

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук _____
(повна назва)
Кафедра _____ Системотехніки _____
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА

Пояснювальна записка

рівень вищої освіти _____ другий (магістерський) _____
«Розробка та дослідження методів формування рекомендацій для системи підбору
лікарських препаратів» _____
(тема)

Виконав:
студент II курсу, групи СПРМ-22-2 _____
Підгорний М.О. _____
(прізвище, ініціали)

Спеціальність 122 Комп'ютерні науки _____
(код і повна назва спеціальності)

Тип програми освітньо-наукова _____
(освітньо-професійна або освітньо-наукова)

Освітня програма Системне проектування _____
(повна назва освітньої програми)

Керівник доц. каф. СТ Ситнікова П.Е. _____
(посада, прізвище, ініціали)


Допускається до захисту
Зав. кафедри _____

(підпис)

Гребеннік І.В. _____
(прізвище, ініціали)

2024 р.

Я, як студент ХНУРЕ, розумію та підтримую політику закладу із академічної доброчесності. Я не надавав і не одержував недозволену допомогу під час підготовки кваліфікаційної роботи. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.

підпис _____ 

Підгорний М. О.

Кваліфікаційна робота не містить відомостей заборонених до відкритого опублікування.

Кваліфікаційна робота виконана у відповідності до стандартів, що діють в Україні.

Попередній захист проведений 21 червня 2024 р.

Керівник кваліфікаційної роботи _____  доц. каф. СТ Ситнікова П.Е.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерні науки
Кафедра Системотехніки
Рівень вищої освіти другий (магістерський)
Спеціальність 122 Комп'ютерні науки
(код і повна назва)
Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)
Освітня програма Системне проектування
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____

(підпис)

« _____ » _____ 2024 р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Підгорному Микиті Олександровичу
(прізвище, ім'я, по батькові)

1. Тема роботи «Розробка та дослідження методів формування рекомендацій для системи підбору лікарських препаратів»

затверджена наказом університету від 01 квітня 2024 р. № 259Ст

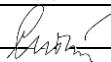
2. Термін подання студентом роботи до екзаменаційної комісії 21 червня 2024 р.

3. Вихідні дані до роботи Дослідити та розробити методи формування рекомендацій для системи підбору лікарських препаратів. Дослідження включає аналіз предметної області, актуальність проблеми дослідження та огляд методів формування рекомендацій. Розробка включає виокремлення найкращого з методів формування рекомендацій для предметної області ліків, його реалізацію та порівняльний аналіз; програмне забезпечення: мова програмування Python; бібліотеки Pandas, Scikit-learn, TensorFlow; текстовий редактор Microsoft Visual Studio Code; онлайн засіб проектування діаграм Draw.io.

4. Перелік питань, що потрібно опрацювати в роботі 4.1 Вступ. 4.2 Аналіз предметної області. 4.3 Актуальність проблеми дослідження. 4.4 Дослідження існуючих методів формування рекомендацій. 4.5 Дослідження методів машинного навчання. 4.6 Розробка методу формування рекомендацій. 4.7 Валідація створеного методу. 4.8 Аналіз результатів. 4.9 Висновки. 4.10 Перелік джерел посилання.

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри)
5.1 Фармацевтичний ринок. 5.2 Діаграма довіри населення до ліків. 5.3 Діаграма моделі формування рекомендацій для системи підбору ліків. 5.4 Діаграма представлення випадкового лісу. 5.5 Діаграма LSTM нейромережі. 5.6 Результати для порівняння методів. 5.7 Діаграма діяльності методу формування рекомендацій ліків. 5.8 Схема валідаційного експерименту. 5.9 Зведення результатів тестового експерименту.


6. Консультанти розділів роботи (п.6 включається до завдання за наявності консультантів згідно з наказом, зазначеним у п.1)


Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата
Аналіз предметної області	доц. каф. СТ Ситнікова П.Е.		

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів	Примітка
1	Отримання завдання кваліфікаційної роботи	01.04.2024	Виконано
2	Аналіз предметної області	12.04.2024	Виконано
3	Дослідження існуючих методів формування рекомендацій	24.04.2024	Виконано
4	Аналіз існуючих рішень для систем підбору лікарських препаратів	30.04.2024	Виконано
5	Створення методу формування рекомендацій	05.05.2024	Виконано
6	Валідація створеного методу	17.05.2024	Виконано
7	Аналіз отриманих результатів	25.05.2024	Виконано
8	Оформлення пояснювальної записки	10.06.2024	Виконано
9	Подача кваліфікаційної роботи на допуск до захисту	19.06.2024	Виконано
10	Подання кваліфікаційної роботи	21.06.2024	Виконано

Дата видачі завдання 01 квітня 2024 р.

Студент 
 (підпис)

Керівник роботи  доц. каф. СТ Ситнікова П.Е.
 (підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка до магістерської кваліфікаційної роботи: 82 арк., 41 рис., 2 додатки, 40 джерел інформації.

СИСТЕМИ ПІДБОРУ ЛІКАРСЬКИХ ПРЕПАРАТІВ, МЕТОДИ ФОРМУВАННЯ РЕКОМЕНДАЦІЙ, МАШИННЕ НАВЧАННЯ, ФІЛЬТРАЦІЯ, КЛАСИФІКАЦІЯ, ЛІКИ, ВИПАДКОВИЙ ЛІС, LSTM, НЕЙРОННІ МЕРЕЖІ

Об'єктом дослідження є процеси аналізу та вибору лікарських засобів з використанням інформаційних технологій у фармацевтиці.

Предметом дослідження є методи формування рекомендацій для автоматизованого підбору лікарських препаратів, з урахуванням індивідуальних характеристик пацієнта та взаємодій між препаратами.

Мета дослідження – розробити ефективний метод формування рекомендацій для системи підбору лікарських препаратів, який би підвищував точність і персоналізацію підбору медикаментів, мінімізуючи ризик негативних взаємодій між препаратами та неефективного лікування.

Результати дослідження – розроблений метод формування рекомендацій, що оцінює сумісність препаратів та потенційну ефективність для широкого спектру стандартних ситуацій, заснованих на фармакологічних принципах та базується на загальнодоступних даних про лікарські препарати та їхній взаємодії.

Галузь застосування – впровадження в аптечні мережі, онлайн-сервіси для самопомоги у виборі безрецептурних лікарських засобів, розробка додатків для інформування пацієнтів про можливості самолікування та профілактики здоров'я, електронні довідники лікарських препаратів.

ABSTRACT

Explanatory note to the master's qualification work: 82 pages, 41 figures, 2 appendices, 40 sources of information.

DRUG SELECTION SYSTEMS, METHODS OF RECOMMENDATION GENERATION, MACHINE LEARNING, FILTERING, CLASSIFICATION, MEDICINES, RANDOM FOREST, LSTM, NEURAL NETWORKS

The object of research is the processes of analysis and selection of medicines using information technology in pharmaceuticals.

The subject of research is the methods of forming recommendations for automated drug selection, taking into account the individual characteristics of the patient and interactions between drugs.

The aim of the research is to develop an effective method for generating recommendations for a drug selection system that would increase the accuracy and personalization of drug selection, minimizing the risk of negative drug interactions and ineffective treatment.

The results of the research – a method for generating recommendations that assesses drug compatibility and potential efficacy for a wide range of standard situations based on pharmacological principles and based on publicly available data on drugs and their interactions.

The field of application includes implementation in pharmacy networks, online services for self-help in selecting over-the-counter medications, development of applications for informing patients about self-treatment options and health prevention, electronic reference books of medications.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ	7
ВСТУП.....	8
1 ДОСЛІДЖЕННЯ ПРОБЛЕМИ ФОРМУВАННЯ РЕКОМЕНДАЦІЙ ДЛЯ СИСТЕМИ ПІДБОРУ ЛІКАРСЬКИХ ПРЕПАРАТІВ	9
1.1 Аналіз предметної області.....	9
1.2 Актуальність проблеми підбору лікарських препаратів.....	12
1.3 Огляд даних про лікарські препарати.....	18
1.4 Схема процесів підбору лікарських препаратів	19
1.5 Постановка задачі формування рекомендацій для СПЛП.....	21
2 ДОСЛІДЖЕННЯ ІСНУЮЧИХ МЕТОДІВ ФОРМУВАННЯ РЕКОМЕНДАЦІЙ.....	23
2.1 Огляд існуючих рішень для СПЛП.....	23
2.2 Системний аналіз задачі формування рекомендацій для СПЛП	32
3. СТВОРЕННЯ МЕТОДУ ФОРМУВАННЯ РЕКОМЕНДАЦІЙ ЛІКІВ	37
3.1 Схема створеного методу	37
3.2 Дослідження та аналіз існуючих методів.....	41
3.3 Методи машинного навчання та гібридні методи.....	47
3.4 Проведення аналізу існуючих методів на основі тестового експерименту ...	66
3.5 Створення методу формування рекомендацій ліків	69
4 ВАЛІДАЦІЯ СТВОРЕНОГО МЕТОДУ	72
4.1 Розробка програмної компоненти для проведення експерименту	72
4.2 Планування експерименту.....	73
4.3 Аналіз отриманих результатів.....	76
ВИСНОВКИ	78
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ	79

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ

ІС – інформаційна система

НМ – нейронна мережа

СПЛП – системи підбору лікарських препаратів

ШІ – штучний інтелект

CDSS – computer decision support system (комп'ютерна система підтримки прийняття рішень)

DSS – decision support system (система підтримки прийняття рішень)

ML – machine learning (машинне навчання)

ВСТУП

В умовах стрімкого розвитку фармацевтичної індустрії та постійного збільшення асортименту лікарських засобів, вибір оптимального лікування стає все більш складним завданням як для медичних фахівців, так і для пацієнтів. Проблема ускладнюється обмеженим доступом до детальних медичних даних та історії лікування, що неминуче веде до потреби у вдосконаленні інструментів для підтримки прийняття рішень. Саме тому актуальним стає завдання розробки та дослідження інноваційних методів формування рекомендацій для системи підбору лікарських препаратів, які б дозволяли здійснити цей вибір ефективно та безпечно, навіть без прямого доступу до особистих медичних даних.

Метою дослідження є розробка методу формування рекомендацій, який покликаний забезпечити вищу точність та персоналізацію вибору медикаментів для кожного пацієнта, враховуючи їх унікальні особливості. Цей підхід не тільки сприяє підвищенню якості медичного обслуговування, але й забезпечує більшу безпеку пацієнтів, базуючись на усвідомленому та обґрунтованому виборі лікарських засобів. Такий підхід підвищує ефективність лікування та мінімізує ризики небажаних наслідків.

Дослідження зосереджене на аналізі існуючих методів, ідентифікації їх обмежень та розробці нового підходу, який спеціально адаптований до потреб медичної галузі. Воно включає оцінку ефективності нового методу через ряд перевірок, що базуються на аналізі інформації про лікарські засоби, їх склад і фармакологічну дію, а також проведення тестів у реальних умовах та порівняння з уже існуючими рішеннями.

Таким чином, це дослідження покликане внести вклад у розвиток фармацевтичної інформатики та медичних інформаційних систем, розширюючи можливості для раціонального використання лікарських засобів у сучасному світі, де технології відіграють все більш значущу роль у підтримці та покращенні здоров'я людей.

1 ДОСЛІДЖЕННЯ ПРОБЛЕМИ ФОРМУВАННЯ РЕКОМЕНДАЦІЙ ДЛЯ СИСТЕМИ ПІДБОРУ ЛІКАРСЬКИХ ПРЕПАРАТІВ

1.1 Аналіз предметної області

У сучасному світі населення планети перевищило 8 мільярдів осіб, і, незважаючи на значне покращення рівня життя та якості медицини, постійно зростає відсоток хворих людей. Це пов'язано не тільки зі збільшенням відносної частини літнього населення, але й з тим, що захворювання, які раніше вважалися характерними для осіб старшого віку, такі як цукровий діабет та серцево-судинні порушення, все частіше діагностуються у молодших вікових групах.

Також важливо відмітити, що попри досягнення у медицині та інформаційних технологіях продовжуються епідемії та пандемії, як остання COVID-19, які виявили неможливість медичних систем більшості країн вирішити проблеми такого типу. Однак, покращення методів створення ліків, поліпшення рекомендацій завдяки залученню інформаційних технологій, а також дослідження захворювань надають умови за яких лікарські препарати зберігають життя людей, тому їхнє застосування є дуже важливим для світу.

Одним з головних напрямків розвитку медицини є впровадження електронних медичних записів (ЕМЗ), що дозволяють зберігати, аналізувати та використовувати медичні дані пацієнтів. ЕМЗ сприяють підвищенню якості медичних послуг, забезпечуючи доступ до актуальної інформації про стан здоров'я пацієнтів, їх історію хвороб, призначення та результати лікування. Це, в свою чергу, допомагає лікарям приймати більш обґрунтовані та ефективні рішення.

Сучасна медицина постійно стикається з новими викликами, пов'язаними зі збільшенням кількості нових лікарських засобів, ускладненням схем лікування та індивідуалізацією медичних підходів до пацієнтів. Сфера медичних препаратів в свою чергу включає широкий спектр продуктів, від традиційних лікарських засобів до біотехнологічних інновацій, що використовуються для лікування, профілактики та діагностики захворювань.

Відповідно до цих факторів фармацевтичні компанії, науковці та держави постійно залучені у створенні нових ліків задля покращення рівня життя і подоланням проблем викликаних хворобами. Тому необхідно оглянути сам процес появи ліків і пов'язаної з цим діяльності. Зрозуміло, що це завдання для окремого ряду досліджень, але найкращим варіантом представлення цієї предметної області є аналіз її економічного впливу та показників.

За останні роки світовий фармацевтичний ринок пережив значне зростання. У 2023 році загальний світовий фармацевтичний ринок оцінювався приблизно в 1,6 трильйона доларів США. Це збільшення на понад 100 мільярдів доларів порівняно з 2022 роком [1].

У всьому світі Сполучені Штати є провідним ринком фармацевтичних препаратів. У той же час, країна відома активністю в дослідженнях та інноваціях, а також високою вартістю лікарських засобів. За ними йдуть країни Європи та ринки, що розвиваються (країни із середнім і низьким рівнем доходу): Бразилія, Індія, Україна, Колумбія та Єгипет. Наприклад, на Латинську Америку припадає найменша частка доходів світового фармацевтичного ринку, в той час як у Європі та Україні зокрема ця частка постійно росте.

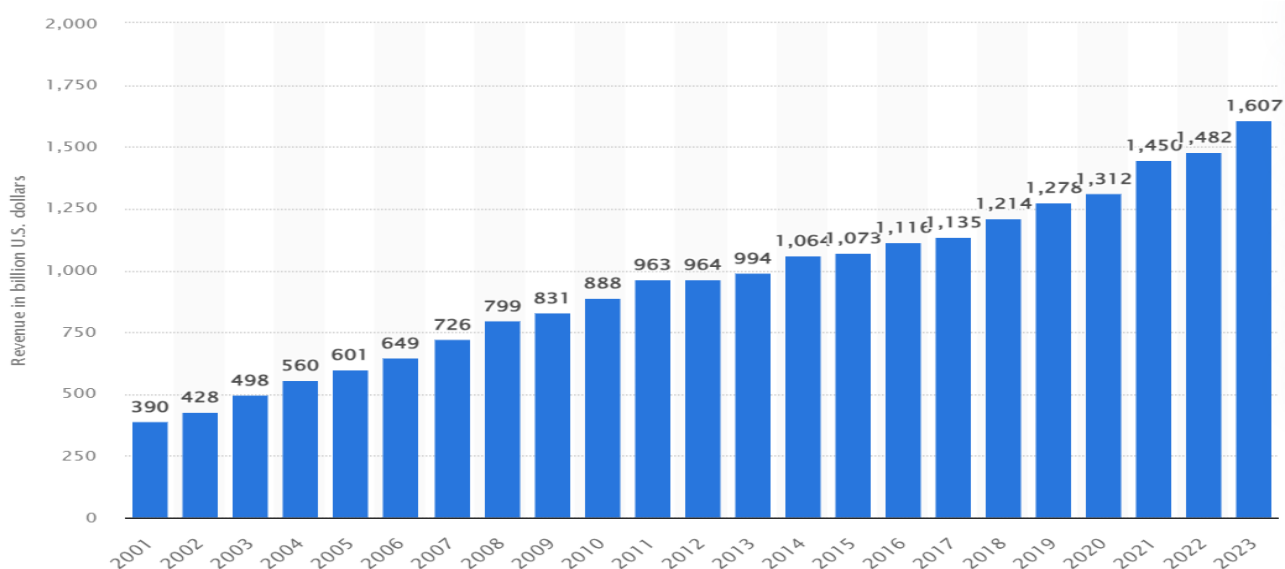


Рисунок 1.1 – Фармацевтичний ринок: світовий огляд за 2001-2023 роки

Глобальні лідери фармацевтичного ринку включають препарати такі як Humira, Eliquis та Revlimid. Основним напрямком росту в терапевтичних сферах є онкологія, яка, як очікується, продемонструє значне зростання у найближчі роки. Аутоімунні захворювання та діабет також є ключовими сегментами витрат на лікарські засоби. У період пандемії COVID-19, вакцина Comirnaty стала найприбутковішим продуктом на світовому фармацевтичному ринку [2].

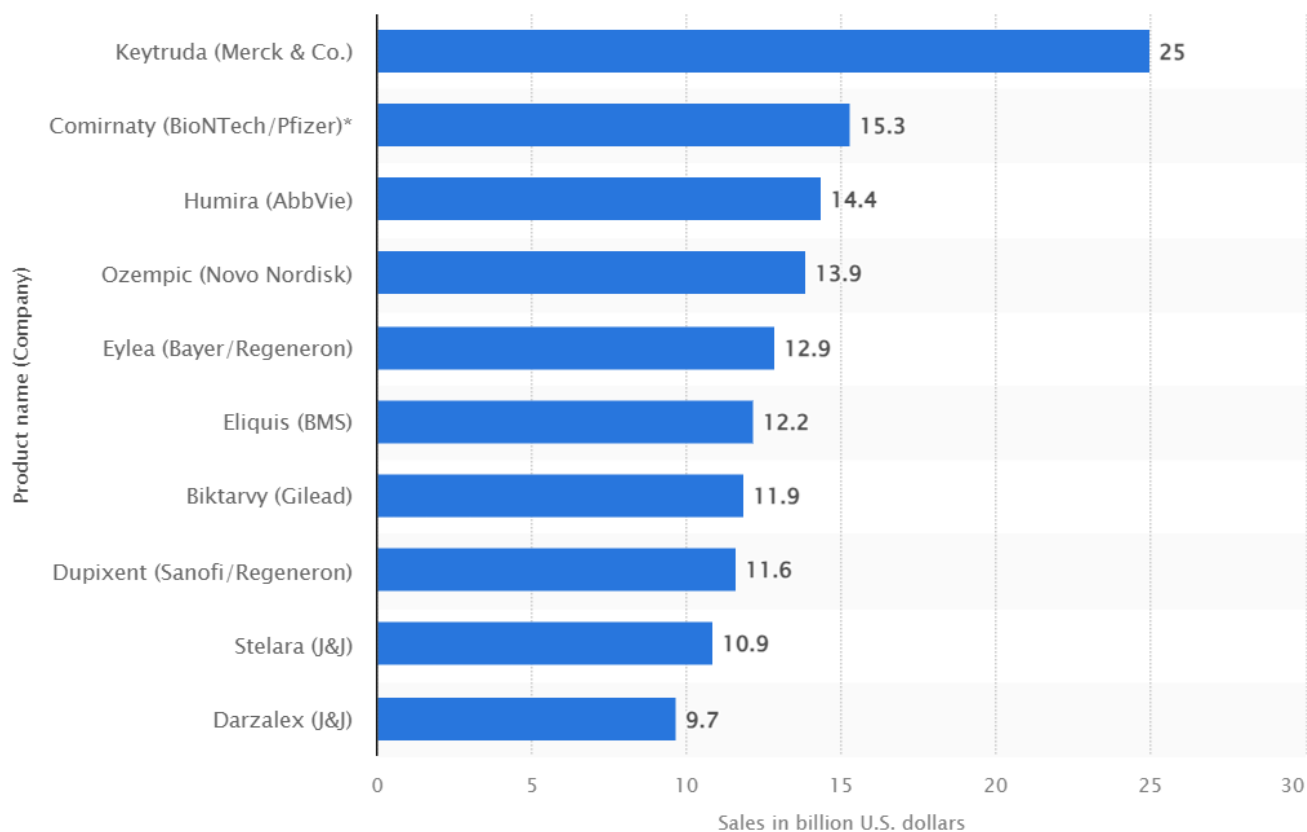


Рисунок 1.2 – Ключові фармацевтичні продукти за продажами у світі

Як і за два роки до цього, вакцина проти COVID-19 Comirnaty не зберегла своє місце як найкращий фармацевтичний продукт у світі за доходами. Ліки Merck від раку Keytruda є новим найбільш продаваним препаратом у світі, прибуток якого становить близько 25 мільярдів доларів США. Humira посіла третє місце в рейтингу препаратів у 2023 році із загальним доходом у 14,4 мільярда доларів.

Humira використовується в основному для лікування ревматоїдного та псоріатичного артриту. Однак, як і багато інших найкращих препаратів, діапазон захворювань, які він може використовувати для лікування, набагато ширший. Так,

препарат також використовують для лікування хвороби Крона, анкілозуючого спондиліту, виразкового коліту, псоріазу, увеїту та деяких інших аутоімунних захворювань. Humira вводиться підшкірно шляхом ін'єкції. Вперше його схвалили в 2002 році в США.

Популярні препарати часто мають великий вплив на виробника. У деяких випадках фармацевтичні компанії дуже залежать від своїх продажів. У випадку Humira компанія AbbVie отримує понад 36 відсотків свого загального доходу від цього препарату та понад 53 відсотки від трьох найпопулярніших продуктів. Компанії, які залежать від невеликої кількості популярних препаратів, можуть опинитися під загрозою значної втрати прибутку. Біосиміляри (аналоги) Хуміри (адалімумаб) уже доступні, і в найближчі роки вони займатимуть все більшу частину ринку. Проте Humira перевищила позначку в 20 мільярдів доларів доходу всупереч деяким прогнозам, зробленим багато років тому.

Тому розробки у сфері рекомендацій ліків є актуальними і важливими одночасно як з економічного боку так і з точки зору покращення рівня людського життя. Таким чином ця предметна області важлива і складна не тільки через саму її актуальність, але й через складність обумовлену ризиками врахування реакції пацієнта, дослідження ефективності препаратів.

1.2 Актуальність проблеми підбору лікарських препаратів

Наприклад з дослідження, що аналізує недовіру до медичних препаратів за групами населення рівень недовіри варіюється від 26% до 43% [4].

Variable	n	Prevalence	χ^2 (df= 1)	P-Value
Race and ethnicity				
White and European American	456	35%	3.21	.07
All minority	386	29%	3.21	.07
Hispanic and Latino	285	26%	8.45	<.01
Black and African American	67	43%	4.16	.04
Asian and Pacific Islander	66	39%	1.74	.19
Native America and Alaska Native	51	29%	0.18	.67
Middle Eastern and North African	11	36%	0.83	.36
Multiracial	100	42%	6.15	.01
Gender				
Male	171	35%	1.01	.32
Female	622	29%	4.83	.03
Transgender and nonbinary or nonidentifying	22	64%	10.48	<.01

Рисунок 1.3 – Результати дослідження недовіри до медичних препаратів за групами населення

Відповідно до вказаного дослідження серед всієї вибірки існує залежність, що більшість людей не мають довіри до рекомендацій лікарів, а ще менше до ліків, що відображено на рисунку 1.4.

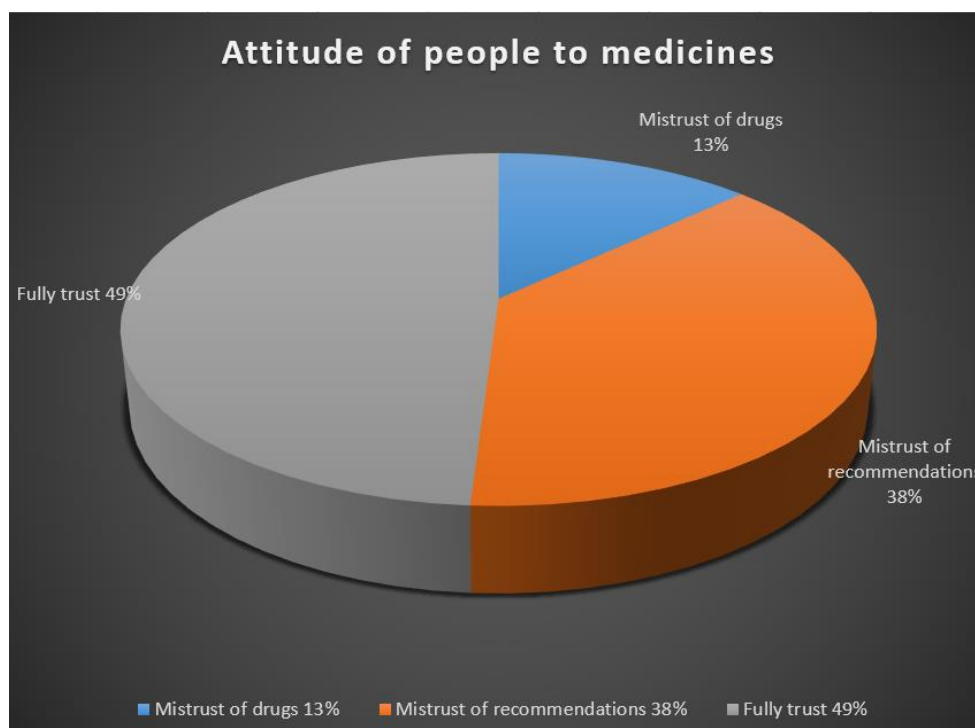


Рисунок 1.4 – Кругова діаграма груп населення до довіри до ліків (недовіра до самих ліків 13%, недовіра до рекомендацій лікарів 38%, повна довіра 49%)

Ця статистика підкреслює існування суттєвих занепокоєнь щодо точності діагностування та призначення медичних препаратів. Завданням медичних працівників, у тому числі лікарів та фармацевтів, є не тільки діагностування, але й відповідальне призначення лікарських засобів. Зі збільшенням кількості людей із проблемами здоров'я зростає й навантаження на них.

Медичні помилки можуть виникнути в секторі охорони здоров'я в результаті того, що медичні працівники призначають ліки на основі неадекватної інформації, пов'язаної з історією пацієнта та побічними ефектами ліків. Таким чином, це дослідження має на меті запропонувати систему рекомендацій лікарських засобів, щоб допомогти постачальникам медичних послуг у прийнятті рішень при призначенні ліків пацієнтам залежно від їхніх діагнозів. Настрої відгуків про ліки аналізуються, щоб визначити ефективність препарату серед споживачів.

Рекомендації щодо лікарських засобів у медичній сфері є складним реальним завданням вибору найкращих ліків для пацієнтів на основі їхніх історій хвороби та симптомів. Тому основним призначенням цієї роботи є формування рекомендацій ліків для пацієнтів. Важливість цього завдання підсилюється постійним накопиченням та оновленням баз даних про лікарські препарати, які є у відкритому доступі.

Однак, першочерговою проблемою, на яку слід звернути увагу, є саме недовіра споживачів. Різниця у списку дозволених ліків між країнами та відмінності у регулюванні обігу ліків ведуть до складнощів у класифікації медичних препаратів національними регуляторами. Значна частина інновацій на фармацевтичному ринку відбувається в Європі, хоча багато країн орієнтуються на стандарти, прийняті в США.

Страхи щодо неефективності ліків та помилок фахівців, а також високі ціни на медичні препарати, додатково підживлюють недовіру серед населення. Вирішенням цієї проблеми стало застосування аналогів дорогих ліків. Таким чином, у 2002 році в рамках Підкомітету з медичної хімії Міжнародного союзу теоретичної та прикладної хімії (IUPAC) була сформована робоча група з розробки ліків на основі аналогів [5].

Відповідно до терміну «аналог», що з'явився у природознавстві у 1791 році, лікарські аналоги відображають хімічну та фармакологічну подібність до оригінальних молекул.

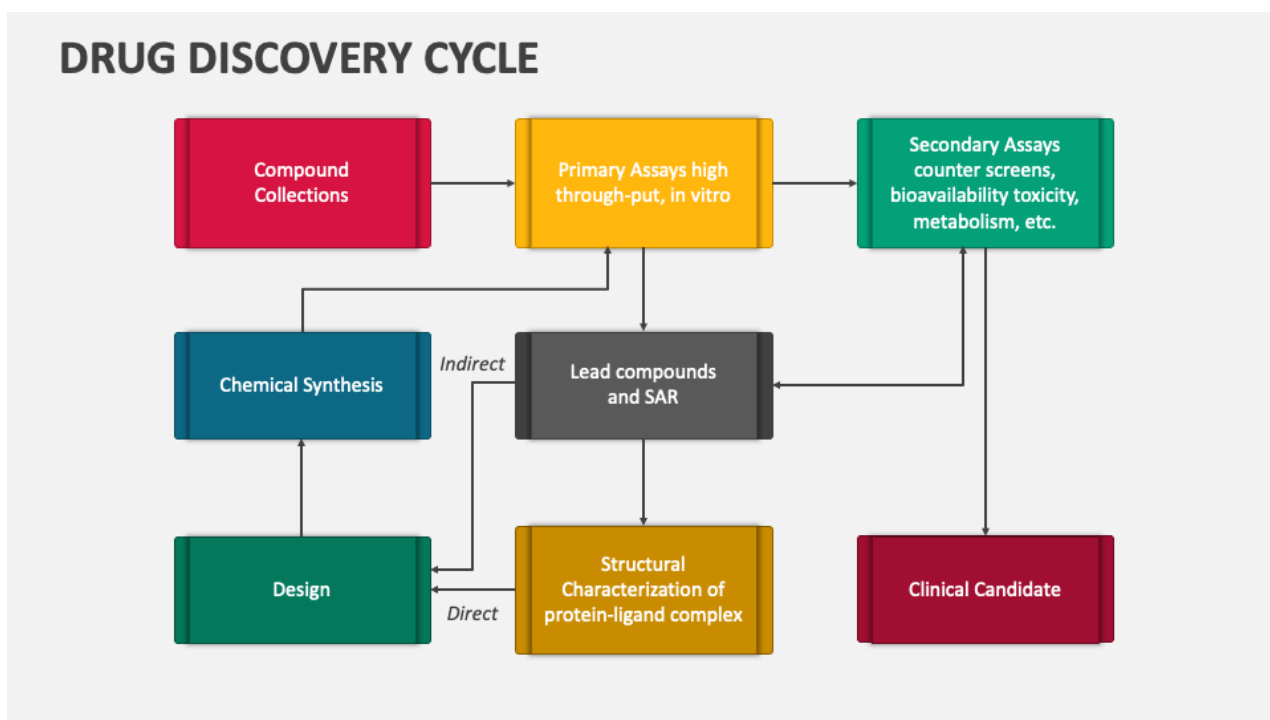


Рисунок 1.5 – Цикл винаходу і розробки лікарських препаратів

Відповідно до цього існує як прямиий, так і непрямий підходи до створення медичних препаратів. Фармацевтичні компанії не лише вводять на ринок нові лікарські засоби, а й розробляють їх аналоги. Цей процес часто включає клінічні дослідження, які дозволяють отримувати дані про ефективність та безпеку ліків. Важливою особливістю лікарських препаратів є їхня активна речовина, яка визначає терапевтичний ефект.

Тому важливим для цього є наявність у медичних інституціях системи підбору ліків. Ключовою проблемою є ефективність та інтелектуальність методів формування рекомендацій. Для дослідження методу підбору ліків важливо мати чітке уявлення про всі характеристики ліків, які можуть впливати на їх ефективність та безпеку для пацієнтів.

Часто у засобах масової інформації та у дослідженнях порушувалися різноманітні питання щодо медичної діагностики та призначення ліків, які

потребують ретельного розгляду. Звісно, медичні працівники відповідають за діагностику пацієнтів і призначення ліків. Однак, виклик полягає в нестачі медичних працівників, особливо з огляду на зростаючу кількість осіб із проблемами здоров'я [6-8].

Коли пацієнтам призначають лікарські засоби, які не відповідають їхньому стану, існує високий ризик медичної помилки, що в крайніх випадках може загрожувати їхньому здоров'ю або навіть життю. Так щороку через медичні помилки в лікарнях реєструється близько 99 тисяч смертей [9]. Ці помилки часто виникають, коли медичні працівники призначають ліки без належної перевірки історії пацієнта та, особливо, побічних ефектів препарату [10]. Часто це виникає внаслідок того, що досвід медичного працівника може бути обмеженим, оскільки він може не знати про всі численні види ліків на ринку, що призводить до медичних помилок. Тому, система рекомендацій щодо ліків мала б вирішальне значення для вирішення цієї проблеми.

Тобто системи рекомендацій – це інформаційні системи, які передбачають уподобання користувачів і надають персоналізовані та суб'єктивні рекомендації щодо продукту чи послуги, що важливо при здійсненні рішень щодо здоров'я [11,12].

Ці системи використовуються в багатьох галузях, зокрема в електронній комерції, щоб надати клієнтам персоналізований досвід. Медична промисловість також може використовувати такий тип рекомендаційної системи, зокрема для рекомендації ліків [13].

Хоча пацієнти не мають змоги самостійно вибрати лікарські засоби без консультації з лікарем, системи рекомендацій стають важливим інструментом для медичних працівників. Вони допомагають спростити процес вибору найкращих ліків, підвищуючи точність медичних призначень. Інтеграція технологій, таких як машинне навчання та аналіз даних, збільшує ефективність та надійність цих систем.

Більшість пацієнтів придбає ліки за порадою лікаря, часто не знаючи про інші доступні на ринку препарати, які можуть бути ефективнішими. Тому вивчення

клінічних досліджень та описів лікарських препаратів є надзвичайно важливим для глибшого розуміння їх ефективності. Такий аналіз дозволяє медичним працівникам краще оцінювати загальну ефективність ліків для різних пацієнтів, забезпечуючи обґрунтовані та точні медичні рекомендації.

І останнє, але не менш важливе: проблеми холодного старту та зростання переваг клієнтів є двома ключовими проблемами, які необхідно розглянути при обговоренні систем рекомендацій. Щоб вирішити проблему холодного старту та зростання переваг клієнтів, знадобиться відповідний алгоритм. Ось де машинне навчання може бути корисним. Щоб медичні працівники мали персоналізований досвід, можна використовувати гібридні методи фільтрації на основі вмісту та спільні методи фільтрації. Ці методи дозволять медичним працівникам краще рекомендувати препарати для певного захворювання. Але ключовою проблемою виступає формування подібних рекомендацій, а конкретніше дослідження, огляд і створення відповідних методів. Так як в подальшому ці методи стають основою для подібного роду систем.

Для цього проведемо декомпозицію задачі підбору ліків, оскільки вона є складним процесом, який включає кілька етапів, що взаємодіють між собою для досягнення точних і персоналізованих рекомендацій для пацієнтів. Декомпозиція цієї задачі передбачає розбиття її на більш дрібні, керовані підзадачі. Основними підзадачами головного процесу підбору ліків є: збір та аналіз даних, розробка моделей, формування рекомендацій та оцінка їхньої ефективності. З огляду на те, що більшість підзадач і сама задача підбору ліків чітко регламентовані відповідними правилами та законами в різних країнах по різних критеріям, то отримаємо, що найскладнішим і найбільш актуальним є певно задача формування рекомендацій. Фармацевтичні компанії, які активно залучені у світовий ринок, підкреслюють важливість універсального підходу до розуміння і лікування розповсюджених захворювань, що веде до інтеграції передових технологій, як-от CDSS, у медичні процеси.

CDSS являють собою зміну парадигми сучасної охорони здоров'я. Вони покликані допомагати лікарям в складних процесах прийняття рішень,

забезпечуючи підтримку на основі накопиченого досвіду та доступних медичних даних. З моменту їхнього початкового впровадження у 1980-х роках, CDSS значно еволюціонували та інтегрувались у багато аспектів медичного обслуговування.

Сучасні CDSS тісно пов'язані з електронними медичними записами та іншими комп'ютеризованими клінічними робочими процесами, що стало можливим завдяки широкому впровадженню цифрових технологій у медицині. Ці системи значно покращили ефективність медичного обслуговування, забезпечивши більшу точність у діагностиці та лікуванні, а також сприяючи оптимізації лікувальних процесів.

Протягом останніх десятиліть було задокументовано численні випадки успішного використання CDSS, які сприяли покращенню досвіду пацієнтів та лікування. Однак, існують і випадки, коли використання CDSS призвело до помилок або невдач, підкреслюючи наявність певних ризиків, пов'язаних із залежністю від автоматизованих систем [14]. Тому важливе подальше вивчення та розробка цих систем з метою мінімізації ризиків і максимізації потенціалу.

1.3 Огляд даних про лікарські препарати

В рамках системи формування рекомендацій, лікарські препарати аналізуються з урахуванням їхнього зв'язку із захворюваннями, які діагностує лікар-спеціаліст. Основною сутністю у процесі рекомендацій є самі медичні препарати, кожен з яких описується через ряд визначених атрибутів. Ці атрибути допомагають зрозуміти не тільки фармакологічні характеристики препарату, але й потенційні побічні ефекти, можливості взаємодії з іншими лікарськими засобами, а також наявність альтернативних варіантів.

Було сформовано оптимальний набір атрибутів, що містить таку інформацію для кожного препарату:

- назва препарату: комерційна або загальноприйнята;
- побічні реакції та ефекти: перелік потенційних небажаних реакцій при застосуванні;

- лікарські взаємодії: опис можливих взаємодій з іншими препаратами, що може впливати на ефективність або безпеку;
- клас ліків: функціональна класифікація, наприклад терапевтичний клас, що вказує на призначення препарату;
- ліки-замінники: альтернативи з аналогічним складом від різних виробників для забезпечення вибору у випадку необхідності;
- активні інгредієнти: основні діючі речовини, відповідальні за лікувальний ефект;
- клас дії (Action Class): класифікація за механізмом дії (напр., блокатори H₂-рецепторів), що допомагає розуміти, як ліки впливають на організм;
- хімічний клас: класифікує препарати згідно з їх хімічним складом;
- формування звички (Habit Forming): інформація про потенційну залежність від препарату;
- застосування (Usage): стани або умови, для яких препарат зазвичай призначається.

Зазначені атрибути надають можливість створити комплексну модель для підбору ліків, яка не лише сприятиме ефективному лікуванню, але й мінімізує ризики, пов'язані з побічними ефектами та неправильною взаємодією препаратів, а також забезпечуватиме можливість вибору альтернативних препаратів.

1.4 Схеми процесів підбору лікарських препаратів

Фактично потрібно вказати схему лікарів роботи з подібними рекомендаційними системами. Тому, так як вона представляє собою інформаційну систему, потрібно провести огляд процесу цієї роботи.

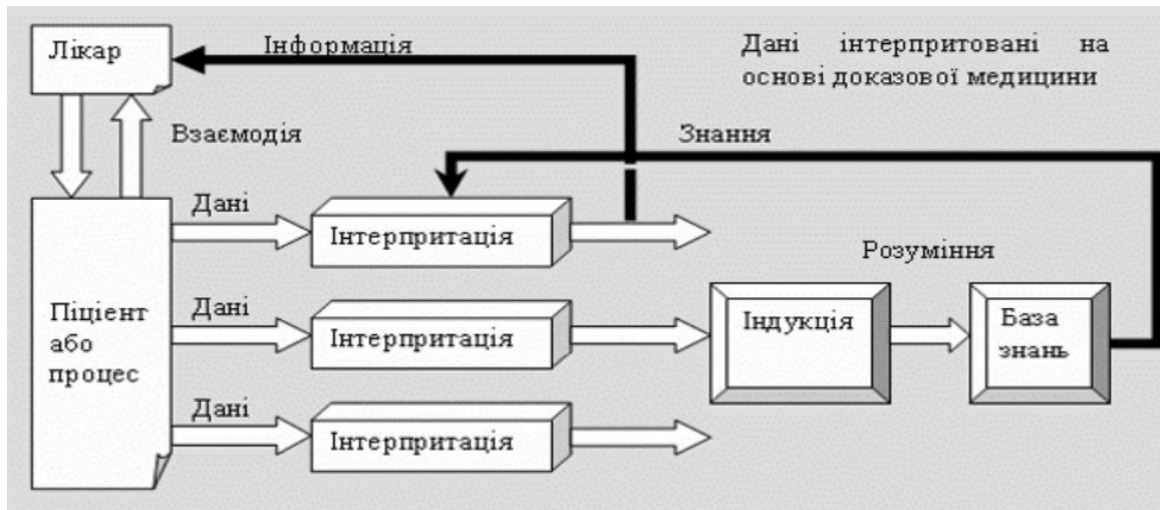


Рисунок 1.6 – Схема роботи лікаря з інформаційною системою рекомендацій

Практично, система вимагає від лікаря розуміння основних відмінностей, ризикових факторів, а також використання передових технологій для аналізу даних про здоров'я відносно пацієнта. Водночас існує потреба в інтерпретації цих даних і після розуміння, тобто поставленого діагнозу, потрібно використовувати систему рекомендацій ліків, основною задачею якої є формування самих рекомендацій відповідно до потреб серед існуючих та доступних ліків. Тоді ця робота включає:

- інтерпретацію медичних показників пацієнта, його симптомів: визначення ключових скарг пацієнта, точне формулювання його симптомів, проведення аналізів для постановки діагнозу і розуміння стану його організму;
- дослідження факторів ризику: аналіз існуючої історії хвороб пацієнта, його хронічних захворювань, проведення аналізів з алергії на певні речовини;
- моніторинг стану пацієнта: регулярна перевірка показників та даних пацієнта для уточнення реакцій на ліки та відповідність поставленого діагнозу;
- підбір лікарських препаратів: використання інформаційної системи для вибору найбільш ефективних та безпечних ліків, з урахуванням діагнозу, ризиків та індивідуальних особливостей пацієнта.

Останній елемент цієї роботи, а саме підбір ліків і є ключовою частиною роботи лікаря не тільки з сучасними інформаційними системами, але й загалом. Ефективність цього кроку безпосередньо впливає на успіх лікування та покращення загального стану пацієнта.

1.5 Постановка задачі формування рекомендацій для СПЛП

Метою дослідження є розробка методу формування рекомендацій, який покликаний забезпечити вищу точність та персоналізацію вибору медикаментів для кожного пацієнта, враховуючи їх унікальні особливості. Метод повинен класифікувати ліки за різними групами ризику і надати рекомендації щодо лікування на основі їх медичних показників, таких як рівень глюкози в крові, індекс маси тіла, а також інших біохімічних та демографічних характеристик.

Об'єктом дослідження є процеси аналізу та вибору лікарських засобів з використанням інформаційних технологій у фармацевтиці.

Для формування рекомендацій система використовує такі дані:

- клінічні показники: як активні речовини, визначені якості, побічні ефекти та інші атрибути ліків;
- визначений запит лікаря: на основі його запиту проводиться фільтрація лікарських засобів для підбору рекомендацій.

Вихідними даними цього методу будуть:

- список рекомендованих ліків з аналогами;
- позначення короткої інформації про дозування;
- перелік які протипоказання були враховані.

Задача полягає у розробці методу формування рекомендацій для системи лікарських препаратів (як у системах CDSS), що застосовуються медичними інституціями. Відповідно, метод має працювати з даними ліків та запитамі лікарів щодо вирішення проблеми підбору ліків. Метод повинен працювати з різноманітними типами вхідних даних і навчатися на них для точної класифікації лікарських препаратів.

Основними етапами вирішення цієї задачі є наступні:

- дослідження даних про існуючі системи підбору;
- вивчення та аналіз існуючих методів формування рекомендацій підбору ліків;
- розробка методу формування рекомендацій;

- створення реалізації методу;
- порівняння ефективності створеного методу;
- валідація створеного методу;
- аналіз отриманих результатів.

Важливими є етапи, що пов'язані зі створенням методу формування рекомендацій, тому для подальшої валідації додаються ще наступні підзавдання:

- збір та підготовка даних: знаходження реальних даних, їхня підготовка для обробки;
- аналіз даних: створення класифікації ліків для подальшого застосування у фільтрації;
- створення моделі методу;
- підбір алгоритмів та моделей для частин створеного методу;
- реалізація методу для проведення валідації.

Це дослідження відкриє шлях до розробки ефективного та точного методу, який послужить основою для системи рекомендацій щодо вибору лікарських препаратів. Метод має забезпечити точність у визначенні рекомендацій для лікарських засобів та сприяти оптимізації процесу вибору ліків, враховуючи індивідуальні потреби пацієнтів та час, необхідний лікарям для роботи.

2 ДОСЛІДЖЕННЯ ІСНУЮЧИХ МЕТОДІВ ФОРМУВАННЯ РЕКОМЕНДАЦІЙ

2.1 Огляд існуючих рішень для СПЛП

Рекомендаційні системи – це клас алгоритмів і програмного забезпечення, призначених для вибору та пропонування користувачам інформації, товарів, послуг або рішень, які, на думку системи, найбільше відповідають їхнім інтересам та потребам. Вони використовують різноманітні методи обробки даних та аналітики для аналізу історії поведінки користувача, їх переваг, а також взаємодій між користувачами та об'єктами системи.

У сучасному світі кількість генерованих даних зростає експоненціально. За даними International Data Corporation (IDC), обсяг світових даних досягне 175 зетабайт до 2025 року. У таких умовах рекомендаційні системи допомагають користувачам орієнтуватися у величезній кількості інформації, вибираючи найбільш релевантний контент або продукти [9].

Завдяки збору та аналізу даних про поведінку користувачів, рекомендаційні системи забезпечують високий рівень персоналізації, що підвищує задоволеність користувачів та їх взаємодію з сервісом. Amazon та Netflix є прикладами компаній, які використовують персоналізовані рекомендації для підвищення продажів та залученості аудиторії. За даними McKinsey, персоналізація може збільшити продажі до 15%.

В свою ж чергу, актуальність таких систем в Україні проявляється у широкому спектрі галузей, від електронної комерції до освітніх сервісів, забезпечуючи користувачам персоналізовані пропозиції та підвищуючи ефективність бізнес-процесів.

З огляду на швидкий розвиток фармацевтичного ринку, де щодня реєструються нові лікарські засоби, рішення про вибір відповідного препарату стає великим викликом для медичних працівників.

Використання систем рекомендацій у медицині важливе не тільки для підтримки лікарів у призначенні лікування. Різні хімічні складки та індивідуальні

реакції пацієнтів на лікарські засоби збільшують складність вибору, тому наявність надійних інструментів для оцінки ефективності та безпеки препаратів є критично важливою. Пацієнти не можуть самостійно вибирати медикаменти без консультацій, однак такі системи можуть допомагати їм в наданні інформації, яка сприятиме підвищенню освіченості.

Також значні потенціали відкривають алгоритми машинного навчання та аналізу даних, які покращують точність та надійність медичних рекомендацій. Завдяки обробці великих даних, медичні працівники можуть отримати доступ до деталізованих описів ефективності препаратів, враховуючи специфічні реакції різних пацієнтів. Це дозволяє краще зрозуміти, як ліки діятимуть в конкретних умовах, мінімізуючи ризики та підвищуючи шанси на успіх лікування [14].

Усвідомлення медичних працівників про новітні препарати та їх властивості є критично важливим, адже в умовах швидкого розвитку фармацевтики традиційні підходи швидко застарівають. Клінічні дослідження та аналізи допомагають медичним працівникам адаптуватися до змін, використовуючи найновіші дані для підвищення якості лікування.

Одним із методів, що використовується в рекомендаційних системах є колаборативна фільтрація, яка дозволяє моделювати та прогнозувати інтереси користувача на основі схожості інтересів інших користувачів з подібними уподобаннями. Цей метод широко застосовується в онлайн-ритейлі для рекомендацій продуктів на електронних платформах торгівлі, таких як Amazon або Netflix, де алгоритми аналізують історію покупок та переглядів користувачів, щоб рекомендувати нові товари або фільми.

У галузі електронної комерції рекомендаційні системи допомагають компаніям аналізувати попит і прогнозувати продажі, оптимізуючи управління запасами. Це забезпечує не лише ефективніше використання ресурсів, але й покращує задоволеність клієнтів за рахунок зменшення часу доставки.

Рекомендаційні системи використовуються також для зниження відтоку клієнтів, аналізуючи їх поведінку та ідентифікуючи ознаки зниження інтересу до

сервісу. Це дозволяє компаніям своєчасно реагувати, пропонуючи акції, знижки або інші ініціативи для збереження користувачів.

У глобальному економічному просторі, де конкуренція зростає з кожним днем, рекомендаційні системи надають компаніям важливу перевагу, допомагаючи їм виділитися завдяки унікальному та персоналізованому підходу до кожного користувача.

Інший популярний підхід – це контент-орієнтована фільтрація, яка використовує характеристики самого продукту або контенту для рекомендацій. Наприклад, якщо користувач слухає певний жанр музики, система може рекомендувати інші пісні з аналогічними музичними характеристиками.

Гібридні системи комбінують елементи колаборативної та контент-орієнтованої фільтрацій, використовуючи переваги обох методів для підвищення точності рекомендацій. Це може допомогти уникнути деяких проблем, таких як холодний старт або недостатність інформації про нових користувачів або продукти.

Також, важливим аспектом сучасних рекомендаційних систем є їх здатність адаптуватися до змінних умов і потреб користувачів у реальному часі, що робить їх незамінними у сферах, де швидкість і точність відіграють критичну роль.

Системи підбору лікарських засобів розроблені з метою покращення ефективності лікування та мінімізації ризиків для пацієнтів шляхом оптимізації процесу вибору та призначення медикаментів. Ці системи використовують алгоритми аналізу великих обсягів даних для ідентифікації найбільш відповідних препаратів на основі інформації про їх властивості, ефективність, можливі побічні ефекти та взаємодії з іншими ліками. Впровадження таких систем має наступні переваги:

- підвищення безпеки пацієнтів: зниження кількості медичних помилок, пов'язаних з призначенням лікарських засобів;
- оптимізація лікувальних режимів: можливість вибору оптимальних комбінацій препаратів для конкретного пацієнта;
- економія часу медичних працівників: автоматизація частини рутинних завдань, пов'язаних з вибором медичних препаратів.

Інтелектуальні системи підбору лікарських засобів вже демонструють свої переваги у медичній галузі, але потенціал штучного інтелекту (ШІ) у сфері охорони здоров'я далеко не вичерпаний. Державні та приватні медичні установи вже сьогодні можуть впроваджувати та використовувати ШІ і таким чином сприяти переходу від наукових розробок до реального застосування. У разі успішного впровадження ШІ може знизити навантаження на медичних працівників і підвищити якість роботи, яка виконується за рахунок зменшення кількості помилок і підвищення точності.

Часто впровадження штучного інтелекту або методів машинного навчання використовується в цілях розширення людських робочих процесів або автоматизації процесу надання медичних послуг як діагностування, підбір ліків, моніторинг стану пацієнта. Це обумовлено тим, що великі обсяги різноманітної інформації не можуть бути завжди ефективно оброблені статистичними методами або іншими методами лінійного типу та важко обробляються людьми, а тому, зі збільшенням обсягів інформації та даних, виникає збільшення помилок. Конкретні програмні компоненти будуть включати автоматизацію приладів, виявлення помилок, прогнозування, інтерпретацію результатів, використання тестів, тому впровадження створеної моделі як ШІ в процеси медичних інституцій буде мати вигляд відображений на схемі [15, 16].

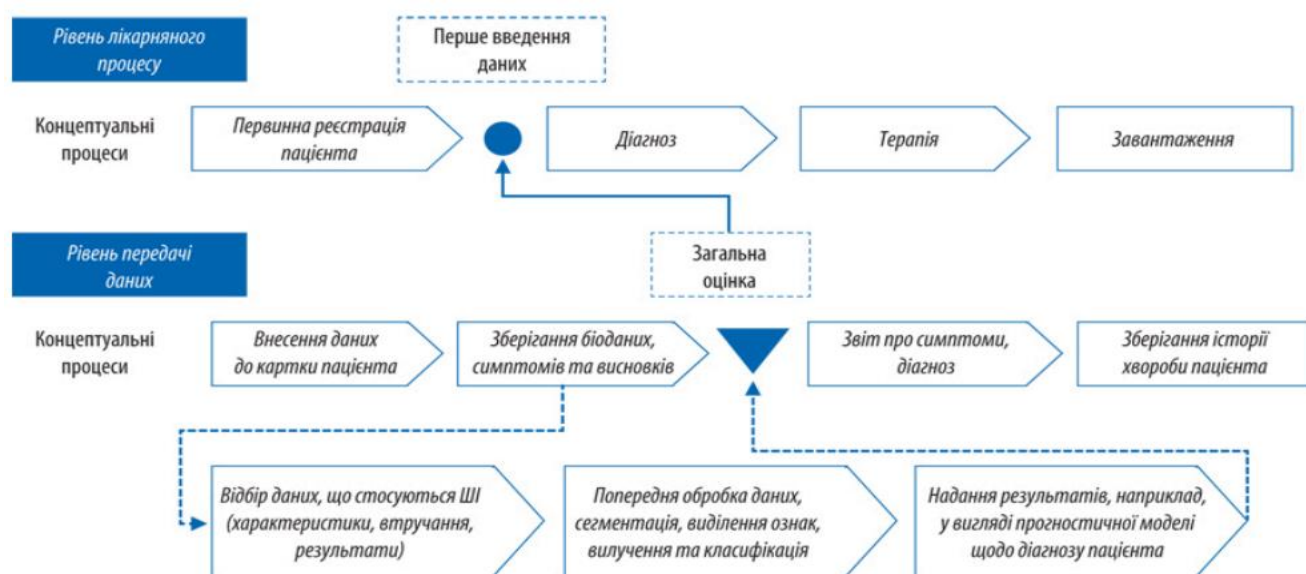


Рисунок 2.1 – Схема впровадження ШІ в процеси медичних послуг

Системи підтримки клінічних рішень (CDSS) стали важливою частиною медичної практики, особливо після їхнього схвалення у законодавчих актах США, спрямованих на охорону здоров'я та медичну допомогу, що стимулюють інтеграцію CDSS у електронні медичні записи (EHR). Наприклад, у 2013 році близько 41% лікарень у США з EHR мали CDSS, а до 2017 року ця цифра стабілізувалася на рівні 40,2% з розширеними можливостями CDSS. Міжнародний досвід також був обнадійливим: у Канаді, завдяки значній підтримці уряду та корпорації Infoway, понад 62% медичних практик впровадили CDSS у 2013 році. Країни, такі як Велика Британія, витратили значні суми на ІТ-інвестиції в охорону здоров'я, досягнувши приблизно 20 мільярдів євро в 2010 році, що підкреслює їхнє лідерство в цій галузі. Данія, Естонія та Австралія були серед країн, яким вдалося запровадити національні медичні записи [17-19]. В Україні також з'являються системи типу EStock, які інтегрують подібні технології для покращення медичного обслуговування.

Обсяг функцій, які надає CDSS, є дуже широким. До основних можна віднести:

- попередження та нагадування: CDSS може автоматично нагадувати медичним працівникам про необхідність проведення певних процедур або перевірок, сприяючи дотриманню клінічних рекомендацій та стандартів догляду;
- діагностична підтримка: системи можуть аналізувати медичні дані та запропонувати можливі діагнози на основі симптомів пацієнта та історії хвороби, що зменшує ймовірність діагностичних помилок;
- підбір лікарських засобів: CDSS допомагає вибрати найефективніші та безпечні лікарські засоби на основі індивідуальних особливостей пацієнта, таких як поточний стан здоров'я, потенційні взаємодії ліків та алергії;
- оцінка ризиків та протоколи лікування: CDSS може аналізувати ризики, пов'язані з різними лікувальними стратегіями, та рекомендувати оптимальні протоколи лікування;

- звіти та аналізи: системи здатні генерувати детальні звіти про стан пацієнтів та ефективність використання медичних ресурсів, що сприяє кращому плануванню та управлінню в медичних установах.

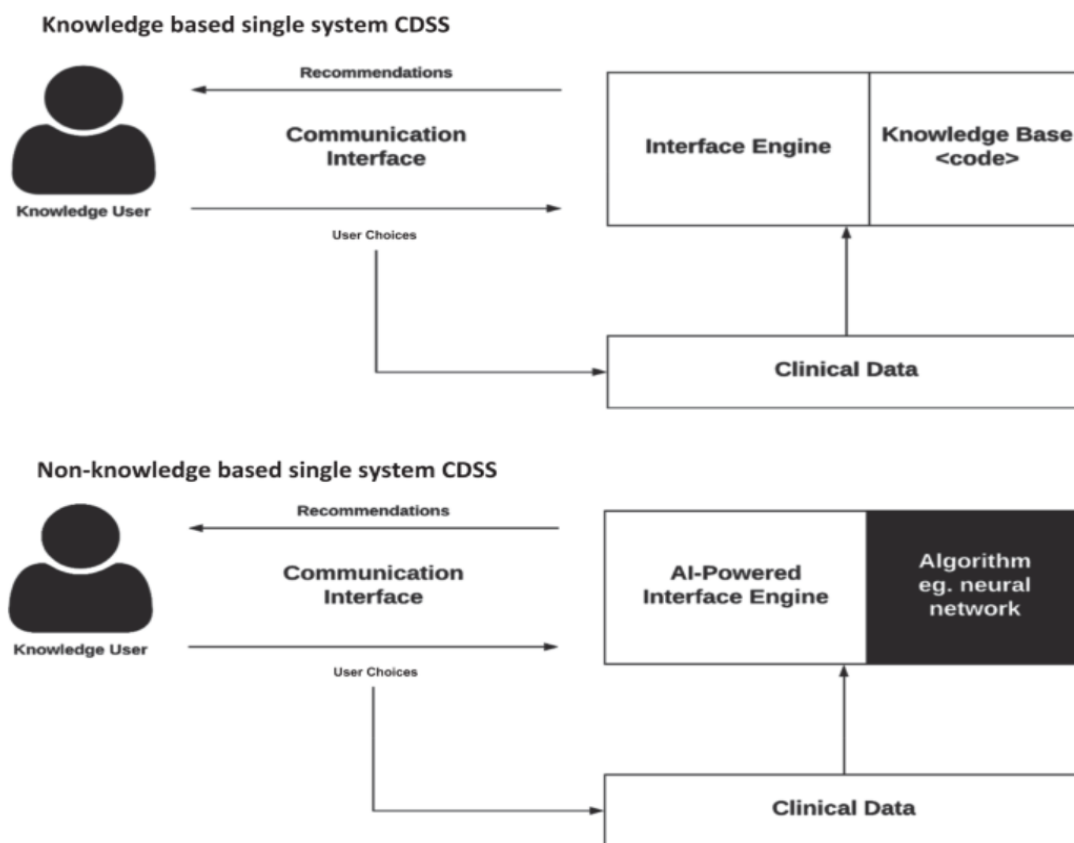


Рисунок 2.2 – Діаграма взаємодій з різними типами CDSS

На основі порівняння, наведеного на рисунку 2.2, головні взаємодії у першому випадку складаються з бази знань, тобто правил, запрограмованих на основі знань в системі, алгоритму, що використовується для моделювання рішення не на основі знань, а також доступних даних. У другому випадку, механізм висновку бере запрограмовані або визначені штучним інтелектом правила та структури даних і застосовує їх до клінічних даних пацієнта для створення виходу або дії, яка представлена кінцевому користувачеві (наприклад, лікарю) через механізм зв'язку (наприклад, веб-сайт, додаток або зовнішній інтерфейс EHR), за допомогою якого кінцевий користувач взаємодіє з системою [19].

Наприклад, часто використовується вебсистема рекомендацій щодо ліків. Такі системи побудовані на трирівневій архітектурі. Двома ключовими

користувачами є медичні працівники та адміністратори. Загалом для створення використовуються фреймворки JavaScript як React JS і відповідні технології бібліотек як Material UI. Так компонентна архітектура React і віртуальний DOM роблять його ідеальним для розробки інтерфейсу. Бібліотека Material UI Library пропонує готові компоненти, які відповідають інструкціям щодо матеріального дизайну, спрощуючи розробку UI, усуваючи необхідність створювати компоненти з нуля. Інтеграція між React і Material UI є безперебійною, що дозволяє налаштовувати компоненти відповідно до конкретних вимог UI [20, 21].

Щоб виконати функції API веб-додатку та включити моделі машинного навчання, застосовують FastAPI. Він служить серверною реалізацією для всіх функцій системи. Крім того, базою даних, яка використовується для такого роду проекту, може бути Mongo DB, яка є базою даних NoSQL. Mongo DB зберігає дані у форматі BSON, який сумісний зі структурою даних про препарати, користувачів у системі. FastAPI діє як посередник між інтерфейсом, моделями машинного навчання та базою даних. Коли користувачі запитують доступ до даних, вони роблять запити API. Подібним чином, для використання моделей машинного навчання запити API надсилаються до FastAPI, передаючи необхідні параметри. Потім серверна частина отримує результати моделі та повертає їх на зовнішню частину для відображення користувачеві [22-24].

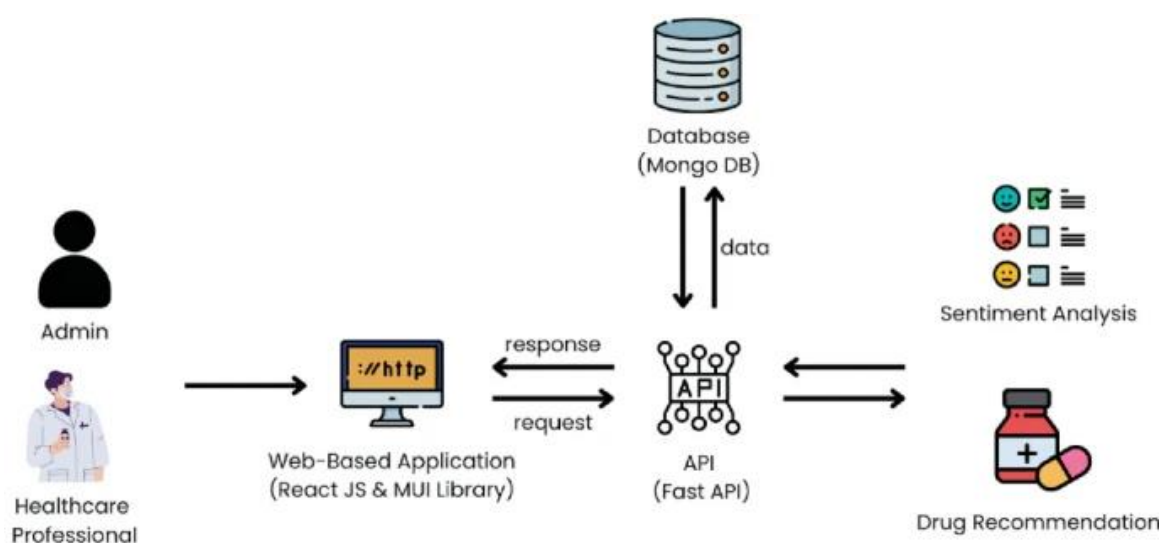


Рисунок 2.3 – Схема типової системи рекомендацій ліків

Подібні системи складаються з чотирьох модулів: модуля керування користувачами, модуля аналізу лікарських засобів, модуля рекомендацій і модуля управління лікарськими засобами. Цільовими клієнтами таких систем є медичні працівники в лікарнях. Вони можуть використовувати цю систему, щоб знайти найкращі ліки для певного діагнозу та рекомендувати їх своїм пацієнтам. Адміністратори керують загальними модулями системи. Медичні працівники та адміністратори, які використовують цю систему, повинні бути зареєстровані.

Ключову роль грають модуль аналітики ліків та модуль рекомендацій, що реалізують засобами Big Data та машинного навчання формування відповідних рекомендацій.

Класична блок-схема (рис. 2.4) такої системи акцентується на взаємодії елементів системи між собою. Її зручно представити як загальну блок-схему системи для цього дослідження. Медичні працівники, як основні користувачі системи, мають спочатку зареєструватися, створивши новий обліковий запис. Після авторизації вони можуть виконувати пошук ліків за назвою чи захворюванням.

У результати пошуку будуть включені всі препарати, які відповідають критеріям пошуку, і вони будуть ранжовані за рівнем відповідності. Ліки з вищими показниками будуть класифікуватися першими. Користувачі можуть перейти на сторінку конкретного препарату у результатах пошуку для отримання детальної інформації про нього, результати аналізу настрою та рекомендовані препарати. Завершивши аналіз відгуків, система надасть зведену інформацію, включаючи загальну статистику використання цього препарату користувачами.

Крім того, на основі вмісту модель фільтрації системи може ідентифікувати препарати, схожі до того, який переглядають або рекомендують користувачу. На сторінці пошуку користувачам пропонуються альтернативні рекомендації, і вони можуть перейти до функції порівняння препаратів. У цій частині системи можна вибрати два будь-які препарати, які доступні в системі, і ознайомитись з результатами їх порівняння.

Користувачі системи мають можливість відвідувати сторінку зі списком бажань, а також сторінку форуму, де можна створювати, редагувати та видаляти дописи, переглядати теми від інших користувачів та залишати коментарі. Ці функції спрямовані на поліпшення взаємодії та комунікації між медичними працівниками. Крім того, існує сторінка для адміністраторів, на якій вони можуть додавати або видаляти інформацію про лікарські препарати в системі.

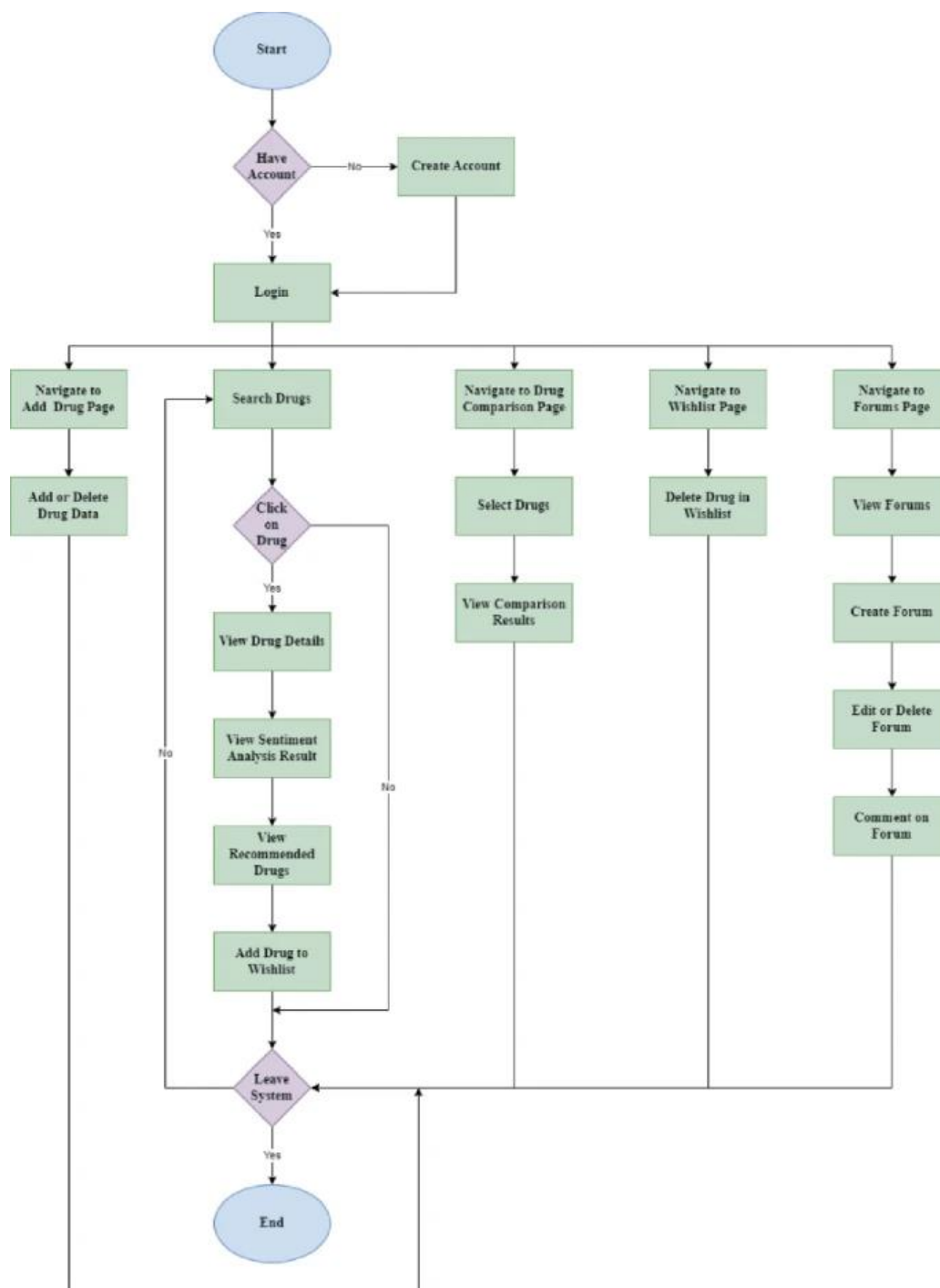


Рисунок 2.4 – Блок-схема системи рекомендацій ліків

2.2 Системний аналіз задачі формування рекомендацій для СПЛП

Так як рекомендаційні системи значно зменшують складність вибору ліків, вони в першу чергу базуються на задачі вибору серед альтернатив та методах фільтрації альтернатив, тому потрібно звернутися до теорії прийняття рішень.

Теорія прийняття рішень – міждисциплінарна галузь, що поєднує знання з математичної статистики, економіки та комп'ютерних наук та має елементи кібернетики. Відповідно, у випадку рекомендаційних систем ліків особою, що приймає рішення, є лікар. Рішення – результат інтелектуальної діяльності людини, що призводить до певного висновку або до необхідних дій. Рішення – той пункт, у якому робиться вибір між альтернативними та, як правило, конкуруючими можливостями [25].

Лікарі часто стикаються з обмеженнями у знаннях про нові та оновлені лікарські препарати через швидке впровадження новинок і значні оновлення у клінічних дослідженнях. Поряд з цим, зростання щорічних витрат на дослідження та розвиток (R&D) і маркетинг у фармацевтичних компаніях сприяє підвищенню якості ліків. Однак, це також може створювати ілюзію щодо ефективності рекламованих препаратів, що не завжди корелює з раціональним вибором, відомим з теорії прийняття рішень [26].

Це підводить нас до необхідності детальнішого розгляду проблеми в контексті прийняття рішень для формування рекомендацій щодо підбору лікарських препаратів. Різні автори з різним ступенем деталізації розглядають послідовність етапів, але загалом зберігається наступна послідовність дій:

- 1) виявлення проблемної ситуації та постановка задачі прийняття рішення;
- 2) формулювання поняття якості рішення та його структуризація до рівня критеріїв;
- 3) опис характеристик зовнішнього середовища, прогнозування можливих результатів дій ППР із подальшим виявленням або конструюванням альтернативних варіантів рішень;

4) оцінювання якості варіантів рішень, порівняння їх між собою та вибір одного чи декількох найвідповідніших меті;

5) аналіз рішень, опрацювання плану реалізації та впровадження рішення.

Відповідно було виявлено і описано проблему, сформовано постановку задачі, структуровано мету, формування альтернатив і розуміння факторів, станів, формування метрик та критеріїв для вибору та використання алгоритму для вибору найкращого варіанту.

У 2008 році К. Л. Го опублікував модель прийняття рішень DECIDE [27], яка складається з шести частин:

- 1) визначити проблему;
- 2) встановити критерії;
- 3) розглянути всі можливі альтернативи;
- 4) визначити найкращу альтернативу;
- 5) розробити та реалізувати план дій;
- 6) оцінити рішення та забезпечити моніторинг із зворотнім зв'язком при потребі.

Ця модель виявилася особливо корисною у медичній сфері, зокрема для прийняття рішень у сфері лікарських препаратів.

Проблема, яку вже визначено, полягає у необхідності розробки методу формування рекомендацій для системи підбору лікарських препаратів, який забезпечує високу точність, персоналізацію і мінімізацію ризиків. Це спонукає до формулювання відповідної мети дослідження.

Відомі фактори, що впливають на прийняття рішень, включають потребу в автоматизації, використанні машинного навчання чи штучного інтелекту, а також наявності атрибутів ліків, ризикових факторів, наявність діагнозу пацієнта та інші, що були детально описані раніше.

Великий обсяг інформації про ліки та їхні характеристики також істотно ускладнює процес створення специфічних рекомендацій для кожного препарату та їх адаптацію до індивідуальних потреб пацієнтів.

Можливо сформулювати математичний опис проблеми вибору для створення необхідного методу.

$$\exists x_d = \bigcap_i^{|C|} \tau_x(P_i) \quad (2.1)$$

В цій моделі x_d представляє собою метод, що відповідає меті роботи. Вибір методу відбувається через логічне об'єднання результатів пошуку методів, керованих набором факторів $|C|$, за кожним з яких існує критерій i . В той же час для кожного методу за кожною категорією P_i , застосовується функція вибору тау Боурбакі τ_x .

Ключовими параметрами було вибрано: протипоказання, активну речовину, хімічну групу, потенціал залежності, застосування при конкретних хворобах. Визначені метрики включають точність, ризик похибки як середньоквадратичне відхилення, застосування препаратів з протипоказами, а також швидкість роботи метода. Для цього вибору альтернатив було проаналізована відповідна фахова література.

Існує декілька підходів до формування рекомендацій для підбору лікарських засобів. Потрібно почати з визначення підходу, а вже потім за ним буде проведено аналіз самих методів.

Рандомізований підхід зосереджується на ймовірності отримання методу коректного для поставлених умов за певну кількість випадкових виборів. Звісно, на цьому підході було сформовано базис відкриття великої кількості самих ліків, але у цьому випадку потрібно створити метод відповідний вимогам безпеки, тому він малопридатний для застосування у цій роботі [7].

В свою чергу інтуїтивний підхід базується на нечіткій логіці, коли особа, що приймає рішення на основі свого досвіду та знань вибирає метод, але ігноруючи більшість достовірної аргументації. Цей підхід не відповідає академічній доброчесності та додатково треба зазначити, що усі медичні дослідження застосовані в цій роботі є аргументованими та доведеними науково в той час як орієнтуватися у відповідних методах для медичної предметної області у студента

комп'ютерних наук так ідеально не вийде і це фактично буде застосування рандомізованого підходу [8].

Науковий підхід, що базується на підтвердженій даних, формуванні гіпотез і доказах найкраще підійде для цієї роботи, але в цей же час є обмеження в часі на її створення та дослідження цього питання, тому пропонується використовувати системний аналіз існуючої проблеми з вибору методу формування рекомендацій для системи підбору ліків.

Спершу ми формуємо гіпотезу про те, що існує хоча б одна варіація відповідає переліку умов достатньо, що відображено у формулі 2.1. На початковому етапі проблема має слабо структурований характер, оскільки включає як якісні елементи, так і маловідомі, невизначені сторони, які можуть домінувати. Це означає, що є такі метрики, як точність діагностики, підбір протипоказаних ліків, швидкість формування рекомендацій та середньоквадратична помилка. Якщо ці метрики розмістити в таблиці разом із відповідними методами, замінивши тау-функції на відповідні функції максимуму та мінімуму відповідно до метрики, то проблема стає добре структурованою.

Таким чином, можна побудувати математичну модель методу та застосувати методи для знаходження оптимальної стратегії управління цілеспрямованими діями. Наприклад для точності отримаємо тау-функцію максимуму і представимо її для кількості методів від 1 до n:

$$\tau_{accuracy} = \max(accuracy_1, \dots, accuracy_i, \dots, accuracy_n)$$

Для середньоквадратичної помилки:

$$\tau_{MSE} = \min\left(\frac{1}{m_1} \sum_{t=1}^{m_1} (y_t - \underline{y}_t)^2, \dots, \frac{1}{m_n} \sum_{t=1}^{m_n} (y_t - \underline{y}_t)^2\right)$$

Для ймовірності призначення протипоказаних ліків використовується мінімум для відповідної тау-функції. Ці критерії будуть використовуватися для порівняльного аналізу існуючих методів шляхом побудови таблиці.

Як вже було зазначено раніше, існують системи, які використовують різні методи, що базуються на статистичних, лінійних та нелінійних моделях і методах. Тепер сформуємо правило для вибору методів. Зокрема, за кожною тау-функцією обираються три найкращі методи. Якщо ці методи також є найкращими за тау-функцією точності та тау-функцією протипоказань, вони отримують пріоритет у виборі.

3. СТВОРЕННЯ МЕТОДУ ФОРМУВАННЯ РЕКОМЕНДАЦІЙ ЛІКІВ

3.1 Схема створеного методу

Розглянемо системи підбору ліків, аналізуючи її складові елементи. Ключовим компонентом для таких систем є датасет ліків, парсинг та обробка цих даних, створення категоризації за пріоритетом використання ліків та їх класифікація. Діаграма на рисунку 3.1 ілюструє ці компоненти [28].

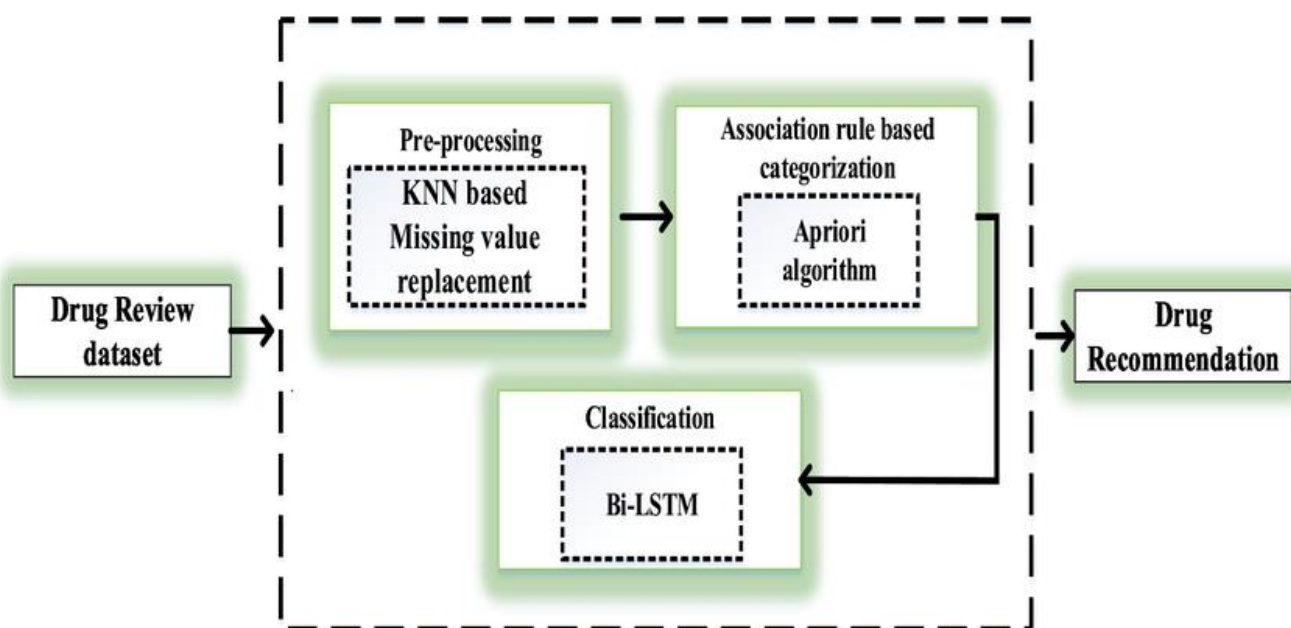


Рисунок 3.1 – Діаграма прикладу CDSS для системи підбору ліків

CDSS для системи підбору ліків відіграє ключову роль в автоматизації процесів, включаючи збір даних про ліки, їхні атрибути та результати клінічних досліджень. Система забезпечує оновлення цих даних та інтегрує необхідні інтерфейси, такі як веб-інтерфейс і API, що управляє взаємодією між компонентами системи. Це включає інтерфейс бази знань, бази даних та представлення рекомендацій щодо ліків. Також система містить модуль відгуків від пацієнтів і лікарів для оновлення і покращення бази знань та інтеграції незалежних джерел інформації.

Основним елементом системи є компонент формування рекомендацій, який базується на певній моделі. На основі цього висновку можна визначити ключові

завдання, які виконує компонент формування рекомендацій у системі підбору ліків. Ця абстракція, яка представляє модель, може використовуватися як основа для розробки методу. Основним завданням є обробка даних із датасету, що включає вирішення проблем з парсингом даних та підготовку даних для аналізу, враховуючи обрані метрики, такі як атрибути ліків.

Щоб цього досягти, визначено наступні задачі:

- класифікація ліків за їх атрибутами, щоб забезпечити високу точність та знизити ризики неправильного лікування.
- фільтрація – процес, який включає вибір ліків згідно з класифікацією та потребами, що впливають з діагнозу.
- формулювання рекомендацій – процес створення списку рекомендованих ліків з урахуванням поставленого діагнозу та наявних протипоказань.

Також треба зазначити, що під час задачі класифікації можливе створення кластерів, з використанням кластеризації, і, в подальшому, виконання класифікації вже за сформованими кластерами.

В цьому контексті, класифікація це «встановлення функціональної залежності між вхідними і дискретними вихідними змінними. За допомогою класифікації вирішується завдання приналежності об'єктів до одного зі заздалегідь відомих класів. Кластеризація – групування об'єктів на основі їх властивостей. Об'єкти всередині кластера повинні бути схожими і відрізнятися від об'єктів, що увійшли в інші кластери. Чим більше схожі об'єкти в середині кластера і чим більше відмінностей між кластерами, тим точніше кластеризація» [29].

На основі розглянутих підзадач можна визначити їхні взаємозв'язки. Першим кроком виступає обробка даних, проте важливим завданням є вибір принципу обробки інформації з моделі, що реалізує метод: вибір між апіорним та апостеріорним підходами. Апіорний підхід припускає відсутність попереднього досвіду і представляє знання самостійно, тоді як апостеріорний передбачає набуття досвіду і формування методу на його основі. Однак, у контексті медицини, яка значно розвинулася за останні 150 років, особливо з появою комп'ютерних наук,

великий обсяг інформації та знань можна використовувати як апіорний фонд, тоді як апостеріорний підхід може збільшити ризики, а нові дані (наприклад, відгуки лікарів) можуть бути інтегровані в систему підтримки прийняття рішень через оновлення датасетів.

Таким чином, класифікація має бути наступним етапом у системі, за яким слідує компонент для пошуку потрібних ліків, використовуючи інтелектуальні методи аналізу, що включають фільтрацію. На основі результатів фільтрації можна автоматично формулювати необхідні рекомендації.

Відповідно до наведеного раніше опису, сформульовано будову цієї моделі і, на основі цього, можна представити її абстрактну модель.

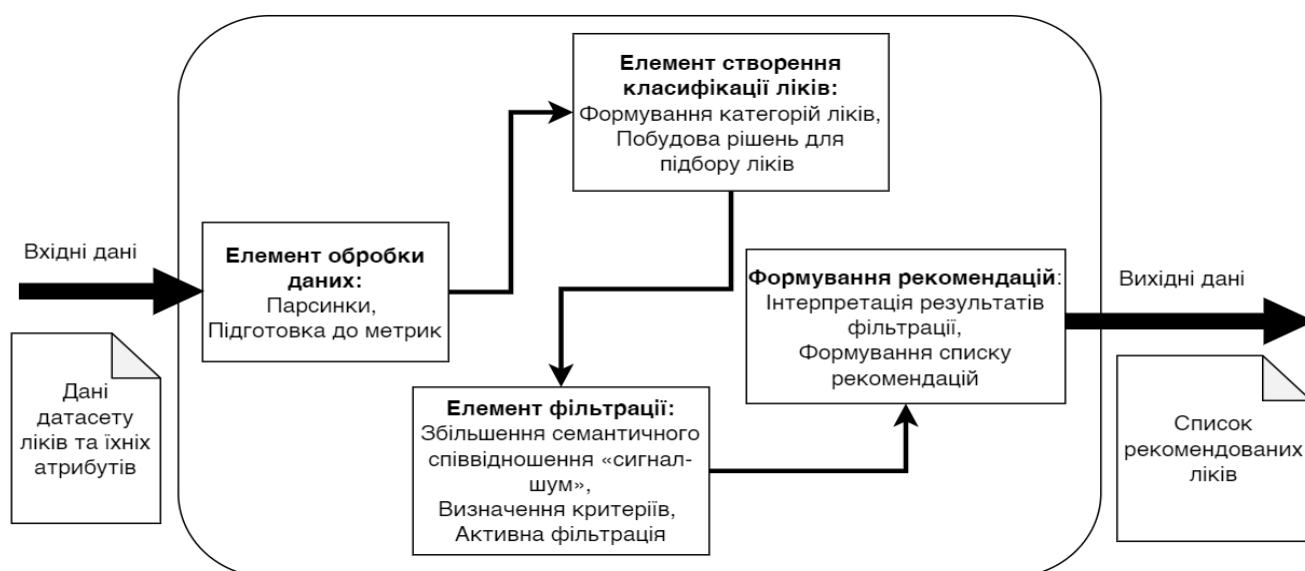


Рисунок 3.2 – Діаграма моделі формування рекомендацій для системи підбору ліків.

Представлення цієї моделі у вигляді послідовних кроків описує метод, спрямований на формування рекомендацій у системі підбору ліків. Тепер, виходячи з описаного процесу прийняття рішень, необхідно вибрати ті інструменти, які найефективніше реалізовуватимуть цей метод.

Таким чином, на базі цієї моделі можна створити математичний опис розробленого методу та, в подальшому, провести додаткову перевірку на предмет його придатності для реалізації з метою уточнення.

$$RCM = Interp(Filt(Prep(X_1, \dots, X_i, \dots, X_n) | Crit_{1..k})),$$

де RCM – результат формування рекомендацій,

$Interp$ – функція формування списку рекомендацій,

$Filt$ – функція, що працює як активна фільтрація, що отримує у вхідні дані підготовлені дані та перевіряє за умовами вказаними як $Crit_{1..k}$,

$Crit_{1..k}$ – список критеріїв від одного до k і вказується відповідно лікарем,

$Prep(X_1, \dots, X_i, \dots, X_n)$ – функція обробки даних, що включає в себе парсинг та підготовку даних і отримує на вхід дані датасетів ліків з кількістю записів з 1 до n .

Існує декілька підходів до формування рекомендацій для підбору лікарських засобів. Вони включають знання на основі правил, контентно-орієнтовані методи, колаборативну фільтрацію, гібридні методи, машинне навчання та інші засоби ШІ. Тому важливо розглянути причини, чому не використовуються лінійні та статистичні методи, такі як лінійна регресія або створення матриць для обчислення коефіцієнту відповідності.

Основною проблемою є те, що ліки мають різноманітні атрибути, які важко обробити як простий текст, а також відсутність лінійної залежності між ними. У контексті системного аналізу, коли основна проблема сприймається як слабо структурована, це підкреслює неможливість встановлення лінійних відносин між деякими атрибутами. Наприклад, варто з'ясувати, чи існує безпосередній зв'язок між застосуванням ліків та протипоказаннями, адже вони можуть бути пов'язані з хронічними захворюваннями, алергіями або навіть вагітністю, маючи одну й ту ж причину, але при цьому виключаючи один одного. Також активна речовина не може бути прямо пов'язана з її застосуванням, оскільки, наприклад, невідомо, як саме активний компонент засобу, як-от парацетамол, функціонує як знеболюючий засіб.

Взагалі, у цій ситуації ми маємо справу з різнотипними даними та нелінійними зв'язками між ними, і, тим більше, між даними та результатом. Тому для реалізації методу потрібно вибирати засоби, що можуть працювати з

нелінійними відношеннями, а також з різнотипними даними, які в подальшому можливо оцінити за поставленими метриками.

3.2 Дослідження та аналіз існуючих методів

Таким чином, необхідно повернутися до послідовного формування методів. Спочатку здійснюється короткий опис і аналіз кожного з методів, після чого, базуючись на цих даних, проводиться тестування для завершальної реалізації розробленого методу.

Методи на основі знань використовують фіксовані правила для прийняття рішень. Вони базуються на експертних знаннях і включають правила, що визначають взаємодії між ліками, їхні протипоказання, дозування та інші фактори. Як приклад можна представити ситуацію з експертною системою для підбору антигіпертензивних препаратів, що враховує наявність хронічних захворювань, таких як діабет або хронічна ниркова недостатність. Такого роду методи формуються на алгоритмах, що формують рішення з фактів через логічні оператори І чи АБО. До них відносяться наступні алгоритми (рис. 3.3):

- алгоритм прямого виведення (Forward Chaining): застосовується для виведення нових фактів на основі вже відомих;
- алгоритм зворотного виведення (Backward Chaining): використовується для пошуку рішень шляхом визначення умов, які необхідно задовільнити.

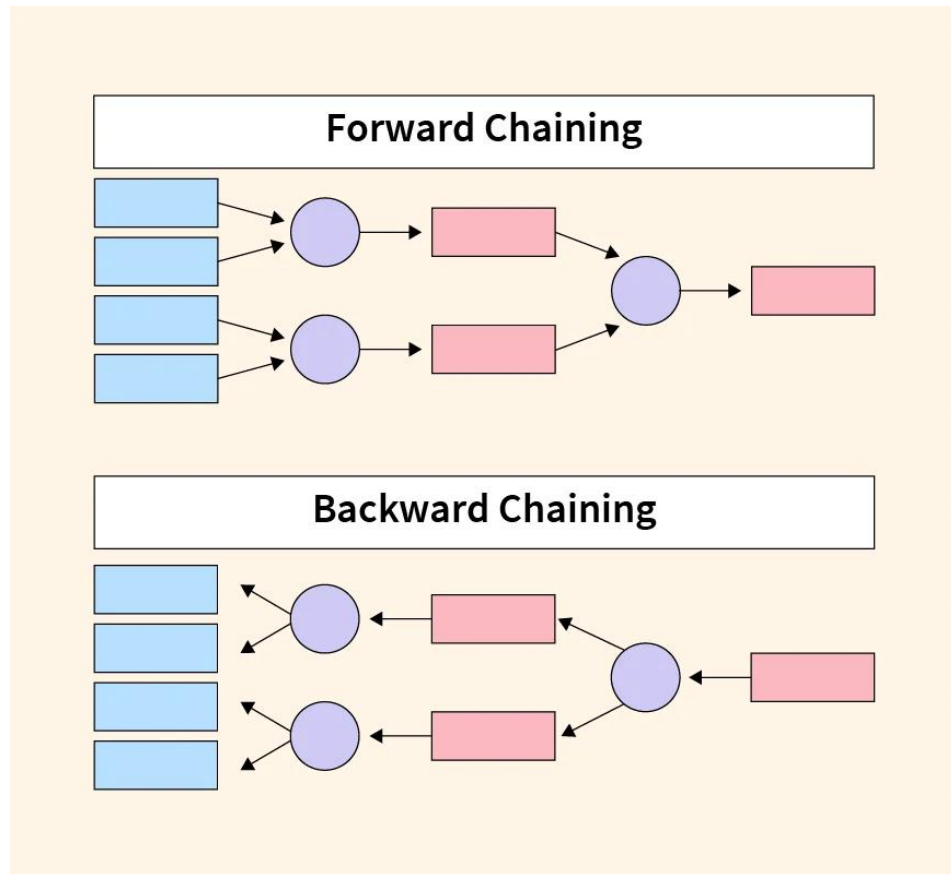


Рисунок 3.3 – Діаграма алгоритмів методів на основі знань

Детальний аналіз таких методів виявляє дві основні проблеми, які породжують ряд інших складнощів. Перша проблема полягає у значній потребі в людських ресурсах, зокрема, у залученні групи експертів для формування правил. Друга проблема – обмеженість використання, що є великим недоліком, оскільки з кожним днем з'являється все більше ліків і препаратів, оновлюється інформація про них та їх застосування, і вона може бути актуальною лише для вузької спеціалізації серед медичних препаратів.

На основі дослідження відповідної літератури, сформовано список переваг та недоліків методів такого типу. До переваг відносяться:

- простота та зрозумілість – легко розуміти та інтерпретувати правила, що керують рекомендаціями;
- контрольованість – лікарі можуть безпосередньо коригувати правила на основі клінічного досвіду.

В свою чергу, недоліками таких методів є:

- жорсткість – складно адаптуватися до нових даних або умов;
- трудомісткість – вимагає значних зусиль для створення та підтримки правил.

До позитивного моменту такого роду методів можна віднести формування чітких правил на основі досвіду експертів-лікарів, але це нівелюється старінням їхніх знань і значним обсягом оновлень ліків. Тому такий тип методів не підходить, бо створений метод повинен базуватися на апріорному принципі, а не на підвищенні ризиків через залучення апостеріорних знань.

Контентно-орієнтовані методи, в свою чергу, аналізують характеристики ліків та підбирають препарати, схожі за властивостями з тими, які вже були призначені пацієнту. Ці методи використовують текстову інформацію, таку як опис препаратів, побічні ефекти та інші атрибути. До прикладу, система, що рекомендує аналогічні препарати на основі активних інгредієнтів та хімічної групи ліків. До таких методів відносяться методи обробки природної мови (Natural Language Processing або NLP), які у цій предметній області можуть бути застосовані для аналізу текстової інформації про ліки та симптоми. Наступні алгоритми часто використовуються для цього:

- TF-IDF (Term Frequency-Inverse Document Frequency) – для визначення важливості слів у текстах про ліки та симптоми (рис.3.4);

c-TF-IDF

For a term x within class c :

$$W_{x,c} = \| \mathbf{tf}_{x,c} \| \times \log \left(1 + \frac{A}{f_x} \right)$$

$\mathbf{tf}_{x,c}$ = frequency of word x in class c

f_x = frequency of word x across all classes

A = average number of words per class

Рисунок 3.4 – Алгоритм TF-IDF

- косинусна схожість (Cosine Similarity) – для порівняння схожості між описами ліків та симптомами пацієнтів (рис. 3.5);

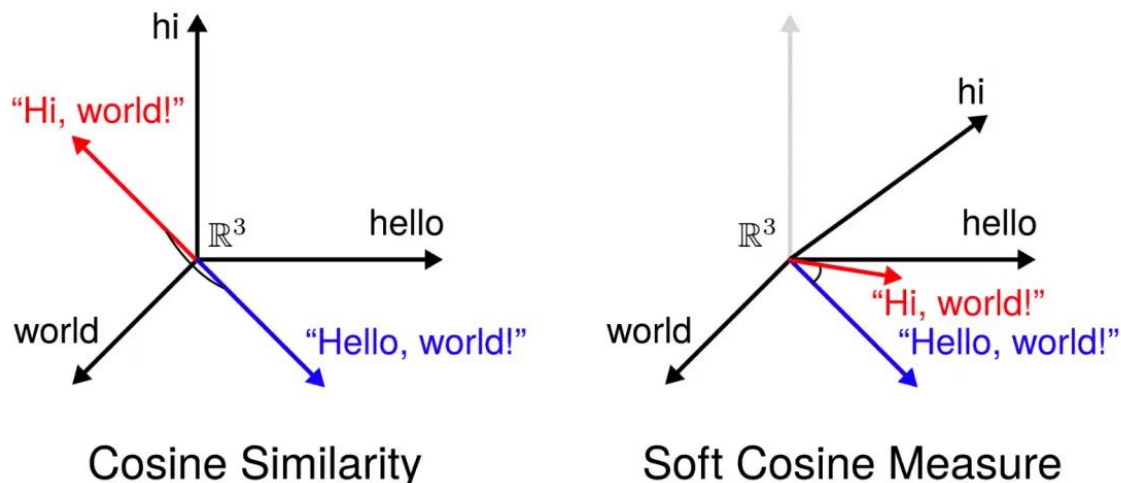


Рисунок 3.5 – Діаграма роботи алгоритму косинусної схожості

При аналізі цих методів є дві проблеми, а саме апостеріорний підхід, що не підходить для сформованого методу, а також відсутність врахування більшості ризиків, що потрібно при фільтрації ліків для створення списку рекомендацій. Загалом такі методи можуть бути корисними при клінічних дослідженнях нових препаратів, або при створенні аналогічних препаратів, але обмеженість і застарілість даних не дають можливості формування якісного підходу через їхнє постійне оновлення.

На основі досліджених наукових публікацій, отримано список переваг та недоліків цього типу методів. До переваг можна віднести:

- індивідуалізація – використовує характеристики препаратів та пацієнтів для персоналізованих рекомендацій;
- самостійність – може функціонувати незалежно від інших користувачів та їхніх даних.

Недоліками ж є:

- обмеженість – може не враховувати взаємодії між препаратами або інші зовнішні фактори;
- застарілі дані – потребує постійного оновлення інформації про препарати.

Колаборативна фільтрація (Collaborative Filtering) використовує аналіз вподобань та виборів інших користувачів для рекомендацій. Цей метод поділяється на два типи (рис. 3.6): user-based (заснований на аналізі поведінки користувачів) та item-based (заснований на аналізі характеристик предметів). Наприклад, можна розглянути систему, яка рекомендує ліки на основі вибору інших пацієнтів з аналогічними симптомами або історією хвороби.

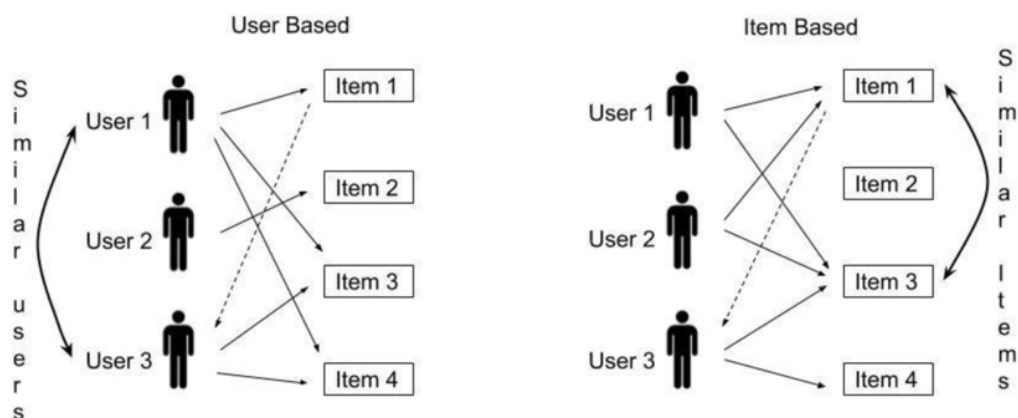


Рисунок 3.6 – Діаграма роботи item-based та user-based фільтрації

З одного боку, цей метод є ефективним для формування рекомендацій, але з іншого боку, він не завжди заснований на ефективності лікування і може зазнавати впливу успішного маркетингу ліків або помилок, пов'язаних з ефектом плацебо. Він орієнтований на задоволення потреб клієнтів, а не на підвищення їхньої результативності лікування.

Відповідно до аналізу існуючих наукових статей, виокремлено перелік переваг та недоліків. Перевагами методів цього типу є:

- широкий охоптя – використовує дані про вподобання та вибори великої кількості користувачів;
- виявлення схожості – може знаходити приховані зв'язки між різними препаратами та користувачами.

В свою чергу, до недоліків відносяться:

- проблема холодного старту – потребує значної кількості даних для початкового налаштування;

- проблема масштабування – може бути обчислювально інтенсивною при великій кількості користувачів та ліків.

Тепер перейдемо до наступного типу методів. Гібридні методи поєднують різні стратегії для забезпечення більшої точності рекомендацій. Вони можуть інтегрувати правила, що базуються на знаннях, з контентно-орієнтованими методами та колаборативною фільтрацією. Наприклад, система може спочатку відфільтрувати ліки за протипоказаннями, а потім застосовувати колаборативну фільтрацію для додаткового уточнення рекомендацій. Такий підхід є характерним для гібридного методу, званого Weighted Hybrid (рис. 3.7).

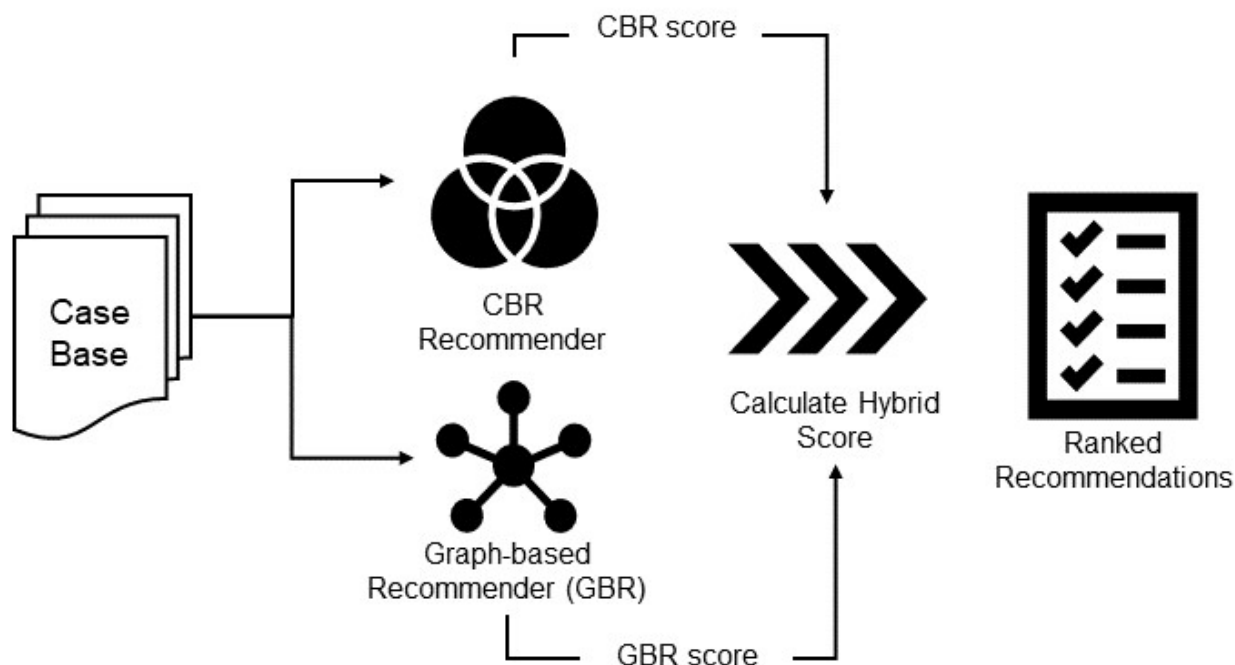


Рисунок 3.7 – Діаграма методу Weighted Hybrid

Під час аналізу подібних методів, виявлено численні дослідження, які, в більшості випадків, демонструють високу ефективність цих підходів. Це залежить від конкретних методів, що використовувалися, та від специфіки завдань, для яких вони були розроблені. Враховуючи, що в роботі розглядається реалізація методів на основі їхньої ефективності, такий тип методів може бути застосований у цьому

дослідженні, і існує значна ймовірність, що розроблений метод буде належати саме до цього типу.

Відносно аналізу наукових публікацій за останні роки, сформовано список переваг та недоліків. Перевагами таких підходів є:

- гнучкість – поєднує різні підходи для досягнення більш точних результатів;
- висока ефективність – може враховувати різні аспекти даних, зменшуючи обмеження окремих методів.

В свою чергу, до недоліків відносяться:

- складність реалізації – вимагає інтеграції та налаштування різних алгоритмів;
- обчислювальні витрати – може бути ресурсомістким в реалізації та підтримці.

3.3 Методи машинного навчання та гібридні методи

Розвиток машинного навчання є важливою складовою у багатьох аспектах сучасних технологій, включаючи автоматизоване розпізнавання зображень і створення систем рекомендацій. У контексті рекомендаційних методів, ефективність машинного навчання має безпосередній вплив на здатність системи забезпечувати релевантні та точні пропозиції.

Методи машинного навчання можна класифікувати за різними критеріями, причому одним з найпоширеніших є поділ на навчання з учителем, навчання без учителя та навчання з підкріпленням.

Першим на огляді є навчання з учителем. Навчання з учителем – «це, як правило, завдання машинного навчання, яке полягає у вивченні функції, що відображає вхідні дані у вихідні на основі зразкових пар вхід-вихід. Воно використовує марковані навчальні дані та колекцію навчальних прикладів для виведення функції. Навчання під контролем здійснюється, коли визначено певні цілі, які мають бути досягнуті на основі певного набору вхідних даних, тобто

підхід, орієнтований на завдання. Найпоширенішими керованими завданнями є «класифікація», яка розділяє дані, і «регресія», яка підганяє дані. Наприклад, передбачення класу або настрою фрагмента тексту, такого як твіт або відгук про товар, тобто класифікація тексту, є прикладом керованого навчання» [39].

Наступним є навчання без вчителя. Навчання без вчителя «аналізує немарковані набори даних без втручання людини, тобто це процес, керований даними. Це широко використовується для вилучення генеративних ознак, виявлення значущих тенденцій і структур, групування результатів і дослідницьких цілей. Найпоширенішими завданнями неконтрольованого навчання є кластеризація, оцінка щільності, вивчення особливостей, зменшення розмірності, пошук правил асоціації, виявлення аномалій тощо» [39].

Наостанок опишемо навчання з підкріпленням. Навчання з підкріпленням – «це тип алгоритму машинного навчання, який дозволяє програмним агентам і машинам автоматично оцінювати оптимальну поведінку в певному контексті або середовищі для підвищення її ефективності, тобто підхід, орієнтований на навколишнє середовище. Цей тип навчання ґрунтується на винагороді або покаранні, а його кінцева мета полягає у використанні інсайтів, отриманих від екологічних активістів, для вжиття заходів, спрямованих на збільшення винагороди або мінімізацію ризику. Це потужний інструмент для навчання моделей ШІ, який може допомогти підвищити автоматизацію або оптимізувати операційну ефективність складних систем, таких як робототехніка, завдання автономного водіння, виробництво і логістика ланцюгів поставок, однак не бажано використовувати його для вирішення базових або простих завдань» [39].

Правильний вибір типу машинного навчання або їх комбінації залежить від специфічних вимог та наявних даних. Кожен тип навчання має свої переваги та обмеження, і ефективне застосування цих методів вимагає детального розуміння як теоретичних, так і практичних аспектів машинного навчання.

Алгоритми машинного навчання є фундаментом для аналітичних систем, які можуть виявляти закономірності та робити прогнози, базуючись на доступних даних. Ці алгоритми класифікуються на основі типу вихідних даних та поставленої

задачі. До найбільш використовуваних алгоритмів належать лінійна регресія, логістична регресія, дерева рішень, випадкові ліси та інші. Кожен з цих алгоритмів має свої унікальні переваги та обмеження. Вибір конкретної моделі залежить від специфіки задачі, наявності та обсягу даних, а також вимог до точності та швидкості обробки. Таким чином, необхідно провести аналіз певних алгоритмів, які б могли відповідати вимогам розроблюваного метода.

Лінійна регресія – це статистична модель, яка використовується для прогнозування значення залежної змінної (Y) на основі однієї або декількох незалежних змінних (X). Модель представляє залежність як лінійну функцію:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n + \varepsilon,$$

де β_0 – це перетин з віссю Y (константа);

$\beta_1 X_1 + \dots + \beta_n X_n$ – коефіцієнти при незалежних змінних, що вказують на міру зміни Y при зміні відповідного X на одиницю;

ε – випадкова помилка.

Лінійна регресія використовується для прогнозування числових значень і оцінки залежностей між змінними, що дозволяє передбачати тенденції. Цей метод широко застосовується для аналізу взаємозв'язків між різними характеристиками, сприяючи оцінці впливу одних змінних на інші.

Логістична регресія – це статистична модель, яка використовується для оцінки ймовірності належності до певної категорії (зазвичай бінарної: 0 або 1). Вона виражає логарифм відношення шансів лінійної комбінації вхідних змінних:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n,$$

де p – ймовірність належності до категорії 1.

Логістична регресія використовується для бінарної класифікації, яка дозволяє визначати ймовірність належності об'єкта до однієї з двох категорій. Цей метод може бути застосований для прогнозування того, чи буде певний продукт цікавим користувачеві, а також використовується для класифікації в різних

галузях, де потрібно визначити наявність або відсутність певних характеристик у об'єктів.

Градiєнтний бустинг є ефективною технікою машинного навчання, що використовується у різних прогнозних задачах, включаючи класифікацію та регресію. Градiєнтний бустинг належить до класу ансамблевих алгоритмів, де кожна наступна модель виправляє помилки попередньої, підвищуючи загальну точність і надійність моделі. Цей метод базується на принципі послідовного додавання нових моделей до ансамблю, кожна з яких виправляє помилки, допущені попередніми моделями, шляхом мінімізації втрат, які визначаються градієнтом функції втрат. Нові моделі в ансамблі концентруються на випадках, які були найскладнішими для класифікації або апроксимації попередніми моделями, тим самим зосереджуючись на їх корекції.

Цей метод виявив високу ефективність у широкому діапазоні застосувань, від конкурсів з аналізу даних до реальних використань у промисловості, фінансах, медицині та інших сферах. Цей метод цінується за свою здатність ефективно обробляти великі обсяги даних та за високу точність прогнозування, що робить його особливо корисним у складних аналітичних завданнях.

За результатами аналізу, наведені методи можуть вільно працювати з великим обсягом різнотипних даних, що наявні у медичній сфері, та, разом з цим, можуть зберігати актуальність для лікарів.

Згідно проаналізованих наукових публікацій та відповідної літератури, отримано список переваг та недоліків. До переваги відносяться:

- адаптивність – моделі можуть навчатися на великій кількості даних та адаптуватися до нових умов;
- прогнозування – може враховувати складні взаємодії та робити точні прогнози.

Недоліками ж є:

- потреба в даних – вимагає великих обсягів якісних даних для навчання;
- інтерпретованість – деякі моделі важко інтерпретувати (наприклад, нейронні мережі).

Відповідно до проведеного аналізу, згадані методи можуть знаходити своє застосування в системах, що пов'язані з медичною сферою. Проте, через недоліки та невідповідності існуючих методів формування рекомендацій до розробленого методу в системі підбору ліків, вирішено відмовитися від застосування методів, заснованих на знаннях.

Враховуючи постійне оновлення даних та зростання обсягів вхідних даних у предметній області, використання виключно методів колаборативної фільтрації стає неможливим, а також виявляється неефективним застосування методів, заснованих на знаннях і контентно-орієнтованих методів. З огляду на це, необхідно переглянути альтернативи, зокрема використання гібридних методів та методів машинного навчання, включаючи їхні моделі та алгоритми, для реалізації в розробленому методі. Значна кількість алгоритмів та моделей машинного навчання вимагає ретельного вивчення тих, що можуть бути ефективно інтегровані у метод підбору ліків у системах DSS. Вибір відповідних методів має базуватися на специфіці завдань та доступних даних, що дозволяє методам машинного навчання вирішувати широкий спектр ключових завдань, таких як прогнозування, класифікація, кластеризація, та побудова асоціативних правил для глибшого аналізу та точнішого прогнозування в медичних застосуваннях.

Методи машинного навчання (рис. 3.8), такі як дерева рішень (Decision Trees), випадкові ліси (Random Forest), підтримуючі векторні машини (SVM) та нейронні мережі, мають свої унікальні переваги та обмеження, які необхідно враховувати при проектуванні систем. Використання гібридних методів може пропонувати більш гнучкі та ефективні рішення шляхом комбінації переваг різних алгоритмів. Адекватний вибір та поєднання методів дозволяють створити систему, яка забезпечує високу точність, персоналізацію та ефективність рекомендацій, що є вирішальним для успішного лікування пацієнтів.

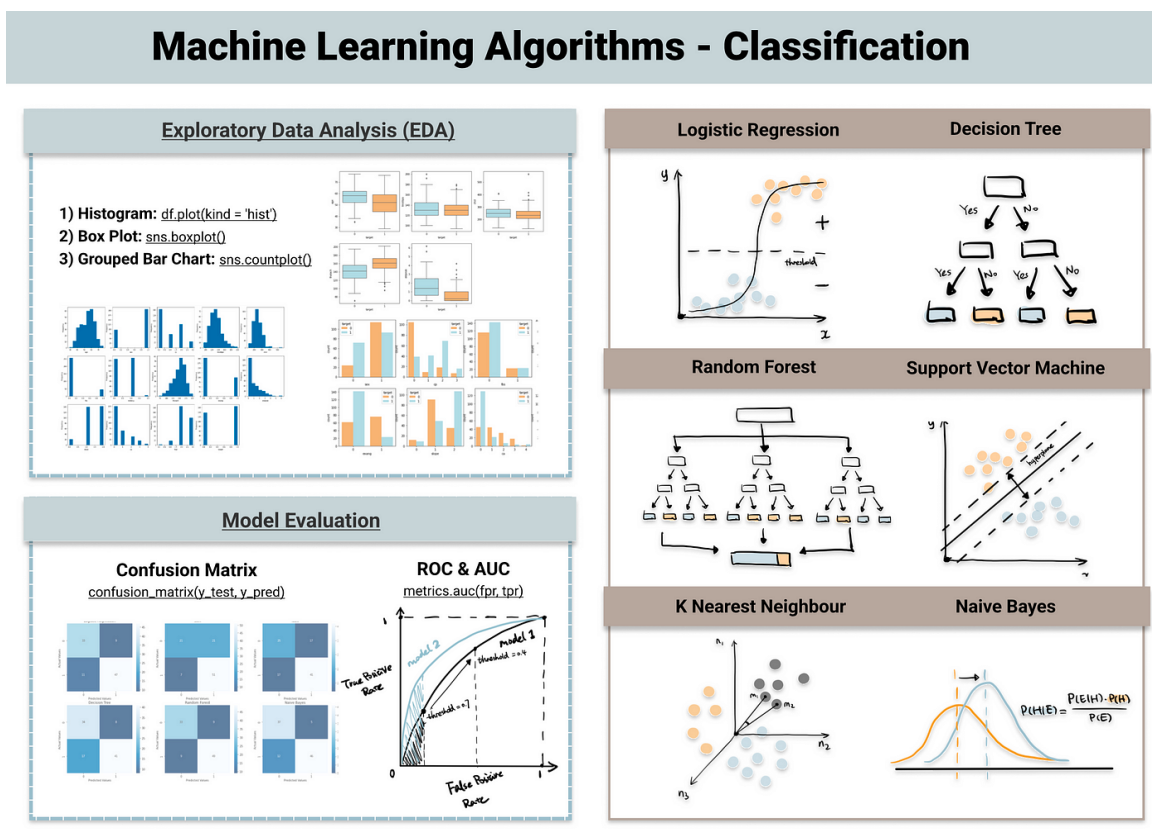


Рисунок 3.8 – Діаграма методів Machine Learning

Очевидно, що застосування логістичної регресії у сучасних цифрових пристроях, здатних обробляти значний обсяг даних і адаптуватися до специфічних запитів лікаря для формування медичних рекомендацій, є складним завданням, тому її використання не є можливим. Подібна проблема виникає і з алгоритмами, заснованими на Naive Bayes.

Дерева рішень становлять один з найпростіших і найбільш інтуїтивно зрозумілих методів машинного навчання, які використовуються для класифікації та регресії. Вони розділяють дані згідно з заданими умовами, формуючи структуру дерева, де кожен вузол символізує атрибут, а гілки представляють рішення.

Дерево рішень як алгоритм машинного навчання інтегрує логічні правила типу if-then у структуру дерева, створюючи ієрархічну систему правил. У дереві рішень (рис. 3.9) кожен вузол виконує перевірку умови, а листя (рішення) вказує на клас або середнє значення. В процесі побудови дерева визначається приріст інформації, який базується на оцінці ентропії. Ентропія позначає ступінь невизначеності або хаосу в системі – чим вища ентропія, тим менш впорядкована

система, і навпаки, інформація є протилежною до ентропії. Ентропія Шеннона розраховується для системи з N можливими станами:

$$H(Y) = - \sum_i^N p_i \log_2 p_i,$$

де p_i - ймовірність стану системи в i -стані.

Принцип жадібної максимізації приросту інформації лежить в основі алгоритмів для створення дерев рішень. На кожному етапі обирається ознака, яка забезпечує найвищий приріст інформації при її розподілі. Цей процес виконується рекурсивно до тих пір, поки ентропія не зменшиться до нуля або до прийнятно малого рівня, що допомагає уникнути ідеального налаштування дерева під тренувальний набір даних та запобігти перенавчанню. Використання різних евристик для «ранньої зупинки» або «обрізання» дерева дозволяє уникнути створення перенавченого дерева, оптимізуючи його структуру для ефективного використання.

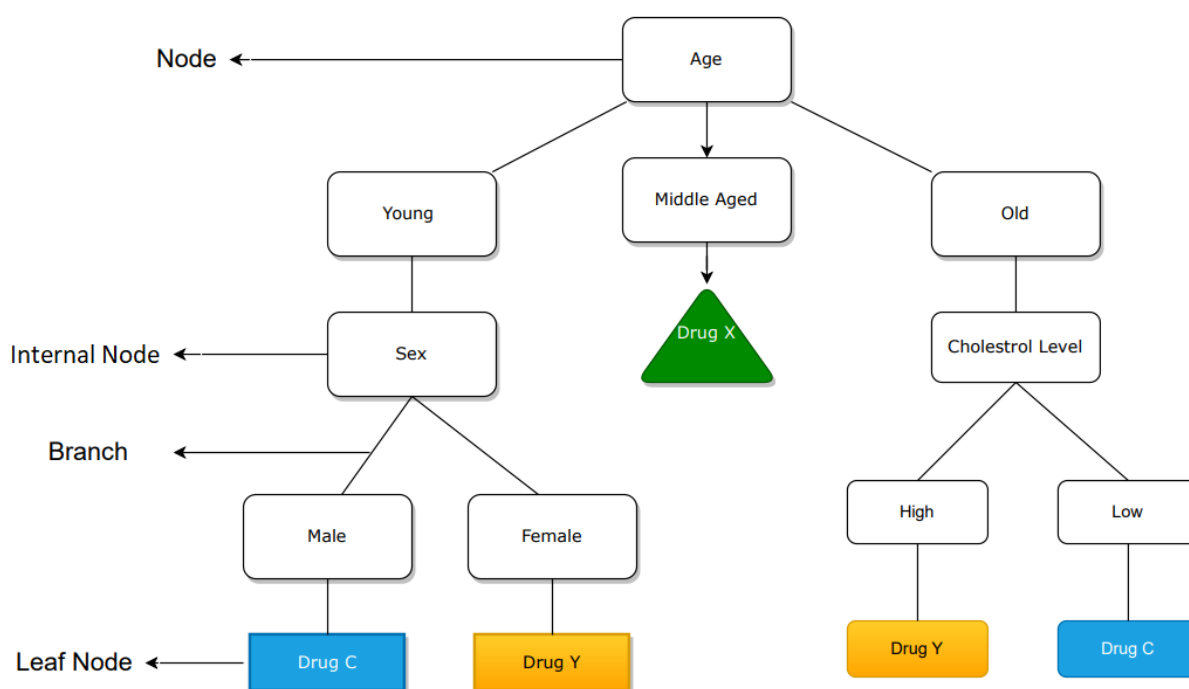


Рисунок 3.9 – Діаграма дерева рішень для підбору ліків

Відповідно до аналізу, отримуємо список переваг та недоліків. До переваг відносяться:

- інтерпретованість;
- простота у використанні;
- швидкість обробки.

Недоліками ж є:

- ймовірність перенавчання - схильність до перенавчання, особливо на малих вибірках або при великій кількості вузлів;
- обмежена точність - порівняно з іншими методами, може мати нижчу точність.

Цей метод можна використовувати для класифікації ліків згідно з їхніми атрибутами. Він дозволяє категоризувати ліки на основі їх властивостей та виявляти протипоказання, тобто ідентифікувати ліки, які не рекомендується застосовувати разом через потенційні негативні взаємодії.

Для підвищення точності моделі з використанням дерев рішень також застосовуються ансамблеві алгоритми машинного навчання, зокрема випадковий ліс, що є однією з реалізацій беггінга.

У методі беггінгу з вихідної вибірки формуються численні випадкові підвибірки за допомогою вибору з повторенням. Кожна модель навчається на своїй підвибірці, а результати їх роботи потім усереднюються. Ефективність беггінга досягається завдяки тому, що базові алгоритми, які навчаються на різних підвибірках, мають різні характеристики і взаємно компенсують свої помилки. Це також знижує вплив об'єктів-викидів, які можуть відсутні у деяких підвибірках.

Беггінг показує високу ефективність, особливо на малих вибірках, де видалення невеликої частини даних може призвести до істотних відмінностей між моделями. Випадковий ліс є однією з форм беггінгу, де використовуються дерева рішень. Під час навчання кожного дерева для розбиття використовується випадкова підмножина ознак.

Random Forest, або Випадкові Ліси, представляє собою ансамбль дерев рішень, де кожне дерево створюється на основі випадкової підмножини даних і

атрибутив. Це поєднання кількох алгоритмів, які навчаються паралельно і допомагають коригувати помилки один одного, що значно підвищує точність і надійність прийняття рішень (рис. 3.10).

До переваг наведеної моделі відносяться:

- висока точність – використання багатьох дерев зменшує ризик переобучення та підвищує загальну точність моделі;
- стійкість до шуму – зменшує вплив шуму та варіативності даних;
- гнучкість – може використовуватися як для класифікації, так і для регресії.

В свою чергу, до недоліків можна віднести:

- менша інтерпретованість – складніше інтерпретувати результати порівняно з окремим деревом рішень;
- обчислювальна вартість – вимагає більше ресурсів для навчання та прогнозування.

Метод може застосовуватися аналогічно до дерев рішень, але з додатковою перевагою у вигляді зниження ризиків неправильного лікування.

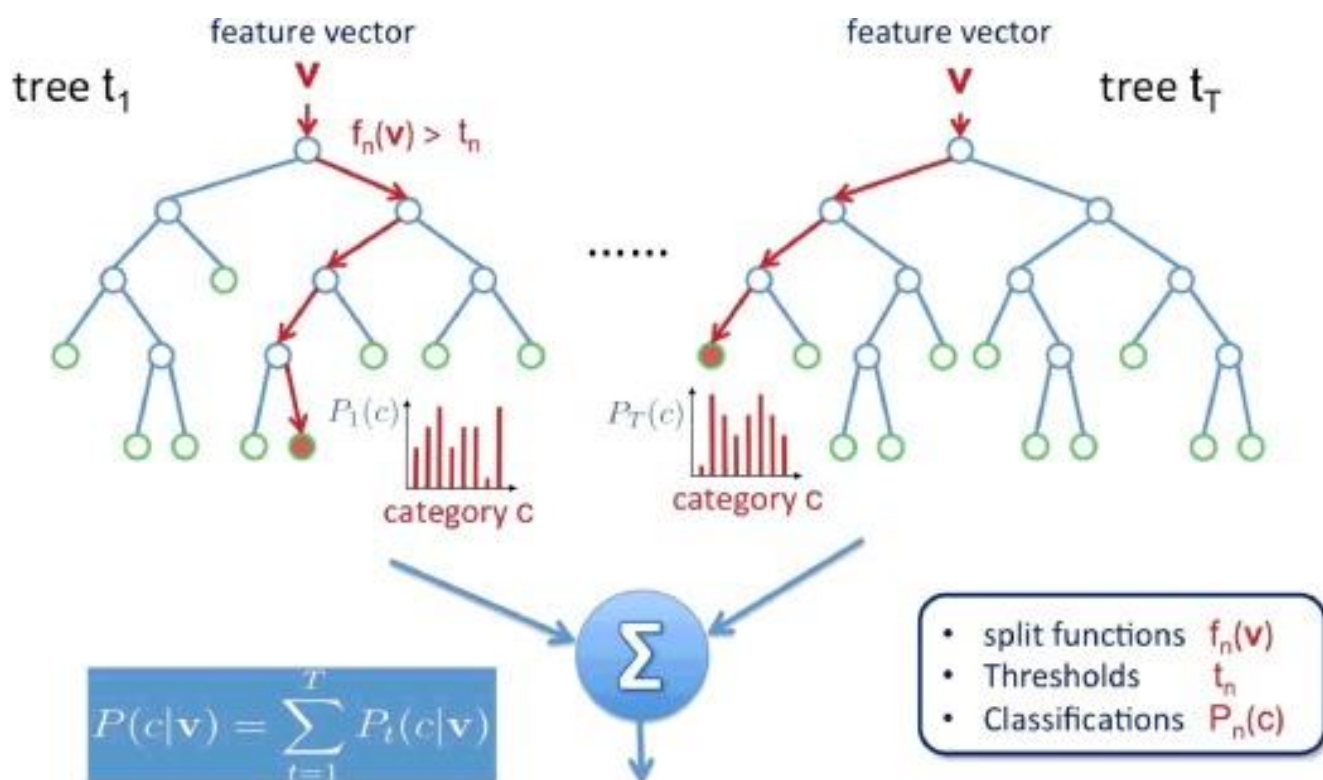


Рисунок 3.10 – Діаграма представлення випадкового лісу

Метод опорних векторів (Support Vector Machines), раніше відомий як алгоритм «узагальненого портрета», використовує гіперплощини для розділення даних на класи, при цьому зусилля зосереджуються на максимізації відстані між цими класами. Основна концепція цього класифікатора полягає у створенні розподіляючої поверхні з використанням обмеженої кількості точок, які знаходяться в критичних позиціях для поділу класів. Точки, що не впливають на позицію гіперплощини, не беруть участь у побудові моделі, але можуть використовуватися в оптимізаційному процесі як «резерв». Точки, що знаходяться безпосередньо на межі або на неправильній стороні відносно межі зазору, називаються опорними векторами. Метод опорних векторів формує класифікаційну функцію F наступним чином:

$$F(x) = \text{sign}(\langle w, x \rangle + b)$$

де w – вектор нормальний, що розподіляє гіперплощини,

b – параметр допоміжний.

Об'єкти у яких $F(x) = 1$ потрапляють до одного класу, об'єкти, для яких $F(x) = -1$ – до іншого. Необхідно знайти такі w і b , що максимізують відстань до кожного класу – $\frac{1}{\|w\|}$. Це можна записати у вигляді задачі оптимізації, яка є стандартною задачею квадратичного програмування і вирішується за допомогою множників Лагранжа (пошук мінімуму $\frac{1}{\|w\|}$ еквівалентний пошуку максимуму $\|w\|$). Задача виглядає наступним чином:

$$\{ \arg \max \|w\|^2, y_i (\langle w, x_i \rangle + b) \geq 1$$

Позитивна визначеність необхідна для того, щоб відповідна функція Лагранжа в задачі оптимізації була обмежена знизу, тобто задача оптимізації була б коректно визначена. Точність класифікатора залежить, зокрема, від вибору ядра. Часто використовують наступні варіації ядра (рис. 3.11):

- Standard Kernel Function Equation, що описується наступним чином:

$$K(\underline{x}) = 1, \text{ if } \|\underline{x}\| \leq 1$$

$$K(\underline{x}) = 0, \text{ Otherwise};$$

- Gaussian Kernel: використовується для виконання перетворення, коли немає попередніх знань про дані та розраховується за формулою:

$$K(x, y) = e^{-\left(\frac{\|x-y\|^2}{2\delta}\right)};$$

- Gaussian RBF Kernel: аналогічно зазначеній вище функції ядра, з додаванням радіального базисного функціоналу для оптимізації процесу трансформації.

- Sigmoid Kernel: функція еквівалентна двошаровій перцептронній моделі нейронної мережі, яка використовується як функція активації для штучних нейронів;

- Polynomial Kernel представляє подібність векторів у навчальному наборі даних у просторі ознак над поліномами вихідних змінних, які використовуються в ядрі.

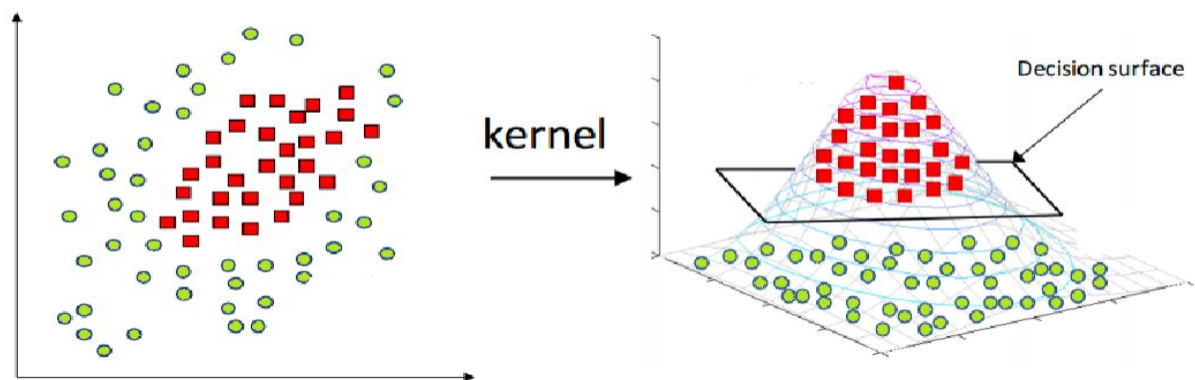


Рисунок 3.11 – Застосування ядра у SVM

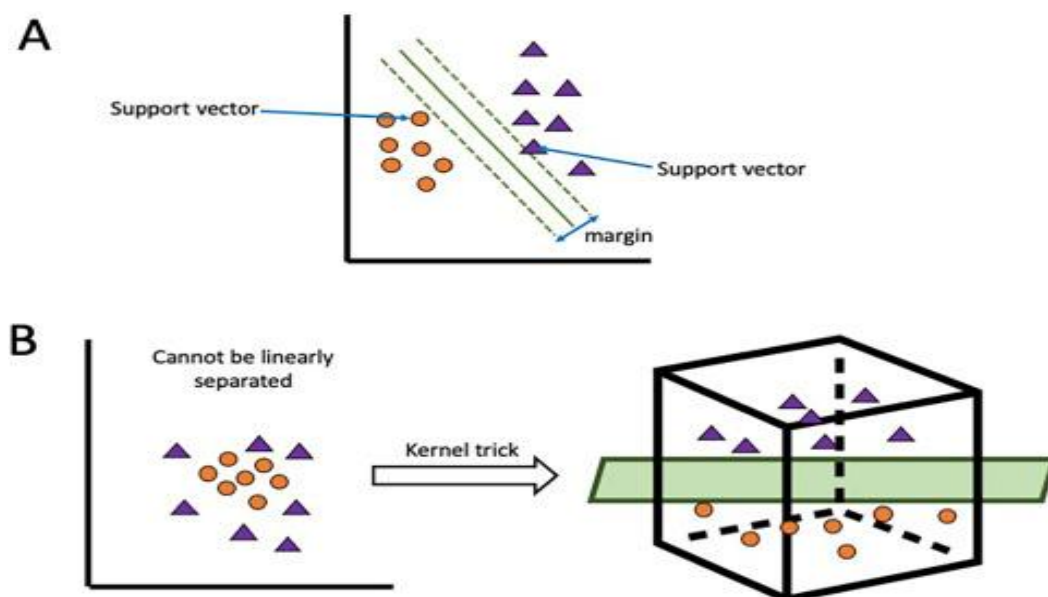


Рисунок 3.12 – Діаграма SVM

Відповідно, маємо сформульований список позитивних та негативних особливостей цієї моделі (рис. 3.12). До переваг відносяться:

- висока ефективність на малих вибірках – може досягати високої точності навіть на невеликих обсягах даних;
- значна узагальненість – добре працює з даними високої розмірності.

В свою чергу, до недоліків можна віднести:

- висока обчислювальна вартість – навчання може бути тривалим, особливо на великих наборах даних;
- потреба в додатковому налаштуванні параметрів – необхідність підбору параметрів (наприклад, вибору ядра), що може бути складним.

Цей метод може бути застосований також для таких алгоритмів, як дерева рішень та випадкові ліси. Він дозволяє покращити ефективність цих моделей, надаючи їм більшу точність та здатність до класифікації та регресії за допомогою структурованого підходу до аналізу даних.

Метод k -найближчих сусідів (k -Nearest Neighbors, k NN) вважається одним із найпростіших і зрозумілих підходів у машинному навчанні, який використовується для розв'язання задач класифікації та регресії. Головна концепція методу базується

на вимірюванні відстаней між об'єктами для визначення їхньої схожості та класифікації.

В основі класифікатора kNN лежить гіпотеза компактності, яка стверджує, що об'єкт тестування d буде мати таку ж класову мітку, як і найближчі об'єкти навчальної вибірки у його локальному оточенні. У варіанті 1NN класифікація об'єкта визначається за найближчим сусідом, тоді як у загальному варіанті kNN об'єкт відноситься до класу, який є домінуючим серед його k найближчих сусідів, де k – параметр, що вибирається в алгоритмі. Правила прийняття рішень у kNN методі визначаються за допомогою діаграми Вороного, яка ділить простір на n опуклих багатокутників, кожен із яких містить лише один об'єкт з навчального набору даних. У багатовимірних просторах межі рішень складаються з сегментів $(n-1)$ -мірних напівплощин, що формуються опуклими багатогранниками Вороного. Алгоритм (див. рис. 3.13) функціонує за принципом «більшості голосів», тобто кінцевий результат – це мітка класу, що отримала більшість голосів.

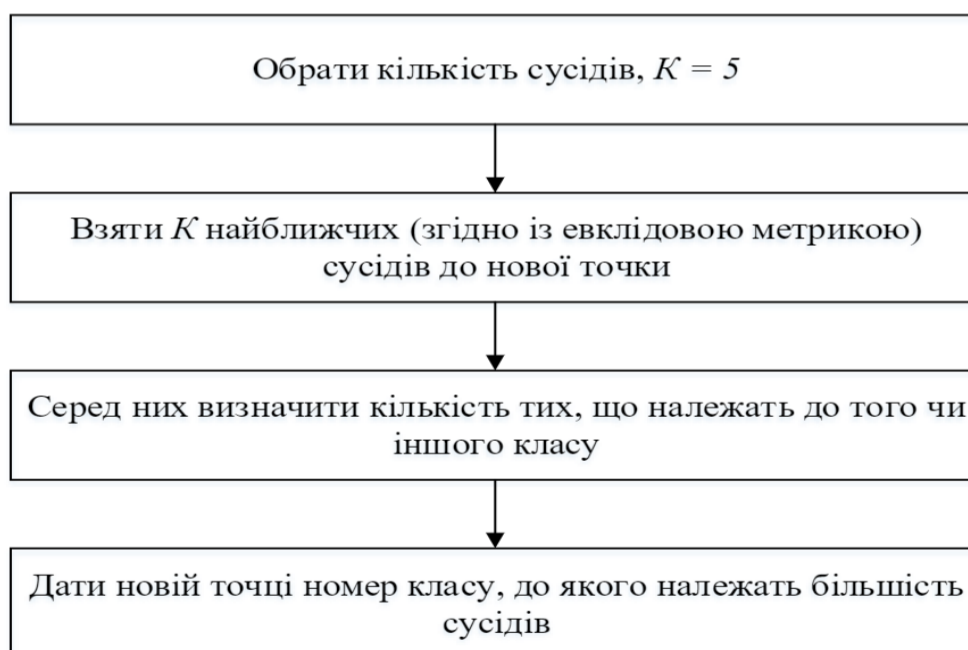


Рисунок 3.13 – Алгоритм методу KNN

Варіант kNN з $k=1$ (рис. 3.14) завжди забезпечує ідеальне розпізнавання для елементів навчальної вибірки, оскільки найближчим сусідом є сам об'єкт, але цей метод часто дає помилки на нових, незнайомих даних. З підвищенням значення k

до певного рівня, якість розпізнавання на контрольній вибірці покращується. При малих значеннях k шум може суттєво вплинути на результати, тоді як великі значення k роблять алгоритм більш вимогливим до обчислень. Один з простих підходів до вибору оптимального k полягає у використанні формули $k = \sqrt{n}$, де n – кількість елементів у навчальній вибірці.

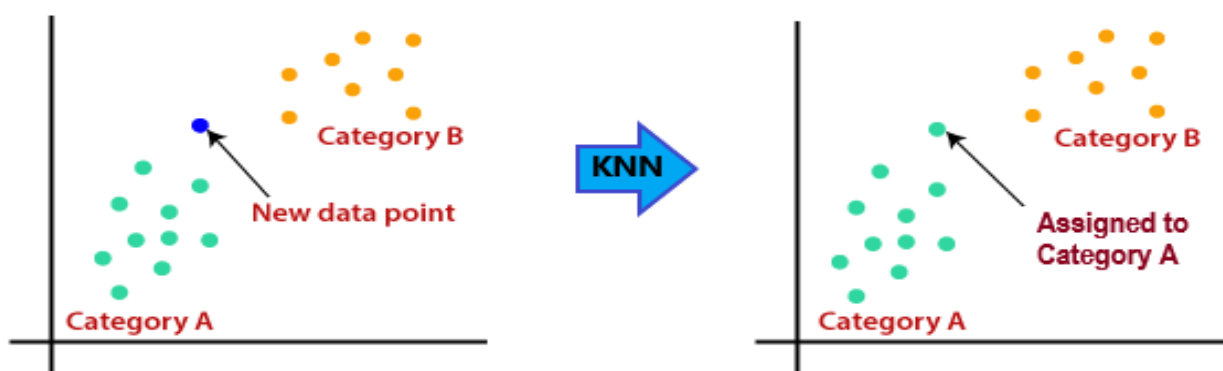


Рисунок 3.14 – Діаграма роботи KNN методу

Метод KNN є ефективним інструментом для класифікації, особливо коли дані мають компакту структуру. Правильний вибір параметра k і метрики відстаней є ключовими для успішного застосування цього методу.

Відповідно до цього, сформульовано список переваг та недоліків цього методу. Перевагами методу є:

- простота реалізації і розуміння;
- висока ефективність на невеликих наборах даних;
- відсутність припущень про розподіл даних.

До недоліків ж відносяться:

- високі обчислювальні витрати на великих наборах даних;
- чутливість до масштабування характеристик;
- вимога зберігання всієї навчальної вибірки.

Адаптивний бустінг (AdaBoost або АВ) є ефективним алгоритмом класифікації, який часто демонструє високу точність. Перед використанням для

прогнозування захворювань, таких як цукровий діабет, критично важливим є вибір відповідних базових моделей та точне налаштування параметрів алгоритму. AdaBoost розпочинає процес з навчання першого слабкого класифікатора на вихідних даних і продовжує навчання наступних класифікаторів, кожен з яких зосереджений на прикладах, що були неправильно класифіковані попереднім. Цей процес повторюється до досягнення заданої точності або поки не буде використано максимально можливу кількість моделей.

Після навчання кожного слабкого класифікатора вони інтегруються в єдиний композитний класифікатор за допомогою зважування їх вкладу на основі точності. Чим вища точність класифікатора, тим більшу вагу він отримує у процесі прийняття рішень. Такий підхід дозволяє алгоритму визначати вагу для кожного слабкого класифікатора і використовувати агреговані результати для визначення класу об'єкта (рис. 3.15).



Рисунок 3.15 – Діаграма методу АВ

Нейронні мережі (Neural Networks) представляють собою розподілені та паралельно працюючі системи, здатні адаптуватися до навчання через взаємодію з зовнішніми стимулами. Ця технологія бере своє натхнення з біології, імітуючи функціональні особливості мозку за допомогою штучних нейронів. Кожен нейрон у мережі має структуру, що включає зважений суматор і нелінійний елемент. Робота нейрона (рис. 3.16) описується наступною математичною моделлю:

$$y = f\left(\sum_i^n w_i x_i, b\right)$$

де f – функція активації,

b – пороговий рівень нейрона,

$\sum_i^n w_i x_i$ – зважена сума вхідних сигналів,

w_i – вагові коефіцієнти ,

x_i – вектор вхідних сигналів .

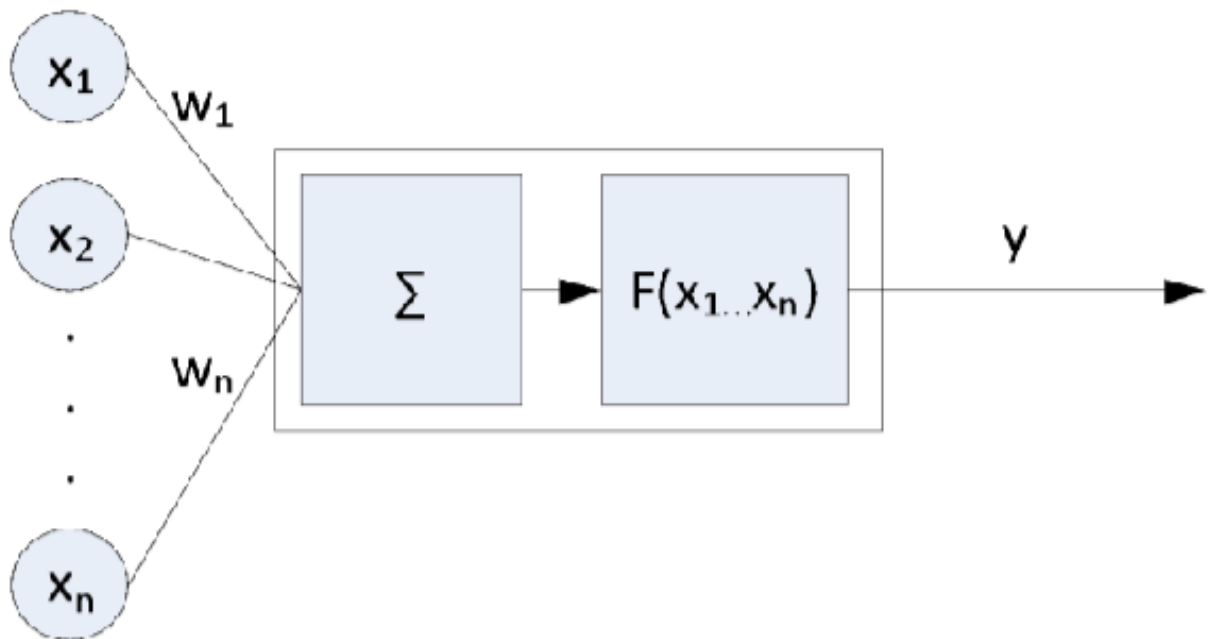


Рисунок 3.16 – Представлення математичного опису роботи нейрона

Нейронні мережі є складними моделями машинного навчання, які включають кілька шарів нейронів, здатних виявляти складні залежності у даних. Ці мережі також є частиною галузі глибокого навчання, оскільки деякі типи нейромереж належать до цієї категорії. У процесі глибокого навчання кожен шар перетворює вхідні дані в більш абстрактні й складні представлення, що дозволяє ефективно вирішувати завдання, які потребують розуміння глибинних особливостей інформації (рис. 3.17).

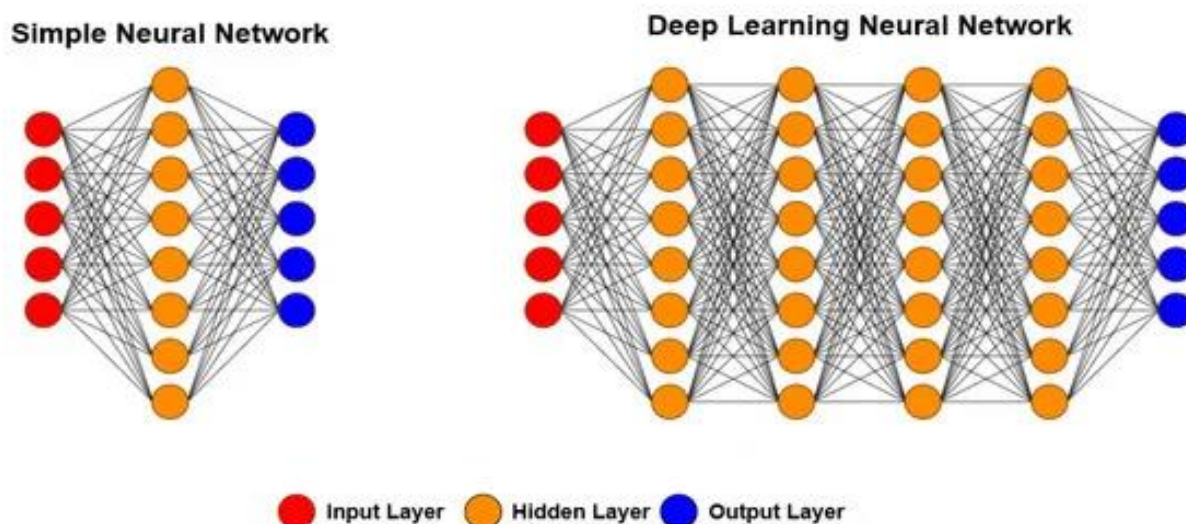


Рисунок 3.17 – Діаграми топологій нейромереж

Відповідно до цього, розроблено список переваг та недоліків. До переваг відносяться:

- висока точність – здатність моделювати складні нелінійні взаємозв'язки між атрибутами;
- адаптивність – можуть навчатися на великих обсягах даних та постійно покращувати свої результати.

В свою чергу, недоліками є:

- висока обчислювальна вартість – навчання та прогнозування можуть вимагати значних ресурсів;
- складність в інтерпретації – важко інтерпретувати результати, що може бути проблемою в медичних додатках.

Ці моделі можуть бути використані для класифікації, прогнозування ефективності медичних рекомендацій, а також виявлення певних знань та залежностей, які мають нелінійний характер. Однак, для глибокого використання та аналізу цих моделей необхідно сформулювати чітке розуміння існуючих типів нейромереж.

Feedforward Neural Networks (FNN) вважаються найбільш розповсюдженим та фундаментальним типом нейронних мереж, які часто описуються у літературі як прості нейромережі. Цей тип мережі містить один або декілька шарів нейронів, які

передають інформацію виключно в одному напрямку – від вхідних шарів до вихідних.

Іншим типом нейронних мереж є Recurrent Neural Networks (RNN) (рис. 3.18), які здатні запам'ятовувати попередні вхідні дані завдяки циклічним зв'язкам між нейронами. Це робить їх ефективними для роботи з послідовними даними і формування нелінійних асоціацій. RNN корисні для прогнозування часових рядів і визначення відповідних ліків на основі контексту.

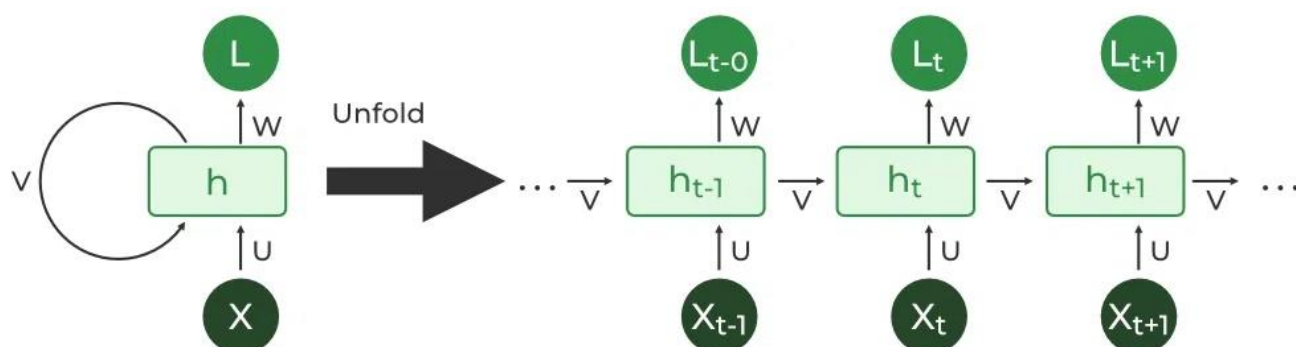


Рисунок 3.18 – Діаграма RNN нейромережі

Існує також тип нейронних мереж під назвою Long Short-Term Memory Networks (LSTM). Ця архітектура (рис. 3.19) була створена для вирішення проблем пам'яті у RNN, які втрачають значну частину свого теоретичного потенціалу. LSTM були розроблені для подолання проблеми зникнення та вибухання градієнтів, дозволяючи ефективно працювати з довготривалими залежностями в даних. Ці нейронні мережі рекомендовані для прогнозування та роботи з тривалими залежностями. Вони мають внутрішній стан комірки, а також прихований стан (h_t), який діє як «інформаційна магістраль» у часі.

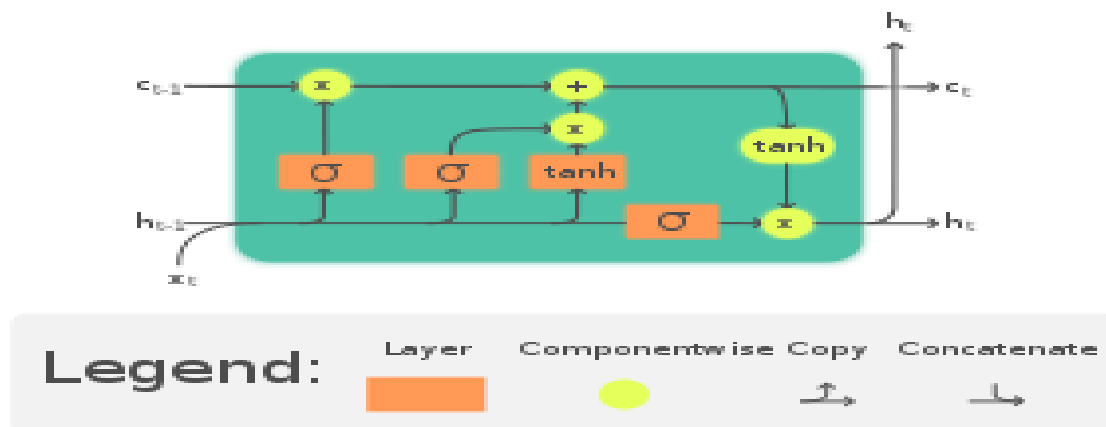


Рисунок 3.19 – Діаграма LSTM неймережі

В свою чергу, Convolutional Neural Networks (CNN) зазвичай використовуються для обробки зображень, але також можуть бути застосовані до послідовних даних за допомогою згорток. Використання CNN (рис. 3.20) дозволяє виявляти властивості та патерни в часових рядах. CNN складаються зі згорткових шарів, які виявляють локальні патерни в даних. Ці шари використовують фільтри для створення ознак, які потім обробляються повнозв'язковими шарами.

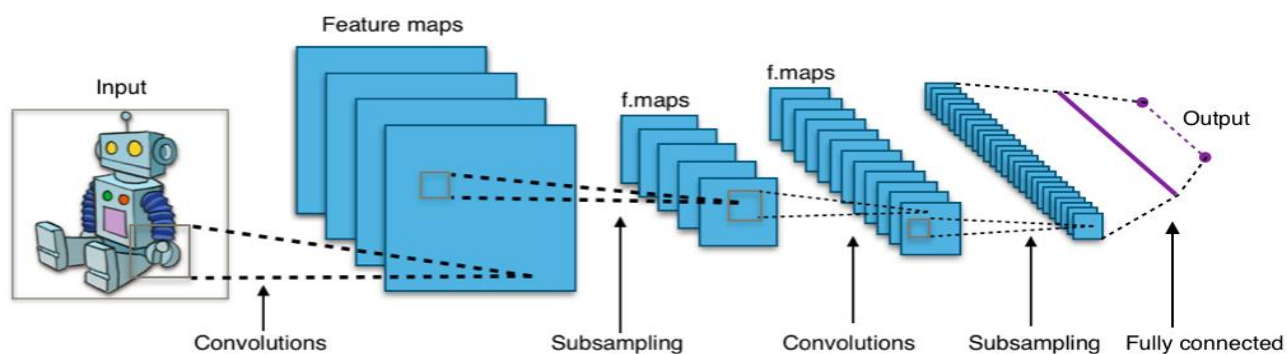


Рисунок 3.20 – Діаграма CNN моделі

Підводячи підсумки проведеного аналізу, гібридні методи поєднують різні підходи для досягнення кращих результатів. Вони можуть комбінувати методи, засновані на знаннях, контентно-орієнтовані підходи, колаборативну фільтрацію та методи машинного навчання, а також включати активну фільтрацію для покращення ефективності.

Відповідно, маємо визначений перелік переваг та недоліків. Перевагами гібридних методів є:

- гнучкість – поєднання різних методів дозволяє враховувати більшу кількість факторів;
- висока ефективність – забезпечення точних та персоналізованих рекомендації, шляхом застосування сильних сторін різних підходів.

До недоліків ж відносяться:

- складність реалізації – вимагає значних зусиль для інтеграції та налаштування різних методів;
- обчислювальна вартість – може вимагати значних ресурсів для навчання та прогнозування.

На основі аналізу методів машинного навчання були досліджені підходи до формування рекомендацій лікарських препаратів. У процесі дослідження визначено критерії для нового методу підбору медикаментів, які передбачають високу точність ідентифікації препаратів та виключення алгоритмів, заснованих на знаннях, через їхні численні недоліки, що підривають ефективність і точність у контексті динамічно змінюваних та оновлюваних даних. В результаті було встановлено, що гібридний метод є оптимальним, оскільки він не піддається обмеженням контентно-орієнтованої і колаборативної фільтрації, уникненню проблем холодного старту та підвищення рівня задоволення пацієнтів. Це сприяє підтримці медичних працівників у прийнятті обґрунтованих рішень щодо лікування та підтримує цифровий процес лікування різноманітних захворювань.

3.4 Проведення аналізу існуючих методів на основі тестового експерименту

Відповідно до дослідження, для оцінки точності можна використовувати два методи валідації: Split-validation та Cross Validation. Перед навчанням будь-якої моделі машинного навчання необхідно відокремити деякі дані, щоб мати можливість перевірити, як модель працює на невідомих їй даних. Це є важливим етапом у будь-якому проекті машинного навчання.

Перехресна перевірка (Cross Validation) – це метод, що використовується для оцінки ефективності прогнозної моделі шляхом поділу набору даних на кілька підмножин та повторного використання різних комбінацій наборів для навчання і перевірки (рис. 3.21).



Рисунок 3.21 – Перехресна перевірка

У машинному навчанні термін «розподіл даних» (рис. 3.22) відноситься до кількох методів, які спеціалісти з обробки даних застосовують перед етапом навчання, щоб розділити набір даних на дві або більше підмножин. Ці підмножини використовуються на різних етапах процесу для навчання, перевірки та тестування моделі.

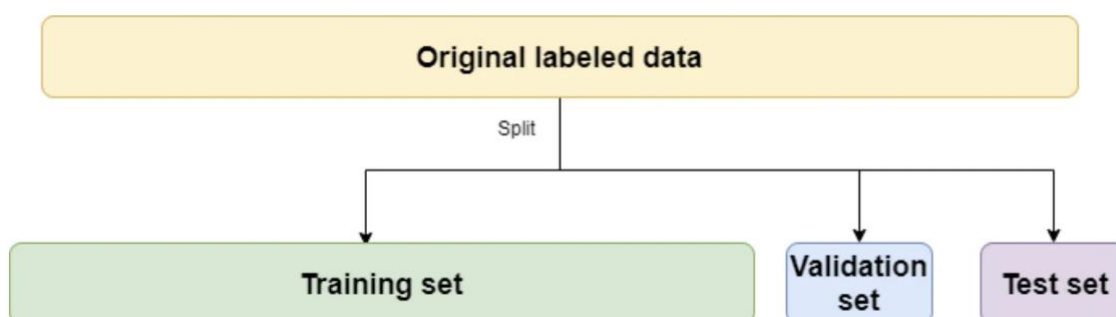


Рисунок 3.22 – Валідація через розподіл даних

З огляду на постійні дискусії у науковій літературі щодо більшої ефективності одного з цих типів валідації, пропонується використовувати обидва методи для порівняння. Такий підхід був застосований, наприклад, у роботі Jobeda Jamal Khanam і Simon Y. Foo (див. рис. 3.23) [31,32].

Classification	Precision	Recall	F-measure	Accuracy
DT (K-fold)	0.739	0.742	0.741	74.24%
DT (Splitting)	0.735	0.731	0.733	73.14%
RF (K-fold)	0.744	0.750	0.746	74.96%
RF (Splitting)	0.779	0.771	0.774	77.14%
NB (K-fold)	0.753	0.755	0.754	75.53%
NB (Splitting)	0.787	0.783	0.785	78.28%
LR (K-fold)	0.761	0.768	0.761	76.82%
LR (Splitting)	0.788	0.789	0.788	78.85%
KNN (K-fold)	0.747	0.751	0.749	75.10%
KNN (Splitting)	0.804	0.794	0.798	79.42%
AB (K-fold)	0.730	0.740	0.730	73.96%
AB (Splitting)	0.792	0.794	0.793	79.42%
SVM (K-fold)	0.761	0.768	0.759	76.82%
SVM (Splitting)	0.774	0.777	0.775	77.71%

Рисунок 3.23 – Порівняння методів машинного навчання двома типами валідації

Відповідно до цієї задачі було обрано датасет «250k Medicines Usage, Side Effects and Substitutes» і метрики точності, Recall, F-measure, Risk probability. Також цей датасет містить вичерпну інформацію про понад 248 тисяч медичних препаратів від усіх виробників, доступних у всьому світі. Дані містять такі деталі, як назви препаратів, активні інгредієнти, терапевтичне використання, дозування, побічні ефекти та замінники. Набір даних має на меті надати корисний ресурс для медичних дослідників, медичних працівників і виробників ліків.

На основі описаного раніше, отримано наступний результат (рис. 3.24) з розрахунку показників.

Algorithm	Accuracy	Risk probability	Recall
KNN (Cross)	76.1%	0.23	0.75
KNN (Split)	76.4%	0.23	0.76
Decision Tree (Cross)	73.3%	0.13	0.74
Decision Tree (Split)	75.5%	0.14	0.73
Random Forest (Cross)	86.6%	0.03	0.79
Random Forest (Split)	87.1%	0.03	0.81
AB (Cross)	73.6%	0.19	0.56
AB (Split)	74.2%	0.2	0.55
Simple Neural Network (Cross)	83.7%	0.24	0.82
Simple Neural Network (Split)	84.1%	0.25	0.83
RNN (Cross)	78.3%	0.22	0.81
RNN (Split)	79.6%	0.22	0.8

Рисунок 3.24 – Результати для порівняння методів

Відповідно до наведеного раніше аналізу, пропонується вибрати найточніші методи, зокрема неймережі та випадковий ліс. Проте, враховуючи нижчу ймовірність ризиків у випадку використання випадкового лісу, основним методом буде використання саме цього алгоритму, який займатиметься класифікацією медичних препаратів.

3.5 Створення методу формування рекомендацій ліків

На основі визначених раніше компонентів майбутнього методу, класифікація лікарських препаратів буде здійснюватися за допомогою алгоритму Random Forest, а для методу фільтрації кращим вибором є використання неймережі типу LSTM, через вищу точність, яка перевершує RNN. Наведені етапи будуть реалізовані за допомогою скрипта (Додаток А), а отриманий метод можна представити через діаграму діяльності алгоритму (рис. 3.25).

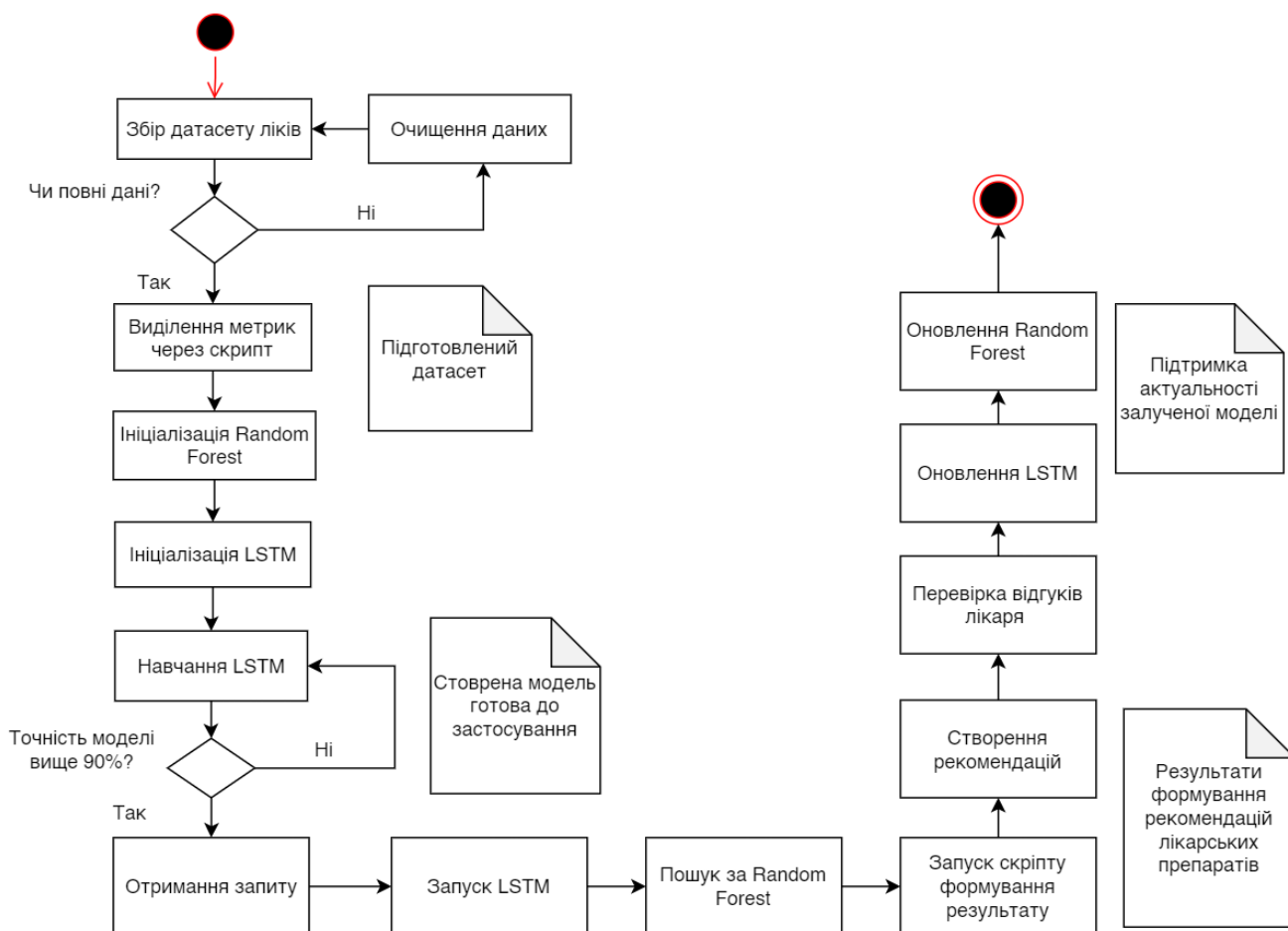


Рисунок 3.25 – Діаграма діяльності методу формування рекомендацій ліків

Цей метод поєднує модель, що включає Random Forest, LSTM та використання бібліотек і скриптів для обробки великих даних (Big Data). Це дозволяє обробляти датасети і окремо представляти список рекомендацій для полегшення інтерпретації. На відміну від більшості інших методів, у цьому випадку застосовується постійне оновлення ключової моделі для підтримання її актуальності, що сприяє підвищенню якості медичних послуг.

Подібний метод був застосований у дослідженні «A Combined Deep CNN: LSTM with a Random Forest Approach for Breast Cancer Diagnosis» [33], яке продемонструвало високу точність, близьку до 100%, у задачах класифікації та прогнозування (рис. 3.26).

Model	Accuracy (%)	Sensitivity (%)
CNN model 1	95.5	91.2
CNN model 2	96.1	91.6
Multilayer perceptron	96.23	93.7
Proposed model	100	99.3

Рисунок 3.26 – Робота подібної моделі у задачах визначення діагнозу

Запропонований метод базується на використанні машинного навчання, зокрема глибокого навчання, що робить модель, яка лежить в основі методу, надзвичайно ефективною у вирішенні завдань, пов'язаних з медичними даними та проблемами сучасної медицини. Застосування методу можна проілюструвати за допомогою наступної діаграми (рис. 3.27).

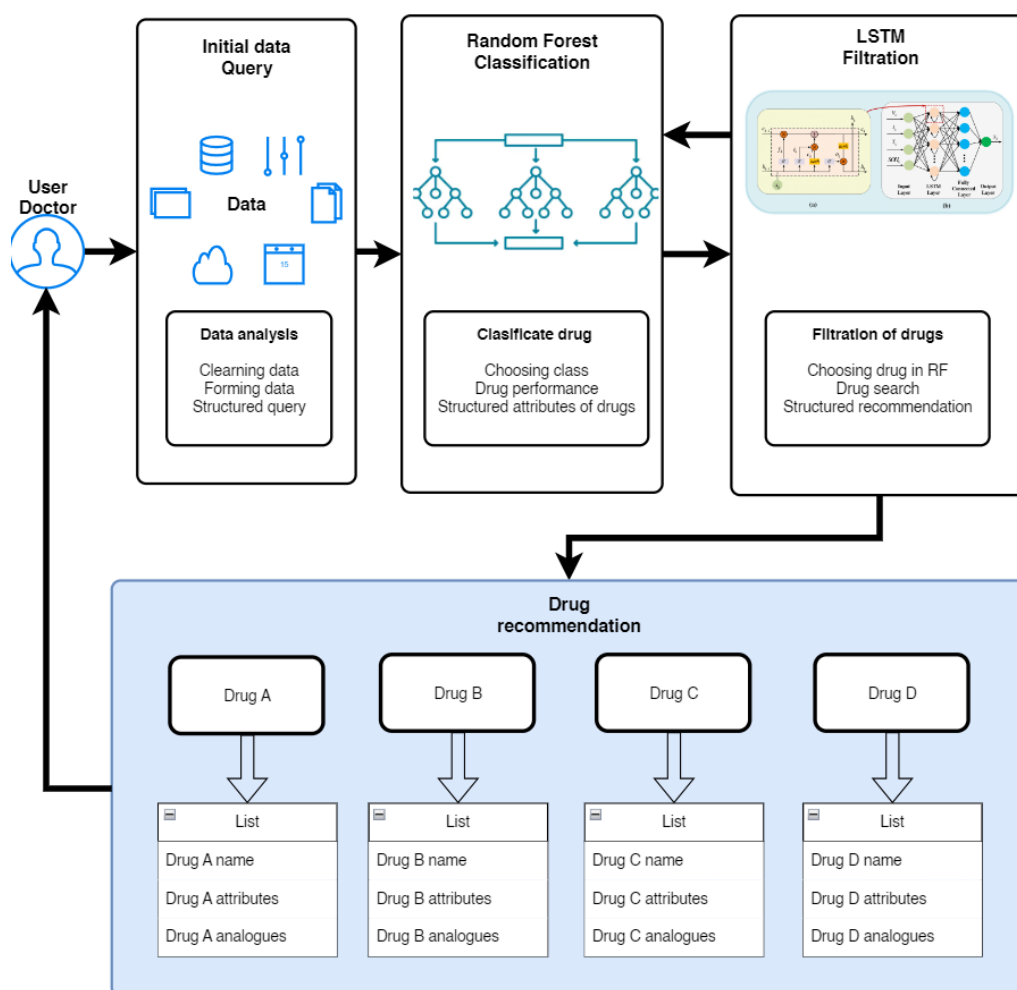


Рисунок 3.27 – Діаграма роботи методу

4 ВАЛІДАЦІЯ СТВОРЕНОГО МЕТОДУ

4.1 Розробка програмної компоненти для проведення експерименту

Для тестування розробленого методу потрібно реалізувати його у програмному коді і порівняти роботу, наприклад, з системою, що здатна обробляти такі ж дані. Тому було обрано мову Python для розробки відповідної компоненти, що буде виконувати завдання формування рекомендації для системи підбору лікарських препаратів. Цей вибір був зумовлений декількома факторами. По-перше, Python відомий своєю швидкістю розробки та простотою, що забезпечує високу ефективність написання коду. По-друге, ця мова програмування має багатий набір бібліотек, які активно використовуються для наукових і медичних досліджень, особливо в областях машинного навчання і обробки даних.

В контексті даного дослідження було вирішено використовувати такі бібліотеки як Pandas, Scikit-learn, та TensorFlow. Pandas, наприклад, є незамінним інструментом для маніпулювання та аналізу даних завдяки своїм високооптимізованим структурам даних і широкому спектру функціональності. Ця бібліотека є незамінною для зручного аналізу великих датасетів, дозволяючи легко фільтрувати, сортувати та виконувати різноманітні агрегації. Scikit-Learn, з іншого боку, надає розробникам зручний доступ до різноманітних алгоритмів машинного навчання, включаючи класифікацію, регресію та кластеризацію, що дозволяє швидко прототипувати та тестувати різні моделі. Її інтерфейс є інтуїтивно зрозумілим та легко інтегрується з іншими бібліотеками Python. TensorFlow, як інструмент для розробки і тренування моделей глибокого навчання, дозволяє реалізувати складні архітектури нейронних мереж, що здатні виявляти складні зв'язки у великих наборах даних. Ця бібліотека визнана за її масштабованість і здатність ефективно обробляти великі обсяги інформації [34, 35].

Відповідна реалізація цього методу представляється як набір у одному Python модулі та його частини коду представлені у додатку А.

В якості аналогів для порівняння ефективності обрано метод, заснований на використанні нейронної мережі, а також комбінований метод, що складається з KNN та нейронної мережі для фільтрації інформації.

4.2 Планування експерименту

Для аналізу та порівняльного оцінювання трьох методів, включаючи розроблений у цій роботі, а також нейромережевий підхід та комбінацію KNN з нейромережею, пропонується здійснити тестовий експеримент для валідації. Відомо, що точність має переважати за 90%, а призначення ліків протипоказаних хворому має становити менше 2% у такого роду системах. Тому план проведення експерименту представлений схематично на рис. 4.1.



Рисунок 4.1 – Схема валідаційного експерименту

Як зазначено у цій роботі, важливим аспектом є вибір датасетів для навчання та тестування розроблених методів. Вибір даних для цих тестів здійснювався через платформу Kaggle, підтримувану Google LLC. Kaggle виступає як динамічна платформа для змагань у сфері аналізу даних, дозволяючи дослідникам і практикам отримувати доступ до численних датасетів, створювати моделі та обмінюватися знаннями. Особливістю Kaggle є її здатність до інтеграції з найсучаснішими

дослідницькими інструментами та технологіями. Наприклад, користувачі можуть використовувати Jupyter notebooks безпосередньо на платформі, що дозволяє їм редагувати, виконувати і ділитися своїм кодом безпосередньо через браузер. Також платформа підтримує різноманітні бібліотеки для машинного навчання, забезпечуючи дослідникам потужні інструменти для обробки та аналізу даних.

Були обрані два датасети:

- «250k Medicines Usage, Side Effects and Substitutes» (рис. 4.2) для підготовки методів і обробки даних про ліки. Цей набір даних містить вичерпну інформацію про понад 248 тисяч медичних препаратів від усіх виробників, доступних у всьому світі. Дані містять такі деталі, як назви препаратів, активні інгредієнти, терапевтичне використання, дозування, побічні ефекти та замітники. Набір даних має на меті надати корисний ресурс для медичних дослідників, медичних працівників і виробників ліків [36];

- «Drug Recommendations» (рис. 4.3) для тестування відповідно до діагнозу підбору ліків. Набір даних про ліки дуже корисний для рекомендацій ліків на основі типу захворювання. Огляд користувачів стане корисним фактором у визначенні настроїв користувачів, побічних ефектів препаратів, а рейтинг буде корисним для призначення препаратів, які дуже підходять для конкретного захворювання та розладів здоров'я. Відповідно з нього береться діагноз і на основі цього ж буде перевірено відповідність сформованої рекомендації методом [37].

medicine_dataset.csv (89.41 MB)



Detail Compact Column

10 of 58 columns

About this file

This file contains all symptoms, usages and alternative medications related to particular medications.


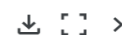
id	name	substitute0	substitute1	substitute2	substiti
Unique Identifier	Name of Medicine	Substitute / Alternative Medicine Name	Substitute / Alternative Medicine Name	Substitute / Alternative Medicine Name	Substitute Medicine I
 1 248k	222825 unique values	[null] 4% Montina-L Tablet 1% Other (236582) 95%	[null] 6% Olox 200mg Tablet 1% Other (232210) 94%	[null] 7% Lecope-M Tablet 1% Other (228366) 92%	[null] Monticope Other (22
1	augmentin 625 duo tablet	Penciclav 500 mg/125 mg Tablet	Moxikind-CV 625 Tablet	Moxiforce-CV 625 Tablet	Fightox Tablet
2	azithral 500 tablet	Zithrocare 500mg Tablet	Azax 500 Tablet	Zady 500 Tablet	Cazithro Tablet

Рисунок 4.2 – Огляд датасету «250k Medicines Usage, Side Effects and Substitues»

Drug_Data.csv (27.3 MB)



Detail Compact Column

6 of 6 columns


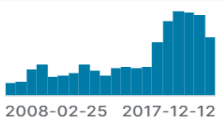

drugName	Prescribed_for	Drug_Review	# User_Rating	Date	# Count
Levonorgestrel 2% Etonogestrel 2% Other (51408) 96%	Birth Control 18% Depression 6% Other (41023) 76%	48280 unique values			
		vomiting, so this was my next ...			
Microgestin Fe 1.5 / 30	Birth Control	"About two months ago I had to switch from LoLestrin Fe to microgestin due to my insurance and pharm...	1	31-Oct-16	4
Wellbutrin	Major Depressive Disorder	"Coming from a very problematic childhood, I've been labeled everything from A to Z by psychiat...	8	06-Mar-15	49
Clonazepam	Anxiety	"Works for me"	6	24-Mar-13	20
Benzoyl peroxide / clindamycin	Acne	"My GP had prescribed this to	4	26-Jan-17	2

Рисунок 4.3 – Огляд датасету «Drug Recommendations»

Також до цього були визначені метрики та методи на основі інших моделей машинного навчання. Тому можна провести тестовий експеримент і проаналізувати отримані результати у наступній частині розділу.

4.3 Аналіз отриманих результатів

Відповідне тестування модулів, що реалізують обрані методи було проведено у IDE Visual Studio Code. Для відображення порівняння було використано окремий модуль, що керує запуском методів, збирає їхні метрики та результати для сумісного відображення окремим вікном.

Тому за планом зображеному на рисунку 4.1, було проведено тестовий експеримент, тобто навчені моделі обраних методів і запусчено їх на датасеті з діагнозами і виписаними ліками, відповідно до цього було отримано результат, що зображено на рисунку 4.4.

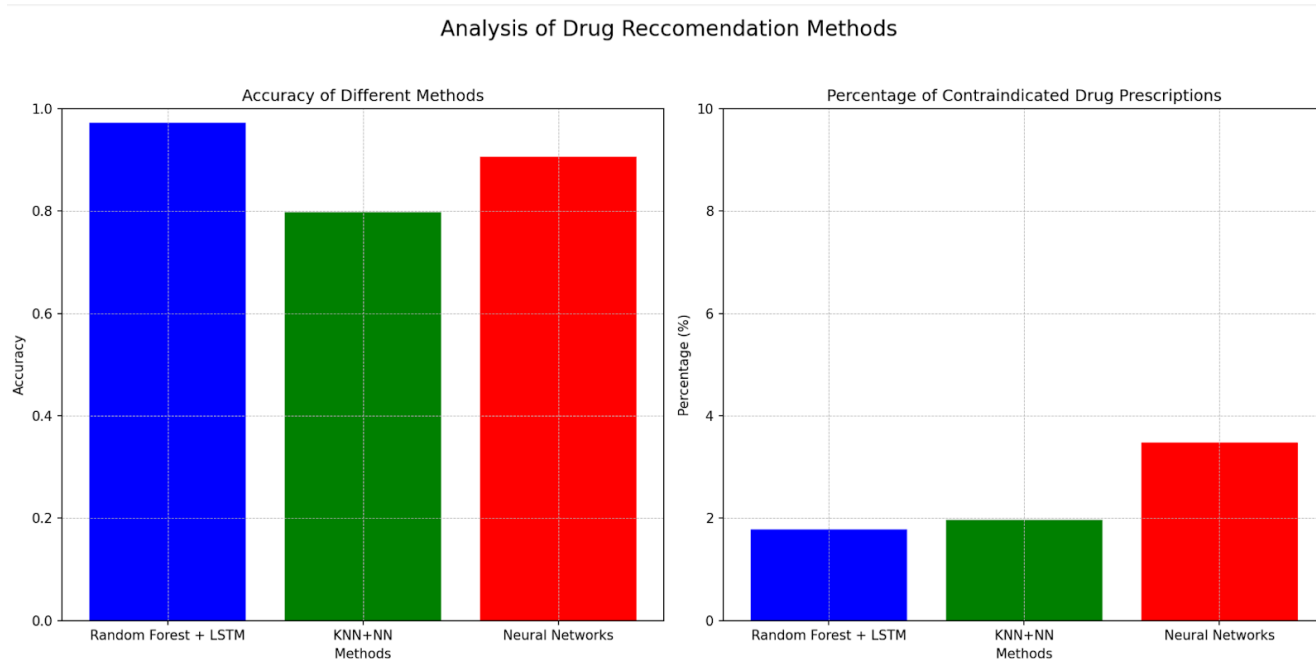


Рисунок 4.4 – Зведення результатів тестового експерименту

У методу Random Forest + LSTM 0.973, тобто 97.3% точність і 1.78 відсоток прописання протипоказаних, аналогічно у KNN + нейромережа – 79.8 % і 1.97%, а у нейромереж – 90.7% і 3.47% відповідно.

Ці результати підтверджують високу точність та низький рівень ризику призначення протипоказаних ліків, що відповідає критеріям, поставленим у другому розділі.

З них видно, що найвищу точність має метод побудований на моделі Random Forest + LSTM, разом з методом на нейромережах, в той час як метод KNN + нейромережа не дав високих результатів. Тобто, з метрики точності тільки два отримали результати вище 90%.

За результатами відсотка визначення протипоказаних рекомендацій, розроблений метод виявився найефективнішим, показуючи найнижчі значення в цій категорії, тоді як підхід на основі нейромереж показав найгірші результати.

Ці результати демонструють потенціал розробленого методу для застосування в системах підбору лікарських препаратів і можливість його подальшого розвитку та удосконалення для підвищення точності та безпеки.

ВИСНОВКИ

Ця науково-дослідна робота підкреслює важливість розробки ефективних методів формування рекомендацій для системи підбору лікарських препаратів, з огляду на постійний розвиток інформаційних технологій у медицині. Актуальність цієї проблеми стає все більш вираженою з розширенням асортименту лікарських засобів і зростанням їхніх складностей і атрибутів.

Проаналізовано не лише саму проблему підбору і рекомендації ліків, а й вивчено численні наукові роботи і публікації, що займаються цією темою. Так були проаналізовані різні методи для формування рекомендацій на основі різних моделей. Визначено, що найкращим варіантом для створення методу є застосування машинного навчання.

Було проведено аналіз різних алгоритмів та моделей машинного навчання визначено найкращі для класифікації та фільтрації і на основі них створено метод. Для методу створено алгоритм його роботи, визначено метрики такі як точність і ризик призначення невідповідних препаратів. В якості оптимального рішення обрано комбінацію методів машинного навчання у поєднанні Random Forest та LSTM.

Використання Python і бібліотек машинного навчання дозволило ефективно реалізувати та випробувати запропонований метод поряд із іншими аналогічними підходами. Результати тестових експериментів підтвердили, що розроблений метод перевершує альтернативні за ключовими показниками точності та безпеки.

Це дослідження не тільки досягло поставлених цілей, а й відкрило шлях для подальшого розвитку і вдосконалення систем рекомендації лікарських препаратів. Подальше впровадження і розширення розробленого методу може значно оптимізувати процес прийняття рішень у медицині, знижуючи навантаження на медичних працівників і покращуючи якість лікування пацієнтів.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Mikulic M. Pharmaceutical market: worldwide revenue 2001-2023, May 22, 2024.
2. Mikulic M. Top pharmaceutical products by sales worldwide 2023, May 22, 2024.
3. Wagner Luiz Ramos Barbosa, Nascimento M. Selecting Medicinal Plants for Development of Phytomedicine and Use in Primary Health. URL https://www.researchgate.net/publication/221922312_Selecting_Medicinal_Plants_for_Development_of_Phytomedicine_and_Use_in_Primary_Health_Care (дата звернення: 11.05.2024).
4. F. Angelo , D. Veenstra, S. Knerr, B. Devine. Prevalence and prediction of medical distrust in a diverse medical genomic research sample. URL: <https://doi.org/10.1016/j.gim.2022.03.007> (дата звернення: 16.05.2024).
5. Proudfoot J.R. Drugs, leads, and drug-likeness: an analysis of some recently launched drugs, *Bioorg. Med. Chem. Lett*, 2002.
6. Fischer J. Timing of analog research in drug discovery *Pharmazie*, 2001.
7. Keikhosrokiani, P., Mustaffa, N. Success factors in developing iHeart as a patient-centric healthcare system: a multi-group analysis. *Telematics Inform.* 35, 753–775, 2018. URL: <https://doi.org/10.1016/j.tele.2017.11.006> (дата звернення: 18.05.2024).
8. Keikhosrokiani, P., Mustaffa, N., Zakaria, N., Abdullah, R. Assessment of a medical information system: the mediating role of use and user satisfaction on the success of human interaction with the mobile healthcare system (iHeart). *Cogn. Technol. Work* 22, 281–305, 2020. URL: <https://doi.org/10.1007/s10111-019-00565-4> (дата звернення: 19.05.2024).
9. Keikhosrokiani, P. Perspectives in the Development of Mobile Medical Information Systems: Life Cycle, Management, Methodological Approach and Application, 2019.

10. Ilvén, H., Savukoski, S.M., Pesonen, P. Association of genetic disorders and congenital malformations with premature ovarian insufficiency: a nationwide register-based study, 2023. URL: <https://doi.org/10.1093/humrep/dead066> (дата звернення: 20.05.2024).
11. Garg, S. Drug recommendation system based on sentiment analysis of drug reviews using machine learning, 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 175–181.
12. Bhat, S., Aishwarya, K. Item-based hybrid recommender system for newly marketed pharmaceutical drugs, International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2013, pp 2107–2111.
13. Xian, Z., Keikhosrokiani, P., XinYing, C., Li, Z. An RFM model using k-means clustering to improve customer segmentation and product recommendation. In: Keikhosrokiani, P. (ed.) Handbook of Research on Consumer Behavior Change and Data Analytics in the Socio-Digital Era, 2022, pp. 124–145. IGI Global, Hershey, PA, USA.
14. Keikhosrokiani, P., Fye, G.M. A hybrid recommender system for health supplement e-commerce based on customer data implicit ratings. *Multimed. Tools Appl*, 2023. URL: <https://doi.org/10.1007/s11042-023-17321-6> (дата звернення: 21.05.2024).
15. Reed T. Sutton, David Pincock, Daniel C. Baumgart, Daniel C. Sadowski, Richard N. Fedorak & Karen I. Kroeker. An overview of clinical decision support systems: benefits, risks, and strategies for success, 2020.
16. Anahtar M.N., Yang J.H., Kanjilal S. Applications of machine learning to the problem of antimicrobial resistance: an emerging model for translational research, 2021. URL: <https://doi.org/10.1128/jcm.01260-20> (дата звернення: 25.05.2024).
17. Висоцький А.А. Суріков О.О. Василюк-Зайцева С.В. Розвиток штучного інтелекту в сучасній медицині, «Український медичний часопис», 2023. URL: <https://doi.org/10.32471/umj.1680-3051.154.241221> (дата звернення: 28.05.2024).
18. HITECH Act Enforcement Interim Final Rule. URL: <https://www.hhs.gov/hipaa/for-professionals/special-topics/HITECH-act-enforcement-interim-final-rule/index.html> (дата звернення: 28.05.2024).

19. Chang, F. & Gupta, N. Progress in electronic medical record adoption in Canada Recherche Les progrès dans l'adoption du dossier médical électronique au Canada, 2015, pp. 1076–1084.
20. Nøhr, C. et al. Nationwide citizen access to their health data: analysing and comparing experiences in Denmark, Estonia and Australia, 2017, pp. 1–11. URL: <https://doi.org/10.1186/s12913-017-2482-y> (дата звернення: 29.05.2024).
21. Matt F. Professional JavaScript for Web Developers, 5th Edition, 2023.
22. Svekis L. JavaScript DOM Projects for Interactive Dynamic Webpages, 2022.
23. Lubanovic B. FastAPI, O'Reilly Media, 2023.
24. Kleppmann M. Designing Data-Intensive Applications, O'Reilly Media, 2017.
25. Ramalho L. Fluent Python, 2nd Edition, O'Reilly Media, 2022.
26. Peterson M. An Introduction to Decision Theory, Cambridge Introductions to Philosophy 1st Edition, 2009.
27. О.І. Кушлик-Дивульська, Б.Р. Кушлик. Основи теорії прийняття рішень, 2014, 94с.
28. Guo, Kristina L. (June 2008). DECIDE: a decision-making model for more effective decision making by health care managers. The Health Care Manager. 2008, pp. 118–127. URL: <https://doi.org/10.1097/hcm.0000000000000299> (дата звернення: 01.06.2024).
29. Sarbani Dasgupta, Banani Saha. Big data analysis on medical field for drug recommendation using apriori algorithm and deep learning. URL: <https://doi.org/10.1007/s11042-024-18832-6> (дата звернення: 01.06.2024).
30. К. Копонова Машинне навчання: методи та моделі, Харківський національний університет імені В. Н. Каразіна, 2020.
31. J.D. Kelleher, Br. Mac Namee, A. D'Arcy. Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies, 2023.

32. J. Rashid, S. Kanwa J. Kim Heart Disease Diagnosis Using the Brute Force Algorithm and Machine Learning Techniques, 2022. URL: <https://doi.org/10.32604/cm.2022.026064> (дата звернення: 03.06.2024).
33. Smith, J.W., Everhart, J.E., Dickson, W.C., Knowler, W.C. Pima Indians Diabetes Database. Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In Proceedings of the Symposium on Computer Applications and Medical Care, 2016, pp. 261-265.
34. A. Begum, V. Dh. Kumar, J. Asghar, D. Hemalatha, G. Arulkumaran A Combined Deep CNN: LSTM with a Random Forest Approach for Breast Cancer Diagnosis, 2022. URL: <https://doi.org/10.1155/2022/9299621> (дата звернення: 03.06.2024).
35. M. Harrison, M. Prentiss, Learning the Pandas Library: Python Tools for Data Munging, Analysis, and Visual, 2016.
36. A. Géron Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition, O'Reilly Media, 2019.
37. Drug Recommendation by type of Disease and Review Analysis. URL: <https://www.kaggle.com/datasets/subhajournal/drug-recommendations> (дата звернення: 05.06.2024).
38. Comprehensive collection of medical drugs info worldwide from various sources: URL: <https://www.kaggle.com/datasets/shudhanshusingh/250k-medicines-usage-side-effects-and-substitutes> (дата звернення: 05.06.2024).
39. Sarker I. H. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science, Vol. 2, no. 3, 2021. URL: <https://doi.org/10.1007/s42979-021-00592-x> (дата звернення: 15.05.2024).
40. Підгорний М. О. РОЗРОБКА РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ ПІДБОР ЛІКАРСЬКИХ ЗАСОБІВ З ВИКОРИСТАННЯМ МЕТОДІВ МАШИННОГО НАВЧАННЯ. РАДІОЕЛЕКТРОНІКА ТА МОЛОДЬ У ХХІ СТОЛІТТІ: МАТЕРІАЛИ ХХVІІІ МІЖНАР. МОЛОДІЖ. ФОРУМУ, м. Харків, 16–18 квіт. 2024 р. Харків, 2024. С. 665–666.