм вибору покупок. Виділення відбувається тоді, коли методи наївного класифікатора Байєса та дерева рішень послідовно застосовуються.

Модуль бази даних пов'язаний з торговим модулем і модулем списку покупок. Цей модуль посилається на базу даних, що містить уже створені списки покупок, а також набір продуктів та посилання між ними. Цей модуль містить запити в SQL, які дозволяють переглядати існуючі списки, видаляти їх або додавати нові.

Розроблена загальна структурна схема функціонування системи підбору покупок зображена на рисунку 2.1. Структура інформаційної технології підбору покупок зображена на рисунку 2.2. Розроблена схема загального алгоритму функціонування системи зображена на рисунку 2.3.

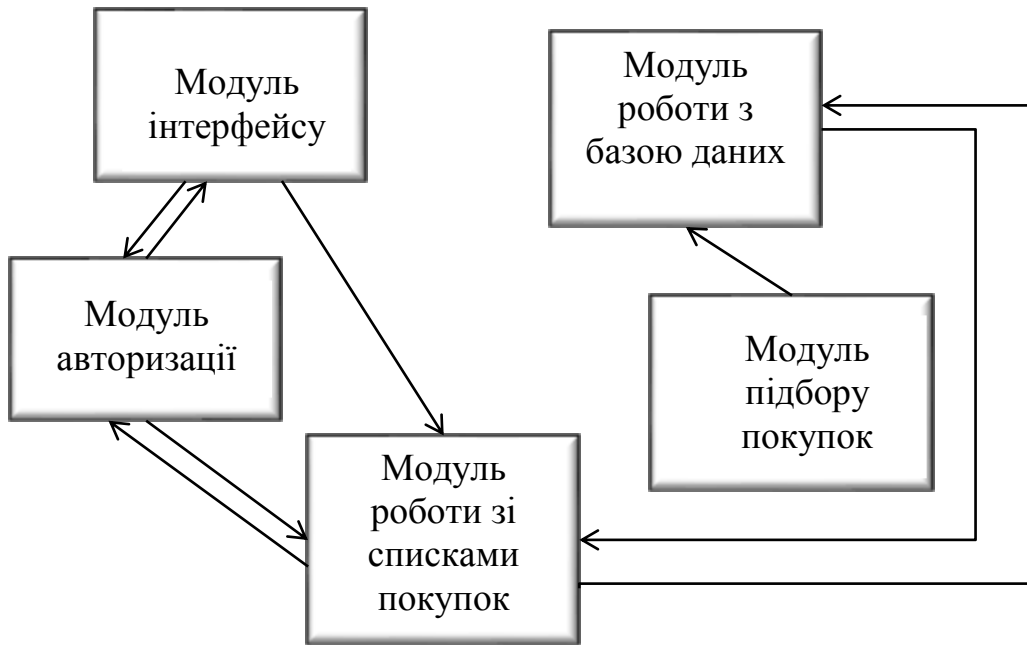


Рисунок 2.1 – Загальна структурна схема системи підбору покупок

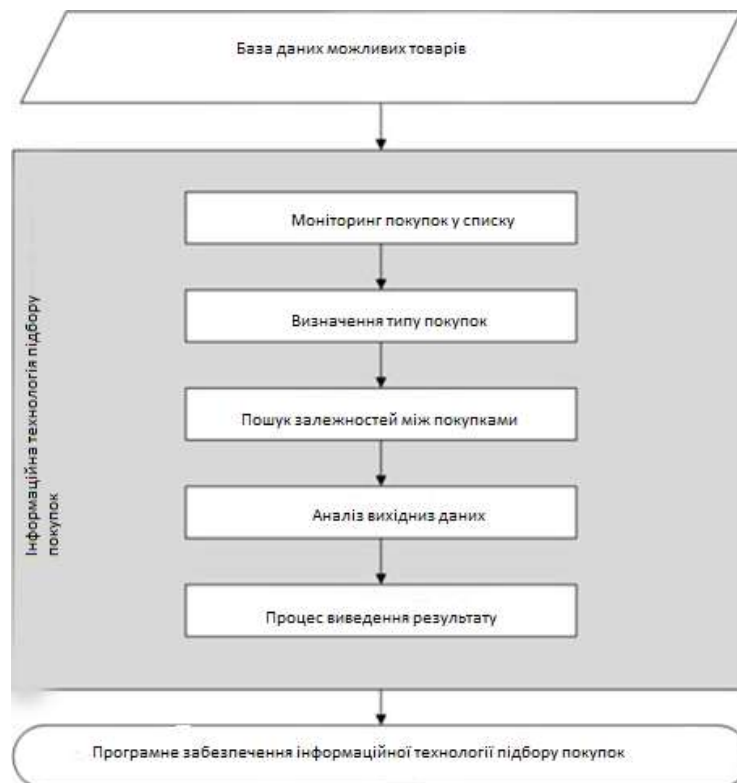


Рисунок 2.2 – Структура інформаційної технології підбору покупок

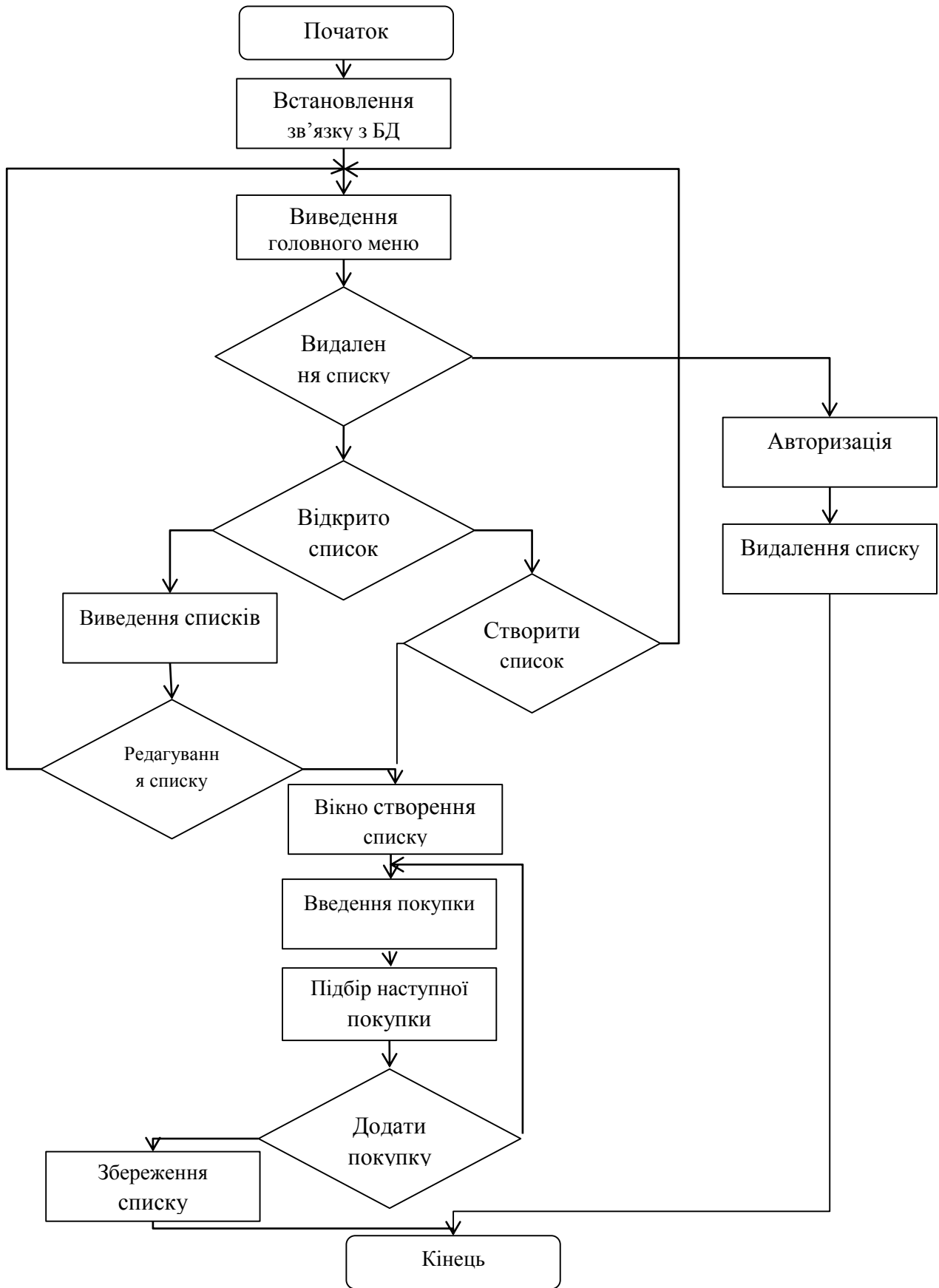


Рисунок 2.3 - Схема загального алгоритму функціонування системи

2.6 Розробка ER-моделі

Перелік сутностей, що необхідні для опису предметної області: користувач, продукти, список покупок.

За правилами побудови концептуальних схем предметної області у вигляді ER-структур сутності зображують позначеними прямокутниками, асоціації – ромбами, а зв'язки між ними – ненаправленими ребрами, над якими може проставлятися ступінь зв'язку і необхідне пояснення.

Представлення ER-моделі для бази даних зображено на рисунку 2.4.

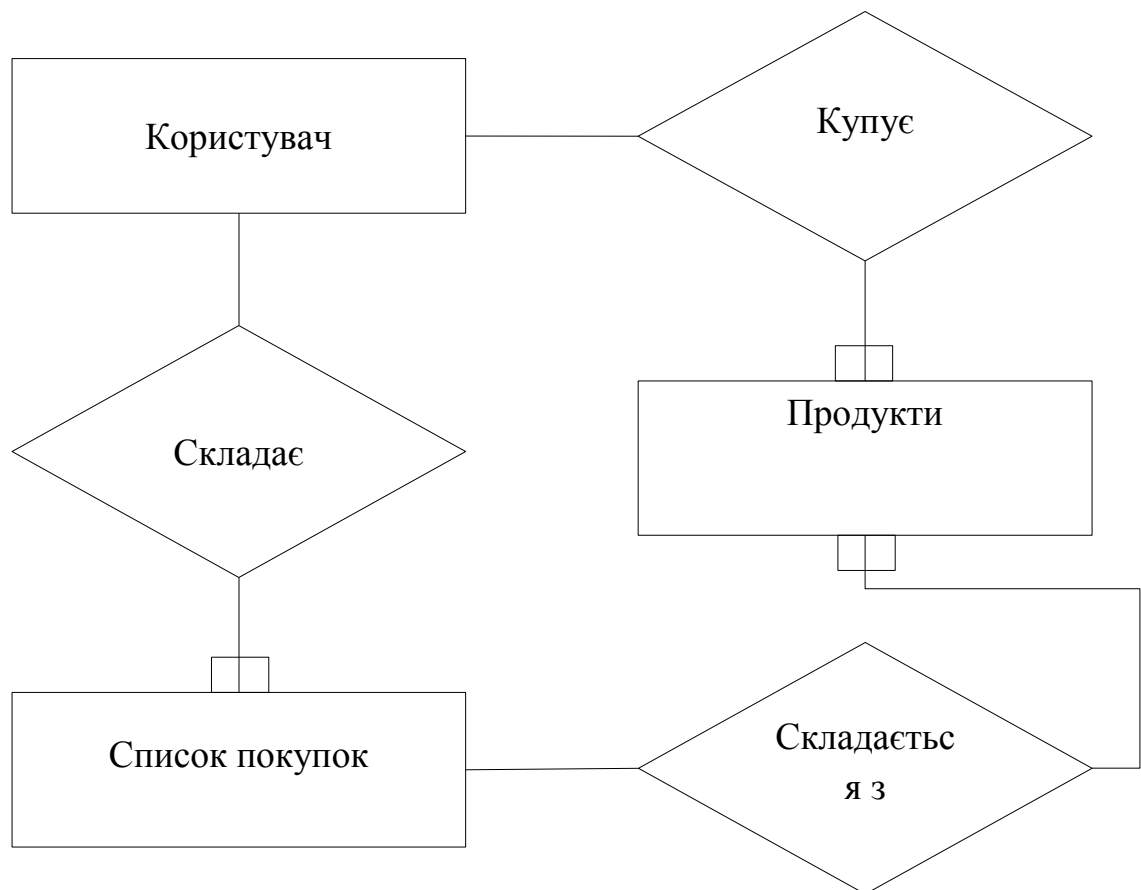


Рисунок 2.4 – ER-модель предметної області

Характеристики зв'язків предметної області представлено у таблиці 2.1.

Таблиця 2.1 - Характеристики зв'язків предметної області

Ім'я сутності 1	Ім'я сутності 2	Тип зв'язку	Ім'я зв'язку	Клас належності
Користувач	Продукти	1:Б	Купує	Обов'язковий
Користувач	Список покупок	1:Б	Складає	Обов'язковий
Список покупок	Продукти	Б:Б	Складається з	Обов'язковий

2.7 Висновок

Аналіз показав, що для вирішення проблеми підбору покупок підходить, що базуються на ймовірнісних методах, зокрема, метод Байєса та дерева рішень повністю задовольняють нашим вимогам.

Обґрунтовано вибір метода дерева рішень в комбінації зі спрощеним алгоритмом Баєса, оскільки це забезпечує можливість перегляду значення цільової функції з додаванням нових параметрів, та зміною уже існуючих параметрів, в наслідок чого запропоновано математичну модель для даної інформаційної технології.

Запропоновано структуру інформаційної технології підбору покупок, яка відрізняється поєднанням алгоритмів Баєса та дерева рішень, що дозволить підвищити швидкість підбору покупок. Обґрунтовано вибір методу дерева рішень та методу Байєса для підбору покупок.

3 ПРОГРАМНА РЕАЛІЗАЦІЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПІДБОРУ ПОКУПОК ТА ЇЇ ПРОГРАМНА РЕАЛІЗАЦІЯ

3.1 Обґрунтування вибору мови програмування

Java (вимовляється Джава; інколи — Ява) — об'єктно-орієнтована мова програмування, випущена компанією Sun Microsystems у 1995 році як основний компонент платформи Java. Зараз мовою займається компанія Oracle, яка придбала Sun Microsystems у 2009 році. Синтаксис мови багато в чому схожий на C та C++. У офіційній реалізації Java-програми компілюються у байт-код, який при виконанні інтерпретується віртуальною машиною для конкретної платформи.

Oracle надає компілятор Java та віртуальну машину Java, які задовольняють специфікації Java Community Process, під ліцензією GNU General Public License.

Мова запозичила значну частину синтаксису C і C++. Зокрема, об'єктна модель C++ була взята, але модифікована. Деякі конфлікти, які можуть виникнути через помилки розробника, були усунені та полегшено процес розробки об'єктно-орієнтованих програм. Ряд дій, які повинні виконати програмісти C / C++, призначені віртуальній машині. Java розроблена головним чином як незалежна від платформи мова, тому має менше апаратного забезпечення. При необхідності Java дозволяє викликати підпрограми, написані іншими мовами програмування [44].

Java вплинула на розробку J++, розроблену Microsoft. Робота над J++ була припинена через причину Sun Microsystems, оскільки ця мова програмування була модифікацією Java. Пізніше нова платформа Microsoft.NET випустила J# для полегшення міграції програмістів J++ або Java на нову платформу. З часом нова мова програмування C# стала основною мовою платформи, перейнявши більшу частину Java. J# востаннє включався в Microsoft Visual Studio 2005. Мова JavaScript має схоже на Java ім'я та синтаксис, але не пов'язана з Java. [45].

Таблиця 3.1 – Порівняння мов програмування C++, C#, Java, PYTHON

Можливість	C++	C#	JAVA	PYTHON
Об'єктно-орієнтована	+	+	+	+
Статистична типізація	+	+	+	-
Динамічна типізація	-	+	-	+
Явна типізація	+	+	+	-
Неявна типізація	-	-	-	+
Можливість компіляції	+	+	+	+
Багатопотокова компіляція	+	-	+	-
Ручне управління пам'яттю	+	+	-	-
Збір сміття	-	+	+	+
Інструкція goto	+	+	+	-
Підтримка try/catch	-	+	+	+
Інтерфейси	+	+	+	+
Перегрузка функції	+	+	+	-
Безкоштовність	-	-	+	-
Сумісність різних версій	-	-	+	-

Враховуючи всі переваги java, а також той факт, що вона є у вільному доступі та має велику кількість безкоштовних бібліотек, її обрали саме за її програмне забезпечення.

Основні оператори мови програмування java

Основними операторами мови програмування, що використовується при розробці системи, є:

- оператор циклу for;
- оператор циклу while;
- оператор вибору if;

Цикли використовують для багаторазового виконання одних і тих же команд, які знаходяться в тілі циклу.

Цикл for, цикл з чітко визначеною кількістю ітерацій. З переходом від ітерації до ітерації, змінюється змінна, яка є параметром циклу (на один менше, або на один більше).

Цикл while, цикл який виконується поки не порушиться умова, яка задана в заголовку циклу, такий цикл не має визначеної кількості ітерацій.

Оператор `if` дозволяє нам виконати чи не виконати певний розділ коду, залежно від того, чи виконується умова.

Ключове слово використовується для забезпечення механізму успадкування `extends`.

Для забезпечення механізму реалізації інтерфейсу використовується ключове слово `implements` [46].

Особливості середовища Eclipse

Eclipse (вимовляється «i-clips», англійською мовою «Eclipse») - це вільне модульне інтегроване середовище розробки програмного забезпечення. Фонд Eclipse був розроблений і підтримується. В основному він написаний на Java і може використовуватися для розробки програм Java.

Eclipse - основа для розробки модульних кросплатформних додатків з низкою особливостей:

- Можливість розробки програмного забезпечення для багатьох мов програмування (рідна Java);
- крос-платформа;
- модульний, призначений для подальшого розширення сторонніми розробниками;
- відкритий код;
- Розроблений та підтримується Eclipse Foundation, до якого входять постачальники програмного забезпечення, такі як IBM, Oracle, Borland.

Проект спочатку був розроблений у IBM як корпоративний стандарт IDE, встановлений багатьма мовами для платформ IBM. Потім проект був перейменований у Eclipse та був наданий громаді для подальшого розвитку [47].

Eclipse - це, насамперед, повноцінна Java IDE, орієнтована на розробку груп інструментаріїв для систем управління версіями (підтримка CVS включена в поставку Eclipse, активно розробляються декілька модулів SVN, доступна підтримка VSS тощо). Eclipse - це корпоративний стандарт розробки програмного забезпечення Java у багатьох організаціях.

Друге призначення Eclipse — служити платформою для нових розширень. Такими стали C/C++ Development Tools (CDT), розроблювані інженерами QNX разом із IBM, засоби для підтримки інших мов різних розробників. Безліч розширень доповнює Eclipse менеджерами для роботи з базами даних, серверами застосунків та інших [48].

Eclipse написана на Java, тому є платформи-незалежним продуктом, крім бібліотеки графічного інтерфейсу SWT, яка розробляється окремо для більшості поширених платформ. Бібліотека SWT використовує інструменти графічної платформи (OS) для забезпечення швидкого та персоналізованого вигляду інтерфейсу користувача.

За даними IDC Eclipse, 2,3 мільйона розробників працюють. Тому для цього проекту була обрана мова програмування Eclipse PPP та Java.

3.2 Програмна реалізація інформаційної технології підбору покупок

На рисунку 3.2 зображено загальну UML-діаграму класів системи підбору покупок.

Були розроблені класи Main, MainGUI, Pass, QVis, DellList TableModalQ, QutList, Create.

Клас Main - первинний клас, відповідальний за запуск програми та підключення до бази даних. Що робити з бібліотекою JDBC. Цей клас зберігає змінну Connection, необхідну для належного функціонування запитів.

Клас MainGUI – відповідає за показ головного меню. Цей клас успадковується від бібліотечного класу JFrame. Клас має методи stvot (), який відповідає за відображення вікна. Існує також метод shov (), який відповідає за приховування вікна після переходу до одного з основних пунктів меню. Клас має три об'єкти в класі JButton, кожен з яких відповідає за відображення та управління кнопками головного меню.

Клас Pass відповідальний за авторизацію користувача. Клас використовує об'єкт класу бібліотеки JPasswordField, який відповідає за безпеку пароля.

Клас DellList несе відповідальність за видалення всіх існуючих списків покупок та видалення одного з них. Цей клас почне свою роботу лише в тому випадку, якщо користувач увійшов у систему. Цей клас містить запит до бази даних, який виконується за посиланням на загальнодоступну загальну змінну маси з'єднання.

Клас QVis відповідає за відображення вже існуючих списків покупок, його коректну роботу забезпечує зв'язок з БД. Даний клас містить SQL запити.

Клас TableModalQ необхідний для коректного відображення вмісту обраного списку. Клас викликається з класу QVis, та може викликати клас Zmina, за умови що користувач хоче внести зміни до обраного списку.

Клас QutList використовується, коли користувач хоче створити новий список покупок. Цей клас асоціюється з класом Create, який взаємодіє зі смарт-модулем для покупок.

Клас Create містить основну частину інтелектуального модуля, який послідовно реалізує алгоритми NSC та дерево рішень. Також цей клас містить запити SQL, які допомагають встановлювати зв'язки між різними продуктами.

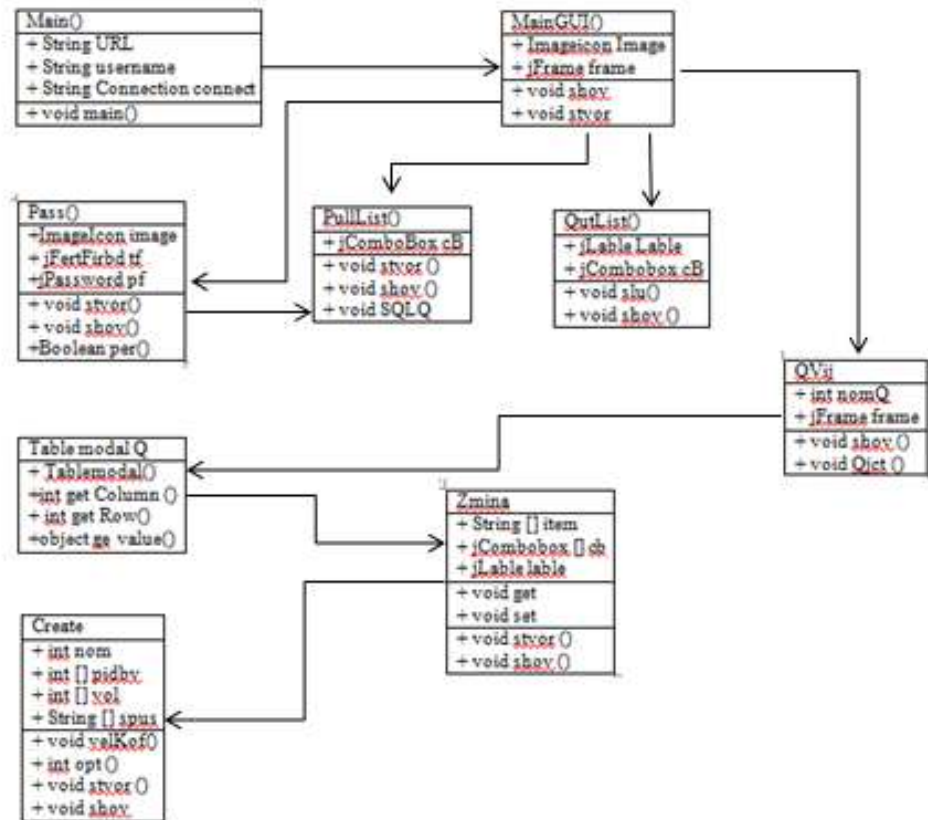


Рисунок 3.2 – Загальна UML-діаграма класів

3.3 Тестування інформаційної технології інтелектуального підбору покупок та аналіз результатів її роботи

Була протестована інформаційна технологія підбору покупок, що підтвердило правильність її роботи.

Було здійснено 1200 запусків програм. Проведено тестування для видалення та видалення наявного списку покупок, його редагування та створення нового.

Вікно початкової активності програми показано на малюнку 3.1. У ньому є кнопки для створення нового списку покупок, видалення та відкриття існуючого.

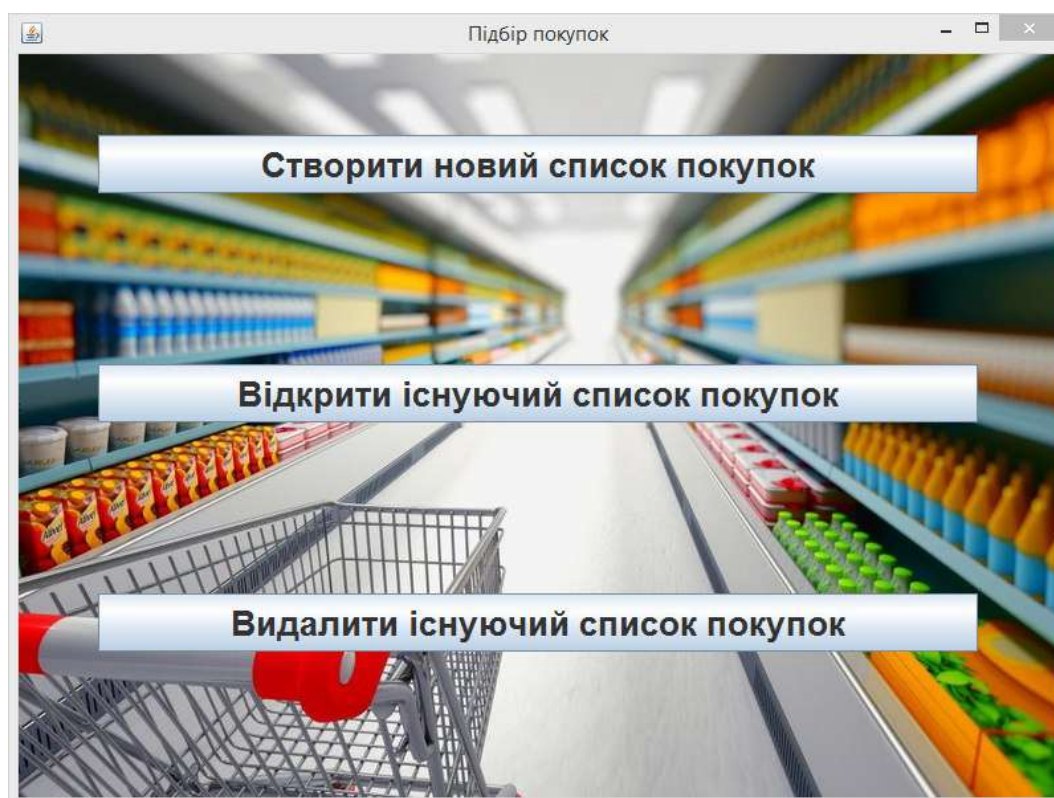


Рисунок 3.3 – Вікно початкової активності програми

Користувач може вибрати режим роботи з програмою, який йому потрібен в даний момент.

Створюючи новий список покупок, користувач вибирає потрібний їм товар, тоді програма надає відповідні рекомендації щодо наступної покупки. Користувач може вибрати продукт із рекомендованого або додати власний. Рекомендації надає інтелектуальний модуль програми (рисунок 3.3).

Модуль розумних покупок надає рекомендації щодо наступної покупки у списку, виходячи з того, які покупки користувач уже додав до списку та їх взаємозв'язки. Відносини складаються між продуктами, які вже є у списку покупок, та іншими можливими.

Продукти, які вже є у списку покупок, кластеруються за допомогою алгоритму НБК. Згодом надсилаються запити до бази даних, яка містить інформацію про взаємозв'язки між різними продуктами.

Процес створення нового списку покупок, видалення наявного або його редагування описаний у Додатку А.

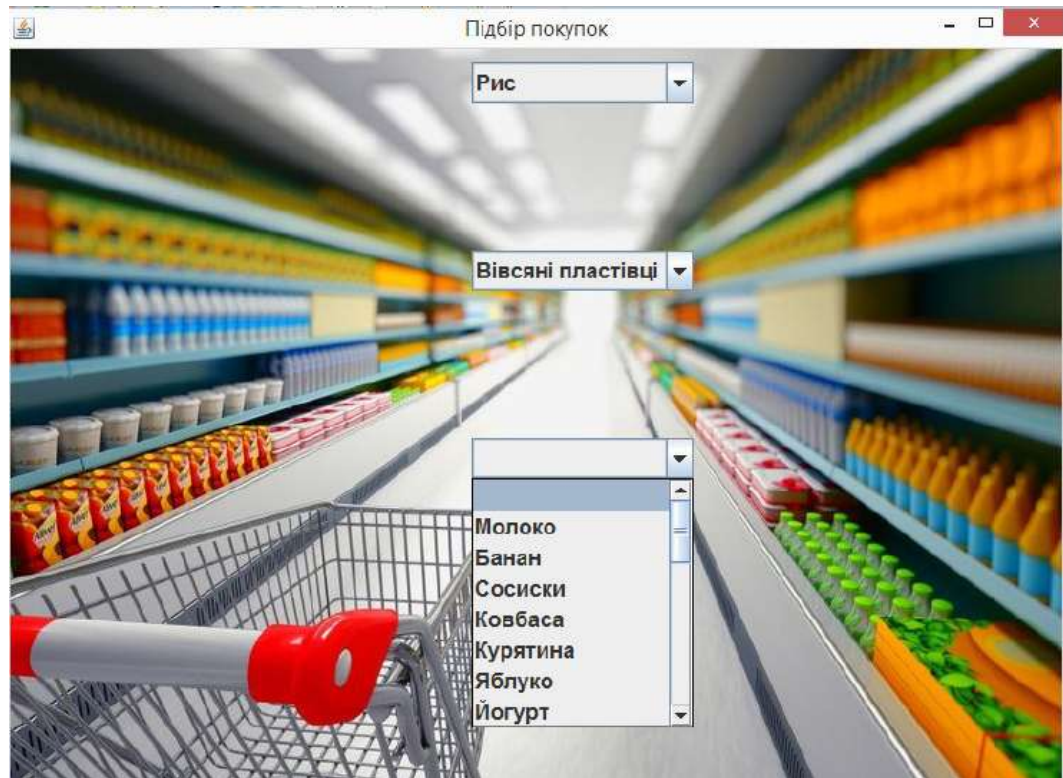


Рисунок 3.4 – Приклад процесу підбору покупок

В таблиці 3.4 наведено порівняльні результати роботи програми з програмами-аналогами.

Таблиця 3.2 – Порівняння результатів роботи програми

Програма	Швидкість підбору покупок, с
Розроблений додаток	$2 \cdot 10^{-2}$
«Купи батон!»	$2,06 \cdot 10^{-2}$
Remember the Milk	$2,2 \cdot 10^{-2}$
За покупками: Listick	$2,1 \cdot 10^{-2}$
Купи мене	$2,08 \cdot 10^{-2}$

Отже, в результаті тестування виявлено, що інформаційна технологія підбору покупок має підвищену швидкодію приблизно на 3% в порівняння з аналогами.

Результати тестування роботи програмного додатку, відповідають результатам, які очікувались.

3.4 Висновок

Здійснено обґрунтування вибору мови програмування для реалізації інформаційної технології підбору покупок – Java, яка є платформо незалежною, об'єктно-орієнтовною та має високу надійність. Обґрунтовано вибір середвища програмування Eclipse.

Розроблено та описано діаграму класів. Програмно реалізовано інформаційну технологію підбору покупок.

Проведено тестування роботи інформаційної технології підбору покупок та проаналізовано її результати. Тестування підтвердило правильність роботи програми та довело збільшення швидкодії майже на 3%.

4 ЕКОНОМІЧНА ЧАСТИНА

4.1 Оцінювання комерційного потенціалу розробки

Для визначення перспектив розробки магістерської кваліфікаційної роботи необхідно виконати наступні етапи економічної частини, які представлені на рисунку 4.1



Рисунок 4.1 – Алгоритм виконання економічної частини магістерської кваліфікаційної роботи та її складові

Метою проведення технологічного аудиту є оцінювання комерційного потенціалу розробки, створеної в результаті науково–технічної діяльності.

Для проведення технологічного аудиту залучено двох незалежних експертів. А саме Сілагін О. В. та Озеранський В. С. Оцінювання комерційного потенціалу буде здійснене за критеріями, що наведені в таблиці 4.1.

Таблиця 4.1 – Критерії оцінювання комерційного потенціалу розробки бальна оцінка

Критерії оцінювання та бали (за 5–ти бальною шкалою)					
Кри-терій	0	1	2	3	4
Технічна здійсненність концепції:					
1	Достовірність концепції не підтверджена	Концепція підтверджена експертними висновками	Концепція підтверджена розрахунками	Концепція перевірена на практиці	Перевірено роботоздатність продукту в реальних умовах

Продовження таблиці 4.1

Ринкові переваги (недоліки):					
2	Багато аналогів на малому ринку	Мало аналогів на малому ринку	Кілька аналогів на великому ринку	Один аналог на великому ринку	Продукт не має аналогів на великому ринку
3	Ціна продукту значно вища за ціни аналогів	Ціна продукту дещо вища за ціни аналогів	Ціна продукту приблизно дорівнює цінам аналогів	Ціна продукту дещо нижче за ціни аналогів	Ціна продукту значно нижче за ціни аналогів
4	Технічні та споживчі властивості продукту значно гірші, ніж в аналогів	Технічні та споживчі властивості продукту трохи гірші, ніж в аналогів	Технічні та споживчі властивості продукту на рівні аналогів	Технічні та споживчі властивості продукту трохи кращі, ніж в аналогів	Технічні та споживчі властивості продукту значно кращі, ніж в аналогів
5	Експлуатаційні витрати значно вищі, ніж в аналогів	Експлуатаційні витрати дещо вищі, ніж в аналогів	Експлуатаційні витрати на рівні експлуатаційних витрат аналогів	Експлуатаційні витрати трохи нижчі, ніж в аналогів	Експлуатаційні витрати значно нижчі, ніж в аналогів
Ринкові перспективи					
6	Ринок малий і не має позитивної динаміки	Ринок малий, але має позитивну динаміку	Середній ринок з позитивною динамікою	Великий стабільний ринок	Великий ринок з позитивною динамікою
7	Активна конкуренція великих компаній на ринку	Активна конкуренція	Помірна конкуренція	Незначна конкуренція	Конкурентів немає
Практична здійсненність					
8	Відсутні фахівці як з технічної, так і з комерційної реалізації ідеї	Необхідно наймати фахівців або витратити значні кошти та час на навчання наявних фахівців	Необхідне незначне навчання фахівців та збільшення їх Штату	Необхідне незначне навчання фахівців	Є фахівці з питань як з технічної, так і з комерційної реалізації ідеї
9	Потрібні значні фінансові ресурси, які відсутні. Джерела фінансування ідеї відсутні	Потрібні незначні фінансові ресурси. Джерела фінансування відсутні	Потрібні значні фінансові ресурси. Джерела фінансування є	Потрібні незначні фінансові ресурси. Джерела фінансування є	Не потребує додаткового фінансування
10	Необхідна розробка нових матеріалів	Потрібні матеріали, що використовуються у військово-промисловому комплексі	Потрібні дорогі матеріали	Потрібні досяжні та дешеві матеріали	Всі матеріали для реалізації ідеї відомі та давно використовуються у виробництві

Продовження таблиці 4.1

11	Термін реалізації ідеї більший за 10 років	Термін реалізації ідеї більший за 5 років. Термін окупності інвестицій більше 10-ти років	Термін реалізації ідеї від 3-х до 5-ти років. Термін окупності інвестицій більше 5-ти років	Термін реалізації ідеї менше 3-х років. Термін окупності інвестицій від 3-х до 5-ти років	Термін реалізації ідеї менше 3-х років. Термін окупності інвестицій менше 3-х років
12	Необхідна розробка регламентних документів та отримання великої кількості дозвільних документів	Необхідно отримання великої кількості дозвільних документів, що вимагає значних коштів та часу	Процедура отримання дозвільних документів у вимагає незначних коштів та часу	Необхідно тільки повідомлення відповідним органам про виробництво та реалізацію продукту	Відсутні будь-які регламентні обмеження на виробництво та реалізацію продукту

Щоб визначити рівень наукового та комерційного потенціалу нашої розробки, скористаємося даними таблиці 4.2, в якій наведено відповідні рекомендації [31].

Таблиця 4.2 – Рівні наукового та комерційного потенціалу розробки

Середньоарифметична сума балів $\overline{СБ}$, розрахована на основі висновків експертів	Рівень комерційного потенціалу розробки
0 – 10	Низький
11 – 20	Нижче середнього
21 – 30	Середній
31 – 40	Вище середнього
41 – 48	Високий

Результати оцінювання комерційного потенціалу розробки інформаційної технології підбору покупок зведемо в таблицю 4.3.

Таблиця 4.3– Результати оцінювання комерційного потенціалу розробки

Критерії	Прізвище, ініціали, посада експерта	
	1 – Сілагін	2 – Колесницький
	Бали, виставлені експертами:	
1	3	4
Ринкові переваги (недоліки):		
2	4	4
3	4	3
4	3	3
5	4	4
Ринкові перспективи		
6	3	4
7	3	3
Практична здійсненність		
8	4	3
9	3	4
10	4	4
11	4	4
12	3	4
Сума балів	СБ ₁ =42	СБ ₂ =44
Середньоарифметична сума балів $\overline{СБ}$	$\overline{СБ} = \frac{\sum_{i=1}^3 СБ_i}{2} = \frac{86}{2} = 43$	

Рівень комерційного потенціалу розробки, становить 43 бали, що відповідає рівню «високий».

4.2 Прогнозування витрат на виконання науково-дослідної роботи та конструкторсько–технологічної роботи

Проведемо прогнозування витрат на виконання науково–дослідної, дослідно–конструкторської та конструкторсько–технологічної роботи для розробки програмного забезпечення, яке складається з таких етапів:

Виконаємо розрахунок витрат приймаючи до уваги те, що розробкою займався один розробник програмного забезпечення.

1. Основна заробітна розробників–дослідника Z_o :

$$Z_o = \frac{M}{T_p} \cdot t \text{ [грн]}, \quad (4.11)$$

де M – місячний посадовий;

T_p – число робочих днів в місяці;

t – число робочих днів роботи розробника–дослідника.

У розробці програмного продукту беруть участь студент-магістрант та науковий керівник роботи.

Отримані розрахунки заробітної плати студента-магістранта та наукового керівника зведемо в таблицю 4.4.

Таблиця 4.4 – розрахунок витрат виконавців даного розділу роботи

Найменування посади виконавця	Місячний посадовий оклад, грн.	Оплата за робочий день, грн.	Число днів роботи	Витрати на опл. праці, грн.
1. Студент-магістрант	5500	261,9	53	13880
2. Науковий керівник роботи	6000	285,7	6	1714,3
Всього:				$\Sigma Z_o = 15594,3$

2. Додаткова заробітна плата Z_d всіх розробників, які брали участь у виконанні даного етапу роботи, розраховується як (10...12)% від суми основної заробітної плати всіх розробників, тобто:

$$Z_d = 0,10 \cdot 15308,5 = 1559,43 \text{ (грн).}$$

3. Нарахування на заробітну плату $H_{зп}$ розробника становить:

$$H_{зп} = (Z_o + Z_d) \cdot \frac{\beta}{100} \text{ [грн]}, \quad (4.12)$$

де Z_o – основна заробітна плата розробника;

Z_d – додаткова заробітна плата розробника;

β – ставка єдиного внеску на загальнообов’язкове державне соціальне страхування – 36,3%.

$$H_{\text{зп}} = (15594,3 + 1559,45) \cdot 0,363 = 6226,81 \text{ (грн)}.$$

4. Амортизація обладнання, комп’ютерів та приміщень А, які використовувались під час виконання даного етапу роботи.

Амортизаційні обчислення А в цілому бути розраховані за формулою 4.13:

$$A = \frac{Ц}{T_{\text{кор}}} \cdot \frac{T_{\text{факт.}}}{12} \text{ (грн)}, \quad (4.13)$$

де Ц – загальна балансова вартість всього обладнання, комп’ютерів, приміщень тощо, що використовувались в процесі виконання даної роботи, грн;

$T_{\text{факт.}}$ – термін, використання обладнання, приміщень, місяці. $T = 4$ місяці.

$T_{\text{кор.}}$ – термін використання, років.

Зроблені розрахунки зведемо до таблиці 4.5.

Таблиця 4.5 – розрахунок амортизації обладнання, які використовувалися у роботі

Найменування обладнання	Балансова вартість, грн.	Термін використання кор.	Термін використання, міс., факт.	Величина амортизаційних відрахувань, грн.
Персональний комп’ютер, принтер	10000	4	3	625
Приміщення	5000	20	3	62,5
Меблі	3000	4	3	187,5
Всього:				A=875

5. Витрати на матеріали M , що були використані під час виконання даного етапу роботи.

Розрахуємо витрати на комплектуючі. Витрати на комплектуючі розрахуємо за формулою:

$$K = \sum_1^n N_i \cdot C_i \cdot K_i, \quad (5.4)$$

де n – кількість комплектуючих;

N_i - кількість комплектуючих i -го виду;

C_i – покупна ціна комплектуючих i -го виду, грн;

K_i – коефіцієнт транспортних витрат (прийmemo $K_i = 1,1$).

Таблиця 4.6 - Витрати на комплектуючі, що були використані для розробки ПЗ

Найменування матеріалу	Одиниці виміру	Ціна, грн.	Витрачено	Вартість витрачених матеріалів, грн.
Флешка	шт.	250	1	250
Папір	уп.	130	1	130
Ручка	шт.	25	1	25
Всього з урахуванням транспортних витрат				445,5

6. Витрати на силову електроенергію V_e , якщо ця стаття має суттєве значення для виконання даного етапу роботи, розраховуються за формулою:

$$V_e = V \cdot P \cdot \Phi \cdot K_n, \quad \text{грн} \quad (4.14)$$

де V – вартість 1 кВт-год. електроенергії, в 2017 р. $V \approx 1,7$ грн./кВт;

P – установлена потужність обладнання, $P=0,6$ кВт;

Φ – фактична кількість годин роботи обладнання, прийmemo $\Phi=250$ год.;

K_{π} – коефіцієнт використання потужності; $K_{\pi} = 0,77$.

$$Ve = 1,7 \cdot 0,6 \cdot 250 \cdot 0,77 = 196,35 \text{ грн.}$$

7. Інші витрати $V_{\text{ін}}$.

Інші витрати $V_{\text{ін}}$ охоплюють: витрати на управління організацією, оплата службових відряджень, витрати на утримання, ремонт та експлуатацію основних засобів, витрати на опалення, освітлення, водопостачання, охорону праці тощо.

Інші витрати $I_{\text{в}}$ можна прийняти як (100...300)% від суми основної заробітної плати розробників та робітників, які виконували дану роботу, тобто:

$$V_{\text{ін}} = (1..3) \cdot (3_{\text{о}} + 3_{\text{р}}) \quad (4.15)$$

$$V_{\text{ін}} = 1 \cdot (15594,3 + 1559,43) = 17153,73 \text{ грн.}$$

8. Сума всіх попередніх статей витрат дає витрати на виконання даної частини (розділу, етапу) роботи – V .

$$V = 15594,3 + 1559,43 + 6226,81 + 875 + 445,5 + 196,35 + 20256,3 = 45153,69 \text{ грн.}$$

Розрахунок загальних витрат на виконання даної роботи. Загальна вартість всієї наукової роботи визначається за $V_{\text{заг}}$ формулою:

$$V_{\text{заг}} = \frac{V}{\alpha} [\text{грн}], \quad (4.16)$$

де α – частка витрат, які безпосередньо здійснює виконавець даного етапу роботи. Для нашого випадку приймемо, що $\alpha = 0,95$. Тоді:

$$V_{\text{заг}} = \frac{45153,69}{1} = 45153,69 \text{ (грн).}$$

Розрахуємо загальні витрати на виконання та можливе впровадження результатів виконаної роботи (4.17).

$$ЗВ = \frac{B_{заг}}{\beta} [грн], \quad (4.17)$$

де β – коефіцієнт, який характеризує етап (стадію) виконання даної роботи.

Так, якщо розробка знаходиться:

- на стадії науково–дослідних робіт, то $\beta \approx 0,1$;
- на стадії технічного проектування, то $\beta \approx 0,2$;
- на стадії розробки конструкторської документації, то $\beta \approx 0,3$;
- на стадії розробки технологій, то $\beta \approx 0,4$;
- на стадії розробки дослідного зразка, то $\beta \approx 0,5$;
- на стадії розробки промислового зразка, $\beta \approx 0,7$;
- на стадії впровадження, то $\beta \approx 0,9$.

Отже, підставимо дані в формулу й отримаємо результат:

$$ЗВ = \frac{45153,69}{0,9} = 50170,8 \text{ (грн)}.$$

Тобто, прогнозовані витрати на виконання та можливе впровадження результатів даної роботи можуть скласти приблизно 50170,8 тис. грн.

4.3 Прогнозування комерційних ефектів від реалізації результатів розробки

Комерційний ефект від можливого впровадження розробленої інформаційної технології полягає у тому, що ця система дозволить значно підвищити доставки товарів по торговій мережі та ефективніше використовувати ресурси.

Ці переваги дозволять реалізовувати розробку інформаційної технології на ринку дорожче, ніж аналогічні або подібні за функціональними характеристиками системи.

Спробуємо кількісно спрогнозувати, яку вигоду, можна отримати у майбутньому від впровадження результатів виконаної наукової роботи. Зрозуміло, що всі зроблені тут розрахунки будуть приблизними і непередбачають деталізації.

В умовах ринку узагальнюючим позитивним результатом, що його отримує підприємство від впровадження результатів тієї чи іншої розробки, є збільшення чистого прибутку підприємства (організації). Зростання чистого прибуткуми можемо оцінити у теперішній вартості грошей.

Розрахунок ефективності вкладених інвестицій передбачає проведення таких робіт:

Оцінка зростання чистого прибутку підприємства від впровадження результатів наукової розробки. У цьому випадку збільшення чистого прибутку підприємства $\Delta\Pi_i$ для кожного із років, протягом яких очікується отримання позитивних результатів від впровадження розробки, розраховується за формулою:

$$\Delta\Pi_i = \sum_1^n (\Delta\Pi_{\text{я}} \cdot N + \Pi_{\text{я}} \Delta N)_i \quad (5.9)$$

де $\Delta\Pi_{\text{я}}$ – покращення основного якісного показника від впровадження результатів розробки у даному році;

N – основний кількісний показник, який визначає діяльність підприємства у даному році до впровадження результатів наукової розробки;

ΔN – покращення основного кількісного показника діяльності підприємства від впровадження результатів розробки;

$\Pi_{\text{я}}$ – основний якісний показник, який визначає діяльність підприємства у даному році після впровадження результатів наукової розробки;

n – кількість років, протягом яких очікується отримання позитивних результатів від впровадження розробки.

В результаті впровадження результатів наукової розробки витрати на виготовлення інформаційної технології зменшаться на 50 грн (що автоматично спричинить збільшення чистого прибутку підприємства на 50 грн), а кількість користувачів, які будуть користуватись збільшиться: протягом першого року – на 200 користувачів, протягом другого року – на 150 користувачів, протягом третього року – 100 користувачів. Реалізація інформаційної технології до впровадження результатів наукової розробки складала 510 користувачів, а прибуток, що отримував розробник до впровадження результатів наукової розробки – 200 грн.

Спрогнозуємо збільшення чистого прибутку від впровадження результатів наукової розробки у кожному році відносно базового.

Отже, збільшення чистого продукту $\Delta\Pi_1$ протягом першого року складатиме:

$$\Delta\Pi_1 = 50 \cdot 510 + (200 + 50) \cdot 200 = 75500 \text{ грн.}$$

Протягом другого року:

$$\Delta\Pi_2 = 50 \cdot 510 + (200 + 40) \cdot (200 + 150) = 109500 \text{ грн.}$$

Протягом третього року:

$$\Delta\Pi_3 = 50 \cdot 105 + (200 + 50) \cdot (200 + 150 + 100) = 138000 \text{ грн.}$$

Вартість інвестицій $PV=3B=50170,8$ грн., що можуть бути вкладені в нашу розробку.

4.4 Розрахунок ефективності вкладених інвестицій та період їх окупності

Визначимо абсолютну і відносну ефективність вкладених інвестором інвестицій та розрахуємо термін окупності.

Абсолютна ефективність $E_{абс}$ вкладених інвестицій розраховується за формулою:

$$E_{абс} = (ПП - PV), \quad (4.19)$$

де $\Delta\Pi_i$ – збільшення чистого прибутку у кожному із років, протягом яких виявляються результати виконаної та впровадженої НДДКР, грн;

t – період часу, протягом якого виявляються результати впровадженої НДДКР, 3 роки;

τ – ставка дисконтування, за яку можна взяти щорічний прогнозований рівень інфляції в країні; для України цей показник знаходиться на рівні 0,1;

t – період часу (в роках) від моменту отримання чистого прибутку до точки 2, 3, 4.

Рисунок, що характеризує рух платежів (інвестицій та додаткових прибутків) буде мати вигляд, рисунок 4.1.



Рисунок 4.1 – Вісь часу з фіксацією платежів, що мають місце під час розробки та впровадження результатів НДДКР

У свою чергу, приведена вартість всіх чистих прибутків ПП розраховується за формулою:

$$\text{ПП} = \sum_{i=1}^{\tau} \frac{\Delta\Pi_i}{(1 + \tau)^t}, \quad (4.20)$$

де $\Delta\Pi_i$ – збільшення чистого прибутку у кожному ізроків, протягом яких виявляються результати виконаної та впровадженої НДДКР, грн;

τ – період часу, протягом якого виявляються результати впровадженої НДДКР, роки;

τ – ставка дисконтування, за яку можна взяти щорічний прогнозований рівень інфляції в країні; для України цей показник знаходиться на рівні 0,1;

t – період часу (в роках) від моменту отримання чистого прибутку до точки „0”.

Якщо $E_{\text{абс}} > 0$, то результат від проведення наукових досліджень та їх впровадження принесе прибуток, але це також ще не свідчить про те, що інвестор буде зацікавлений у фінансуванні даного проекту (роботи).

Розрахуємо абсолютну ефективність інвестицій, вкладених у реалізацію проекту. Ставка дисконтування τ дорівнює 0,1. Отримаємо:

$$ПП = \frac{50170,8}{(1 + 0,1)^0} + \frac{75500}{(1 + 0,1)^2} + \frac{109500}{(1 + 0,1)^3} + \frac{138000}{(1 + 0,1)^4} = 289098,8(\text{грн}).$$

Абсолютний ефект від впровадження результатів нашої розробки протягом 3-х років складе:

$$E_{\text{абс}} = 289098,8 - 50170,8 = 238928,8 \text{ (грн)}.$$

Оскільки $E_{\text{абс}} > 0$, то вкладання коштів на виконання та впровадження результатів НДДКР є доцільним.

Розраховуємо відносну (щорічну) ефективність вкладених в наукову розробку інвестицій E_v . Для цього використовуємо формулу:

$$E_v = \sqrt[T_{\text{ж}}]{1 + \frac{E_{\text{абс}}}{PV}} - 1, \quad (4.20)$$

де $E_{\text{абс}}$ – абсолютна ефективність вкладених інвестицій, грн;

PV –теперішня вартість інвестицій $PV = 3B = 50170,8$ грн;

$T_{\text{ж}}$ – життєвий цикл наукової розробки, роки.

Тоді маємо:

$$E_v = \sqrt[3]{1 + \frac{238928,8}{50170,8}} - 1 = 0,8 \text{ або } 80\%$$

Далі, розрахована величина E_v порівнюється з мінімальною (бар'єрною) ставкою дисконтування τ мін, яка визначає ту мінімальну дохідність, нижче за яку інвестиції вкладатися не будуть. У загальному вигляді мінімальна (бар'єрна) ставка дисконтування $\tau_{\text{мін}}$ визначається за формулою:

$$\tau = d + f,$$

де d – середньозважена ставка за депозитними операціями в комерційних банках; в 2019 році в Україні $d = 0,2$;

f – показник, що характеризує ризикованість вкладень, величина $f = 0,1$.

$$\tau = 0,2 + 0,1 = 0,3$$

Оскільки $E_B = 80 \% > \tau_{\min} = 0,3 = 30\%$, то у інвестор буде зацікавлений вкладати гроші в дану наукову розробку.

Розраховуємо термін окупності вкладених у реалізацію наукового проекту інвестицій. Термін окупності вкладених у реалізацію наукового проекту інвестицій $T_{ок}$ можна розрахувати за формулою:

$$T_{ок} = \frac{1}{E_B}. \quad (4.22)$$

Якщо $T_{ок}$ буде менше 5–ти років, то фінансування даної наукової розробки загалом є доцільним. В інших випадках потрібні додаткові розрахунки та обґрунтування. Для нашої розробки термін окупності вкладених у реалізацію проекту інвестицій $T_{ок}$ складе:

$$T_{ок} = \frac{1}{0,8} = 1,25 \text{ (року)}$$

Обрахувавши термін окупності даної наукової розробки, можна зробити висновок, що фінансування даної наукової розробки буде доцільним.

4.5 Висновок

На основі зроблених підрахунків в економічній частині магістерської кваліфікаційної роботи досягнуті наступні результати:

– визначено, що рівень комерційного потенціалу розробки є середнім. Для успішної конкурентоспроможності програмного продукту на ринку планується збільшити універсальність програми, що забезпечить більший попит у користувачів.

– витрати на розробку та її впровадження складають 50170,8 тис. грн.;

– абсолютний ефект від впровадження результатів нашої розробки протягом 3-х років складе 238928,8 тис. грн.

– вартість інвестицій, що можуть бути вкладені в нашу розробку становить 50170,8 тис. грн;

– термін окупності системи, що розробляється складає 1,25 року, що вписується в задані у ТЗ часові рамки та є показником доцільності розробки.

Таким чином, всі визначені у технічному завданні основні техніко-економічні показники розробленої інформаційної технології підбору покупок повністю виконані.

ВИСНОВКИ

У ході виконання магістерської кваліфікаційної роботи вдосконалено технологію підбору покупок в тому числі для розумного дому, що відрізняється від відомих застосуванням поєднання методу наївного Баєсівського класифікатора та методу дерева рішень для підбору покупок та забезпечує ефект підвищення швидкості підбору покупок.

У першому розділі розглянуто та проаналізувати існуючі методи та технології розв'язання задачі підбору покупок. Проведено аналіз сучасних програм-аналогів, які використовуються для автоматизації процесу підбору необхідних покупок, наведено порівняльну характеристику знайдених програм-аналогів та визначено їх недоліки, що дало змогу обрати прототип. В результаті аналізу аналогів та методів підбору покупок сформульовано проблему підбору покупок та здійснимо постановку задачі.

Другий розділ присвячено математичним методам та алгоритмам, що використовуються для задач складання списків покупок та дослідженню їх можливостей. Розроблено математичну модель, засновану на комбінації методів наївного Баєсівського класифікатора та дерева рішень для підбору покупок, , що за рахунок простоти і швидкості роботи алгоритмів призводять до збільшення швидкості підбору покупок. Запропоновано структуру інформаційної технології, спроектовано схему алгоритму роботи розроблюваної програми та ER-діаграму.

Третій розділ присвячено проектуванню та розробці інформаційної технології підбору покупок. Обґрунтовано вибір мови програмування Java яка є платформо незалежною, об'єктно-орієнтовною та має високу надійність; розглянуто особливості середовища Eclipse.

На основі отриманих результатів у попередніх розділах створено UML-діаграму класів. Відповідно до математичної моделі, схеми алгоритму та UML-діаграми класів реалізовано інформаційну технологію підбору покупок.

Проведено тестування програми, що довело доцільність проектування та використання розробленої інформаційної технології. Порівняно з обраним

прототипом швидкість підбору покупок збільшилась майже на 3%, що означає досягнення поставленої мети.

У четвертому розділі здійснено оцінювання комерційного потенціалу розробки. Проведено технологічний аудит із залученням експертів. Згідно висновків експертів, рівень комерційного потенціалу розробки вище середнього. Головним конкурентом розробки є мобільний додаток «Купи батон!». Здійснено прогнозування витрат на виконання науково-дослідної роботи. Розраховано витрати на заробітну плату та амортизаційні відрахування, витрати на силову електроенергію. Загальні витрати становлять 238928,8 тис. грн.

За результатами досліджень опубліковано тези доповіді у збірнику праць всеукраїнської науково-практичної інтернет-конференції [1, 2, 3]. Також подано заяву на оформлення авторського права.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Замковий О. Д., Петришин С. І. Аналіз основних методів інтелектуального аналізу даних для задачі підбору покупок / Молодь в науці: дослідження, проблеми, перспективи (МН-2019).
URL: <https://conferences.vntu.edu.ua/index.php/mn/mn2019/paper/view/8184>
2. Замковий О. Д., Петришин С. І. Оцінка актуальностей методів інтелектуального аналізу даних для задачі підбору покупок / XLVII Науково-технічна конференція факультету інформаційних технологій та комп'ютерної інженерії (2018)
URL: <https://conferences.vntu.edu.ua/index.php/all-fitki/all-fitki-2018/paper/view/5274>
3. Замковий О. Д., Петришин С. І. Аналіз актуальності методів інтелектуального аналізу даних для задачі підбору покупок / XI міжнародній науково-практичній конференції «ІОН-2018»
URL: <https://conferences.vntu.edu.ua/index.php/all-fitki/all-fitki-2018/paper/view/5274>
4. А.П. Челенков. — М.: ОАО Типография «Новости», 2000. 256 с.
5. Акентьева, П.А. Модель потребительского поведения / П.А. Акентьева // Экономика и управление производством. Межвузовский сборник / Под ред. проф. Е.Л. Кантора и др. СПб.: 2005.
6. Акерлоф, Дж. Рынок «лимонов»: неопределенность качества и рыночный механизм // THESIS, 1994. вып. 5. - С.91-104.
7. Алешина, И.В. Поведение потребителей / И.В. Алешина. М.: ФИАР ПРЕСС, 1999.-394 с.
8. Аллэ, М. Поведение рационального человека в условиях риска: Критика постулатов и аксиом американской школы / М. Аллэ // THESIS, 1994, вып. 5.
9. Андрианов, В.Д. Конкурентоспособность России в мировой экономике /
10. В.Д. Андрианов // Мировая экономика и международные отношения. — 2000. -№3.-С. 47-57.

11. Антосик, Л.В. Информационная асимметричность рыночного пространства как условие монополизации экономики / Л.В. Антосик // Дис. канд. экон. наук: 08.00.01. Волгоград, 2006.
12. Аристотель, Категории // Этика. Политика. Риторика. Поэтика. Категории. Минск: Литература, 1998. 438 с.
13. Арнольд, В.Ф. Теория потребительского поведения и спроса / В.Ф. Арнольд. — СПб: Экономическая школа, 1992.
14. Ахмеджанов, Л.А. Социально-экономические проблемы взаимосвязи производства и потребления при социализме / Л.А. Ахмеджанов // Производство и потребление при социализме. — Уфа: Башкир, ун-т, 1974. — 185 с.
15. Бакшеев, С.Л. Потребительский рынок как форма реализации потребностей населения / С.Л. Бакшеев. Изд. 2-е. - Тюмень: Экспресс, 2004.
16. А.В.Вишнеков. Методы принятия проектных решений в САД/САМ/САЕ системах электронной вычислительной техники: Учебное пособие. - МГИЭМ. - М., 1999. - 78 с.
17. Методы экспертных оценок. - Методические указания к лабораторной работе по курсу "Системы поддержки принятий решений". - МГИЭМ. Сост.: А.В.Вишнеков, И.Е.Сафонова, Н.С.Курилова, В.И.Бадулин. - М., 2001. - 24 с.
18. Продукционная модель представления знаний в экспертных системах. - Методические указания к лабораторной работе по курсу "Системы поддержки принятий решений". - МГИЭМ. Сост.: А.В.Вишнеков, И.Е.Сафонова, Н.С.Курилова. - М., 2001. - 20 с.
19. Принятие решений в условиях неопределенности на основе аппарата нечетких множеств. - Методические указания к лабораторной работе по курсу "Системы поддержки принятий решений". - МГИЭМ. Сост.: А.В.Вишнеков, И.Е.Сафонова, Н.С.Курилова, А.В.Киселев. - М., 2001. - 33 с.
20. Методы принятия решений. Методы Дельфи и ЭЛЕКТРА. - Методические указания к лабораторной работе по курсу "Системы поддержки принятий

- решений". - МГИЭМ. Сост.: И.Е.Сафонова, А.В.Дробышев, К.Ю.Мишин, С.В.Цыганов: М., МГИЭМ, 2007. - 26 с.
- 21.J. Ross Quinlan. C4.5: Programs for Machine learning. Morgan Kaufmann Publishers 1993.
 - 22.S.Murthy. Automatic construction of decision trees from data: A Multi-disciplinary survey.1997.
 - 23.Machine Learning, Neural and Statistical Classification. Editors D. Mitchie et.al. 1994.
 - 24.К. Шеннон. Работы по теории информации и кибернетике. М. Иностранная литература, 1963.
 - 25.С.А. Айвазян, В.С Мхитарян Прикладная статистика и основы эконометрики, М. Юнити, 1998.
 - 26.Дрейпер Н., Смит Г. Прикладной регрессионный анализ. М.: Издательский дом «Вильямс». 2007.
 - 27.Айвазян С.А. Прикладная статистика и основы эконометрики. М.: Юнити. 2001.
 - 28.МакКай, D. Hyperparameters: optimize, or integrate out? Neural Computation. 1999.
 - 29.Nabney, Yan T., Netlab: Algorithms for pattern recognition. Springer. 2004.
 - 30.Grunwald, P D., Myung, I. J. (eds.) Advances In Minimum Description Length: Theory And Applications. Springer. 2005.
 - 31.Burnham, K., Anderson, D. R. Model Selection and Multimodel Inference. Springer. 2002.
 - 32.Купи батон! – самый удобный список покупок [Электронный ресурс] URL: <https://appleinsider.ru/obzory-prilozhenij/app-store-hd-kupi-baton-samyj-udobnyj-spisok-pokupok.html>
 - 33.Обзор Remember the Milk. Молочный список задач с мозгами [Электронный ресурс] URL: <http://gadgetick.com/newsnew-4176.html>
 - 34.Обзор Clear [Электронный ресурс] URL: <https://clean-master-for-pc.ru.softonic.com>
 35. За покупками: Listick [Электронный ресурс] URL: <https://captain-droid.com/apps/tools/za-pokupkami/>

36. Купи меня: обзор плюсов и минусов [Электронный ресурс] URL: http://4pda.ru/2013/03/12/92579/?utm_source=older
37. 13. Ахиезер Н. И. Лекции по теории аппроксимации. – М.: Наука, 1965.
- 38.14. В.И. Бердышев, Ю.Н. Субботин. Численные методы приближения функций. – Средне-Уральское книжное издательство, 1979.
- 39.15. Мак – Кракен Д., Дрон У. Численные методы и программирование на фортране. – М.: Мир, 1977. – 584 с.
- 40.16. Бахвалов Н. С. Численные методы . Т. И. Анализ, алгебра, обычные дифференциальные уравнения. – М.: Наука, 1975. – 631 с.
- 41.Гмурман В. Е. Теория вероятностей и математическая статистика, — М.: Высшее образование. 2005
- 42.Берд Киви. Теорема преподобного Байеса. // Журнал «Компьютер» августа 2001 г.
- 43.Judgment under Uncertainty: Heuristics and Biases / Daniel Kahneman, et al. — 21st. — Cambridge University Press, 2005. — 555 p. — ISBN 978-0-521-28414-1
44. Bayes, Thomas, and Price, Richard (1763). «An Essay towards solving a Problem in the Doctrine of Chance. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, M. A. and F. R. S.». Philosophical Transactions of the Royal Society of London 53: 370—418.
45. Seidman.C. Data Mining with Microsoft SQL Server 2000: Technical Reference. — Microsoft Press, 2001.
46. Ville.B., de. Microsoft Data Mining. — Digital Press, 2001
47. Java основні відомості [Электронный ресурс] – режим доступа: <https://uk.wikipedia.org/wiki/Java>
48. Порівняння мов програмування [Электронный ресурс] – режим доступа:https://ru.wikipedia.org/wiki/Сравнение_языков_программирования
49. Брюс Эккель – «Философия Java».
50. Eclipse основні відомості [Электронный ресурс] – режим доступа: [https://ru.wikipedia.org/wiki/Eclipse_\(среда_разработки\)](https://ru.wikipedia.org/wiki/Eclipse_(среда_разработки))

51. А. Н. Васильев – Java. «Объектно-ориентированое программирование».
52. Т. Конноли, К. Бегг, А. Страчан. Базы данных: проектирование, реализация и сопровождение. – М.: Диалектика, 2000. – 1139 с.
53. Мейер М. Теория реляционных баз данных. - М.: Мир, 1987. - 608 с.
54. Гайна Г.А. – Основи проектування баз даних: Навчальний посібник. – К.: КНУБА, 2005. – 204 с.
55. Джудит С. Боуманб Сандра Л. Эмерсон, М. Дарновски. Практическое руководство по SQL 4-е издание. – Москва: Вильямс, 2001. – 352с.