

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ
І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

Факультет інформаційних технологій

УДК 004.4:005.93:614.94

«ПОГОДЖЕНО»

Декан факультету
інформаційних технологій

Болбот І.М., д.т.н., професор

«ДОПУСКАЄТЬСЯ ДО ЗАХИСТУ»

Завідувач кафедри комп'ютерних наук

Голуб Б.Л., к.т.н., доцент

_____ 2024 р.

_____ 2024 р.

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

на тему Програмне забезпечення системи підтримки прийняття рішень
керівництвом собачого притулку

Спеціальність 121 Інженерія програмного забезпечення

(код і назва)

Освітня програма Програмне забезпечення інформаційних систем

(назва)

Орієнтація освітньої програми освітньо-наукова

(освітньо-професійна або освітньо-наукова)

Гарант освітньої програми

доктор технічних наук

(науковий ступінь та вчене звання)

Семко В.В

(підпис)

(ПІБ)

Керівник магістерської кваліфікаційної роботи

кандидат технічних наук, доцент

(науковий ступінь та вчене звання)

Голуб Белла Львівна

(підпис)

(ПІБ)

Виконав

Кищук Олексій Миколайович

(підпис)

(ПІБ студента)

КИЇВ-2024

ЗМІСТ

ВСТУП	3
1 Системний аналіз предметної області.....	9
1.1 Опис предметної області	9
1.2 Аналіз існуючих рішень та їх проблем.....	10
1.3 Постановка завдання для магістерської роботи.....	12
2 Моделювання системи.....	15
2.1 Опис та аналіз методології системного аналізу.....	15
2.2 Загальні відомості про UML та моделювання	17
2.3 Діаграма прецедентів	18
3 Розробка системи	25
3.1 Структура джерела інформації для проведення інтелектуального аналізу ..	25
3.2 Загальні поняття технології Data Mining	31
3.4 Реалізація задач Data Mining	34
3.5 Реалізація ключових показників ефективності	47
3.6 Дані для аналізу	51
4 Результати дослідження	54
4.1 Дослідження використання задач класифікації	54
4.2 Дослідження використання методу асоціативних правил	63
4.4 Дослідження використання методу часових рядів для прогнозування.....	66
4.5 Апаратні та програмні вимоги.....	71
Висновки	73
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	74

ВСТУП

З року в рік проблема безпритульних тварин, зокрема, собак стає все більш актуальною. Збільшення кількості бездомних собак у містах породжує значні проблеми для муніципальних притулків, волонтерських організацій та державних служб. Притулки для собак відіграють вагомую роль у підтримці безпеки громадян і збереженні екологічної рівноваги, забезпечуючи тим самим тимчасове житло для тварин, які не мають домівки. Натомість ефективне управління такими установами часто ускладнюється обмеженими ресурсами та недостатнім фінансуванням. У сучасних умовах, якщо притулки потребують оптимізації своєї діяльності для раціонального використання ресурсів, постає потреба в інструментах, котрі б забезпечували оперативний аналіз та підтримку прийняття управлінських рішень на основі аналізу даних [1].

На разі переважна більшість притулків продовжує використовувати ручні або застарілі підходи для управління ресурсами, що перешкоджає швидко обробляти дані та надавати інформацію для прийняття обґрунтованих рішень. Цей спосіб може породити неефективний розподіл ресурсів, що, в свою чергу, зменшить рівень обслуговування собак, а також збільшить час їхнього перебування в притулку та ускладнить процес їх усиновлення. Недостатній рівень автоматизації процесів ускладнює не лише облік тварин, але й виявлення закономірностей, які б могли сприяти оптимізації роботи притулку та підвищенню ефективності його функціонування.

Актуальність даного дослідження ґрунтується на використанні сучасних технологій обробки та аналізу даних, що може значно підвищити ефективність роботи притулків для собак. Розробка системи підтримки прийняття рішень, котра базується на інтелектуальному аналізі даних *Data Mining* та багатовимірному аналізі *OLAP*, дає можливість інтегрувати в один інтерфейс інформацію про стан собак, процеси усиновлення та інші показники. Дане програмне забезпечення надає змогу керівникам притулків приймати обґрунтовані рішення, швидко адаптуватися до змін, планувати ресурси і

підвищувати рівень якості обслуговування тварин. Актуальність дослідження підтверджується тим, що в умовах обмежених ресурсів, інструменти для аналізу даних допоможуть покращити управління притулками, підвищити рівень усиновлення, зменшити середній час перебування тварин у притулках, а також забезпечити оптимальний розподіл ресурсів.

Об'єктом дослідження є діяльність притулків для безпритульних собак як комплексного процесу управління ресурсами та опіки над тваринами. Домівки для собак виконують важливі соціальні функції, однак водночас мають значні обмеження у фінансуванні та управлінських ресурсах, тому що потребують оптимізації. У цьому контексті предметом дослідження виступає аналіз даних, пов'язаних з тваринами, які потрапляють до притулку, перебувають у ньому, а також знаходять нових господарів.

Метою дослідження являється створення ефективного програмного забезпечення для підтримки прийняття рішень у собачих притулках, що б дозволило керівникам притулків своєчасно отримувати важливу інформацію для оцінки стану тварин, ефективності процесів усиновлення та раціонального використання ресурсів. Дана розроблена система об'єднує сучасні методи аналізу даних і базується на принципах багатовимірного аналізу, а саме *OLAP* і технологіях *Data Mining*. Саме тому, ця система забезпечує можливість швидкого отримання аналітичних звітів, оцінки ефективності управління та ухвалення обґрунтованих рішень.

Завданням дослідження є здійснення аналізу предметної області притулків для безпритульних собак, для того щоб виявити ключові потреби та проблеми, які постають перед керівниками у сфері обліку, управління ресурсами і усиновлення собак. На основі згаданого аналізу необхідно сформулювати вимоги до системи підтримки прийняття рішень, оскільки це дозволить забезпечити своєчасне отримання інформації для ефективного керування собачим притулком.

Наступним завданням є моделювання системи, яка включає розробку функціональної і логічної моделей, крім того побудову діаграм прецедентів та

розгортання. Саме це дозволить відобразити основні процеси, котрі відбуваються в притулку, та забезпечити наочне представлення функціональних компонентів системи і їх взаємодії.

На етапі розробки системи передбачається створення структур для зберігання необхідної інформації про собак та процеси притулку. Також необхідно розробити механізми обробки й передачі даних між оперативною базою даних і сховищем даних. Важливим завданням являється реалізація методів *Data Mining* для виявлення закономірностей у процесах, які відбуваються у притулку, і розробка ключових показників ефективності для оцінки результатів діяльності.

Заключне завдання представляє собою проведення дослідження використання системи в умовах тестового середовища, що включає випробування методів класифікації, асоціативних правил, кластеризації та аналізу часових рядів для прогнозування. Це дозволить оцінити корисність розробленої системи, її вплив на якість управлінських рішень та можливості покращення процесів притулку.

В роботі для проведення аналізу даних, підтвердження гіпотез і розрахунку ключових показників продуктивності застосовується технологія *OLAP*, котра забезпечує можливість швидкого та гнучкого багатовимірного аналізу великих обсягів даних. Тому це особливо важливе для керівництва собачих притулків, тому що потребує оперативного доступу до аналітичної інформації для прийняття обґрунтованих рішень.

Задля формування нових гіпотез і виявлення прихованих закономірностей у даних застосовано методи інтелектуального аналізу даних. До них належать: метод класифікації Наївного Байєса, що є реалізованим за допомогою пакету *ml-naivebayes*, метод кластеризації, який реалізований з використанням пакету *ml-kmeans* й аналіз часових рядів, де використовується пакет *ARIMA*, котрий реалізує відповідний метод аналізу і пошук асоціативних правил, для якого застосовано пакет *node-fpgrowth*. Дані методи дозволяють

знаходити зв'язки та тенденції у даних, які можуть залишатися непомітними при використанні традиційних методів обробки інформації.

Розробка інтерфейсу, а також забезпечення функціональності системи здійснюється з використанням сучасних веб-технологій, зокрема *Node.js*, *Express.js* та *Bootstrap*, що дозволяє створити зручний, масштабований і адаптивний інструмент для підтримки прийняття рішень у собачих притулках. Задля зберігання та управління даними застосовується *MSSQLServer*, тому що забезпечує надійне зберігання великих обсягів інформації. Саме завдяки такому поєднанню технологій досягається ефективний аналіз і відображення даних, бо дозволяє визначати найкращі стратегії управління ресурсами притулку.

В ході дослідження було розроблено систему підтримки прийняття рішень для керівництва собачих притулків. Ця система об'єднує методи інтелектуального аналізу Data Mining (знаходження в сирих даних нових знань) та технологію аналізу даних в режимі реального часу OLAP (підтвердження або спростування знайдених нових знань) у єдине рішення. Розроблене програмне забезпечення, яке адаптовано для специфічних потреб притулків дозволяє виявляти важливі закономірності та оптимізувати процеси усиновлення. Дана програмна система дозволяє отримувати глибокий аналіз даних, тому що вона значно підвищує рівень обґрунтованості рішень, прийнятих керівництвом притулку. Саме тому запропоноване рішення сприяє не лише раціоналізації витрат і вирішення в розумінні потреб в персоналі, але й створенню програм усиновлення для собак, які перебувають у притулку, а також підвищенню ймовірності їх усиновлення.

Апробація відбувалась:

- 1) На науково-практичній конференції студентів і аспірантів «Теоретичні та прикладні аспекти розробки комп'ютерних систем» були опубліковані тези «Використання Data Mining у системі підтримки прийняття рішень керівництвом собачого притулку». Тези наведено у ДОДАТКУ А.

- 2) А також на XV Міжнародній науково-практичній конференції молодих вчених «інформаційні технології: економіка, техніка, освіта» були опубліковані тези «Програмне забезпечення системи підтримки прийняття рішень керівництвом собачого притулку». Тези є наведені у ДОДАТКУ А.

Структура магістерської роботи складається з чотирьох основних розділів, кожен з яких висвітлює окремий етап розробки та дослідження програмного забезпечення системи підтримки прийняття рішень керівництвом у притулках для собак.

У першому розділі представлено системний аналіз предметної області, який починається з опису загальних аспектів і особливостей роботи притулків для собак, та включає процеси, пов'язані з обліком, управлінням ресурсами та доглядом за собаками. Наступним кроком проведено аналіз наявних рішень та патентів, звертаючи увагу на їхні проблеми і недоліки, а також це створює передумови для розробки нової системи. Даний розділ завершується постановкою завдання для магістерської роботи, яке було сформульовано на основі проведеного аналізу.

Другий розділ магістерської роботи присвячений моделюванню системи підтримки прийняття рішень. В ньому описуються обрані методології системного аналізу та підходи до моделювання, які забезпечують структуроване представлення системи. У розділі подано діаграми прецедентів, що відображають основні функціональні сценарії взаємодії користувачів із системою, а також діаграми розгортання, котрі ілюструють архітектурну структуру системи й основні вузли для забезпечення роботи програмного забезпечення.

У третьому розділі робота зосереджена на розробці самої системи. В цьому розділі докладно описано структуру джерела інформації для аналізу, де було включено розробку оперативної бази даних та структури сховища даних, в тому числі механізмів вилучення, обробки та передачі даних. Надалі у розділі розглядаються основні поняття й інструменти *Data Mining*, тому що вони

необхідні для реалізації задач класифікації, кластеризації, аналізу часових рядів, та асоціативних правил, а також для визначення ключових показників ефективності. Особлива увага в цьому розділі приділена обраним технологіям, котрі застосовуються для реалізації функціональних можливостей системи.

Четвертий розділ описує результати дослідження та функціональні випробування системи. Було досліджено використання задач класифікації, таких як: методи *One-Rule* та Наївного Байєса. Крім того розглянуто метод асоціативних правил для пошуку взаємозв'язків при виборі собак. Також виконано дослідження кластеризації для групування даних за спільними характеристиками і застосовано метод часових рядів для прогнозування надходжень та усиновлень собак.

Обсяг та структура даної роботи висвітлюються у записці, яка охоплює 77 аркушів. Записка сформована з чотирьох основних розділів, а також включає в себе 20 рисунків. Крім того вона містить список з 35 використаних джерел, на які зроблено посилання в тексті роботи.

1 СИСТЕМНИЙ АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Опис предметної області

Предметна область дослідження стосується діяльності притулків для собак, які надають тимчасовий прихисток та догляд. Притулки для собак є спеціалізованими установами, в яких волонтери та працівники забезпечують харчуванням, соціалізацією, а також працюють над усиновленням собак, щоб знайти для них постійні домівки. Основна мета функціонування притулку це не лише забезпечення базових потреб тварин, а й створення оптимальних умов для їхнього усиновлення [2].

У притулках відбувається реєстрація новоприбулих тварин, облік тварин ведення стану здоров'я, вакцинацій, поведінкових характеристик [3]. Велика частина цих даних потребує регулярного оновлення та аналізу. Крім того, усиновлення собак є одним з основних індикаторів ефективності роботи прихистку. Метою цих закладі є усиновлення та повернення собак до нормальних умов життя у нових родинах. Через це особливого значення набуває збір та аналіз даних щодо тривалості перебування, вікових і породних характеристик тварин, а також факторів, які впливають на ймовірність усиновлення.

Окрему роль у роботі притулків відіграє розподіл ресурсів, зокрема матеріальних: корм, ліки, медичні послуги та кадрові ресурси. В умовах обмежених ресурсів важливим завданням для керівництва притулків стає оптимізація процесів [4]. Це потребує якісного аналізу та обробки наявних даних. Наприклад, кількість тварин у притулку може змінюватися залежно від сезону. Тому притулки мають бути готовими до періодів значного зростання кількості новоприбулих тварин. В цих випадках стає актуальним планування додаткових ресурсів. В сезонах, коли притулки мають низьке значення усиновлення, вони потребують проведення цільових кампаній з популяризації усиновлення в певні періоди.

Для більш ефективного керування притулком необхідна система, яка б могла автоматично обробляти та аналізувати великі обсяги даних, котрі накопичуються в результаті щоденної роботи [5]. Система підтримки прийняття рішень, що була розроблена для цієї предметної області, повинна забезпечувати можливість моніторингу ключових показників діяльності притулку, надавши аналітичну інформацію про тенденції в усиновленнях та надходженнях. Ця система також повинна дозволяти приймати обґрунтовані рішення на основі глибокого аналізу даних.

Таким чином, предметна область включає не лише фізичну та операційну діяльність притулків, а й аналітичну складову, яка потрібна для ефективних стратегій управління.

1.2 Аналіз існуючих рішень та їх проблем

Під час аналізу існуючих рішень у сфері систем для керівництва притулків для собак, було розглянуто кілька відомих систем, котрі частково відповідають завданням.

Було розглянуто такі рішення: *Shelter Animals Count*, *Microsoft Power BI* які, демонструють певні можливості з управління даними та підтримки прогнозування, але вони мають обмеження щодо цілісного підходу, відносно аналізу великих обсягів даних з використанням методів *OLAP* та *Data Mining* у притулках для собак.

Shelter Animals Count – відома платформа, яка дозволяє притулкам аналізувати показники та прогнозувати результати. В той же момент система здебільшого орієнтована на статистичний облік та базове прогнозування, що охоплює тільки один аспект, який може бути розв’язаний за допомогою методів *OLAP* та *Data Mining*. Зображення інтерфейсу *Shelter Animals Count* на рис. 1.2.1.

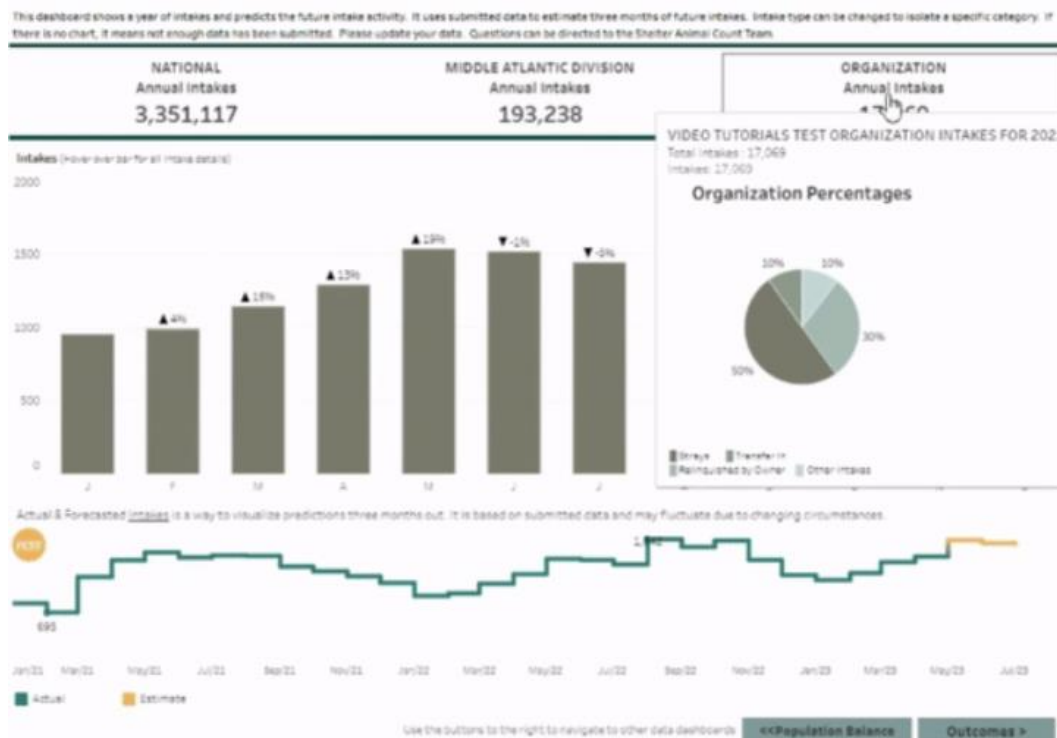


Рис. 1.2.1 Інтерфейс *Shelter Animals Count*

Обмеженість методів знижує ефективність даного рішення для більш глибокого аналізу даних. В тому числі відсутність можливості формувати куби для гнучкого аналізу великих обсягів даних з різних перспектив обмежує її практичну цінність у порівнянні з розробленою системою, яка має інтегровані інструменти для глибшого аналізу й обробки даних.

Microsoft Power BI надає широкі можливості для побудови інтерактивних панелей керування, а також має функції, які можна застосувати до OLAP і Data Mining. Це дозволяє створювати гнучкі звіти на основі багатовимірних кубів. Але *Power BI* невиключно побудований та потребує вищої технічної підготовки користувачів.

Ще одним корисним ресурсом у сфері є патент *Decision Support System for Selection of Dog Breeds*, код патенту – 10.1109/ISRITI.2018.8864327. Цей патент демонструє переваги у сфері вибору порід собак. Хоча цей інструмент надає підтримку при виборі порід, він обмежений виключно цим аспектом, не охоплюючи функцій управління та прогнозування, необхідних для комплексного аналізу притулків.

Наступним рішенням є *Pawlytics*. Хоча це рішення й має елементи *KPI*, але воно зосереджене на управлінні обліком собак. Це забезпечує основні функції для організації роботи притулків, але не включає методи аналізу, такі як *Data Mining*, для глибшого аналізу впливу різних факторів. Головним едоліком є відсутність інтегрованого інструментарію для побудови багатовимірних кубів і реалізації *KPI* на основі даних, які є важливими для стратегічного управління в притулках. Зображення інтерфейсу *Pawlytics* на рис. 1.2.2.

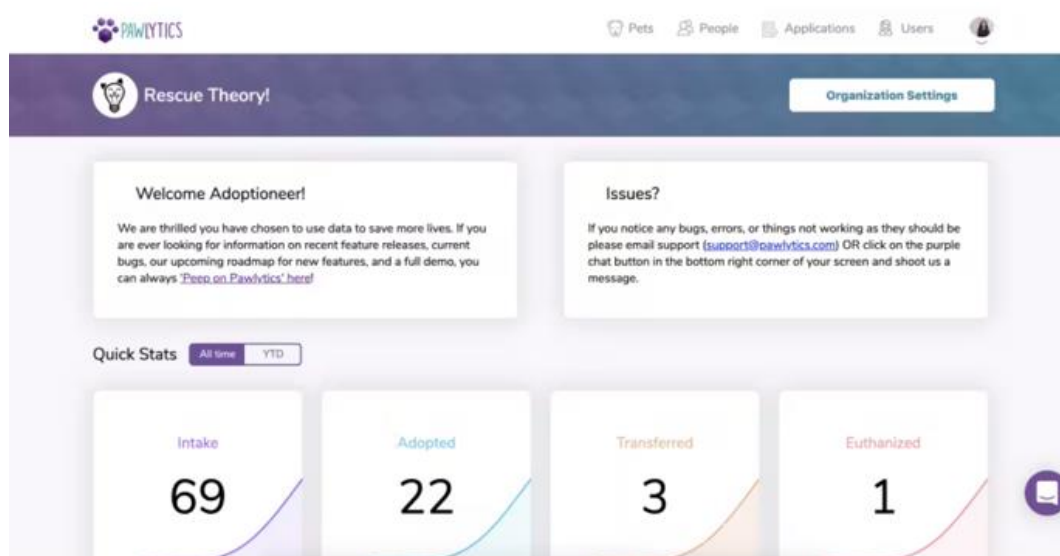


Рис. 1.2.2 Інтерфейс *Pawlytics*

1.3 Постановка завдання для магістерської роботи

Постановка завдання магістерської роботи охоплює комплексну мету розробки системи підтримки прийняття рішень для ефективного управління притулками для собак. Завдання сформульоване таким чином, щоб охопити основні потреби притулків у зборі, зберіганні, аналізі та візуалізації даних. Це повинно дозволити ухвалювати обґрунтовані рішення. У рамках дослідження було визначено ключові гіпотези, які перевіряються за допомогою методів *Data Mining* і *OLAP*, а також *KPI*, котрі підтверджують доцільність і надійність висновків.

Перша гіпотеза досліджує вплив віку собаки на тривалість перебування в притулку до моменту усиновлення. Таким чином завдання дозволяє виявити, як різні вікові групи впливають на ймовірність і швидкість усиновлення, що

важливо для ефективного розподілу ресурсів і організації роботи притулку. Для перевірки цієї гіпотези використовується алгоритм *One-Rule*, що дозволяє проаналізувати вплив віку як основного чинника на тривалість перебування. Додатково застосовується кластеризація для групування собак за тривалістю перебування, це створює основу для виявлення вікових груп, які мають високу ймовірність затримки в притулку. Це, в свою чергу, дозволяє керівникам притулків здійснювати спрямовані дії для просування усиновлення у відповідних вікових групах.

Друга гіпотеза полягає у визначенні шансів на усиновлення для різних порід та статей у певному регіоні. Тим часом, досліджується, чи існують породи, які мають вищі шанси на усиновлення в конкретних регіонах, що може бути пов'язано з культурними або економічними особливостями. Для аналізу цього питання застосовуються асоціативні правила та алгоритм Наївного Байєса, що дозволяють встановити зв'язок між породою собаки, її статтю й ймовірністю усиновлення. Використання асоціативних правил має дозволяти виявляти комбінації порід та статі, котрі частіше всього приводять до успішного усиновлення, а Наївний Байєс допомагає оцінити ймовірність усиновлення для кожної комбінації параметрів. Таким чином це дозволяє керівникам розробляти стратегії, спрямовані на підвищення шансів усиновлення для конкретних порід, тому що це є важливим для збільшення загального рівня усиновлення.

Третя гіпотеза досліджує сезонні коливання у надходженні тварин до притулків у певних регіонах. У даному випадку увага приділяється виявленню сезонних періодів, коли кількість новоприбулих тварин значно зростає або знижується. Для аналізу цієї гіпотези застосовуються методи часових рядів (ARIMA), які дозволяють передбачити потенційні зміни у кількості надходжень протягом року. Виявлення даних сезонних коливань є важливим для оптимізації ресурсів у притулку, зокрема підготовки персоналу та матеріальних запасів на певні періоди. За результатами такого аналізу керівництво може

підготуватися до пікових періодів, для цього забезпечивши належний догляд та обслуговування тварин.

Четверта гіпотеза стосується аналізу сезонних коливань у кількості усиновлень. Цей аспект важливий для визначення періодів, коли усиновлення зменшується чи збільшується, що допомагає формувати стратегії просування та збільшення рівня усиновлень. Користуючись часовими рядами, можна створювати прогнози для кількості усиновлень у різні сезони року. Таким чином це дозволяє керівникам притулків виявляти тенденції, які можуть впливати на динаміку усиновлень, та спрямовувати зусилля на активізацію рекламних заходів, пропонуючи програми, що стимулюють усиновлення в тих періодах, де спостерігається зниження інтересу до тварин.

Кожна з гіпотез перевіряється та підтверджується за допомогою *KPI*. Це забезпечує надійність результатів і дозволяє об'єктивно оцінити ефективність системи. Основні *KPI* включають середній час перебування собак різного віку, ймовірність усиновлення за породою і статтю, кількість надходжень і усиновлень у певні періоди. Наприклад, аналіз середнього часу перебування собак певних вікових груп дозволяє визначити, які групи можуть потребувати додаткової підтримки для прискорення процесу усиновлення. Аналогічно, показники усиновлень за сезонами дозволяють виявити періоди, в які потрібно активізувати рекламні або громадські ініціативи, спрямовані на збільшення кількості усиновлень.

2 МОДЕЛЮВАННЯ СИСТЕМИ

2.1 Опис та аналіз методології системного аналізу

Системний аналіз є одним із ключових етапів створення будь-якої інформаційної системи, оскільки саме на цьому етапі визначаються цілі, вимоги, функціональність та методи реалізації [6]. Основним завданням системного аналізу являється глибоке дослідження предметної області, моделювання її основних процесів та вибір оптимальних інструментів для досягнення поставлених цілей [7]. У контексті розробки системи підтримки прийняття рішень для управління собачими притулками системний аналіз дозволяє ефективно поєднати методології аналізу даних з функціональними вимогами системи.

Системний аналіз базується на двох основних підходах: функціональному та об'єктно-орієнтованому. Кожен з цих підходів має свої особливості, переваги та недоліки, котрі необхідно враховувати залежно від специфіки системи.

Функціональний підхід:

- Орієнтований на аналіз процесів і функцій системи.
- Основна увага приділяється тому, як система працює, які функції вона виконує та які процеси забезпечують досягнення цілей.
- Використовує діаграми потоку даних та діаграми прецедентів.
- Дозволяє чітко зрозуміти, які функції повинна виконувати система, і як ці функції взаємодіють між собою.

Переваги функціонального підходу:

- Легкість у розумінні та застосуванні для простих систем.
- Чіткий фокус на функціональності, що дозволяє деталізувати окремі процеси.
- Можливість швидкого моделювання процесів без глибокого аналізу структури даних.

Недоліки:

- Недостатня увага до структури даних.
- Можливість утруднень при розробці складних систем із багатьма взаємозалежними компонентами.

Об'єктно-орієнтований підхід:

- Базується на об'єктах як основних елементах системи.
- Розглядає дані й функціональність як єдине ціле.
- Основний акцент робиться на структурах даних та їхній взаємодії через методи.
- Використовує такі діаграми *UML*, як діаграми класів, послідовностей та активностей.

Переваги об'єктно-орієнтованого підходу:

- Зручний для розробки великих систем із високим ступенем модульності.
- Полегшує повторне використання компонентів системи.
- Забезпечує краще структурування даних.

Недоліки:

- Вища складність при первинному моделюванні.
- Вимагає глибшого розуміння принципів об'єктно-орієнтованого програмування.

Розробка системи підтримки прийняття рішень для управління собачими притулками передбачає інтеграцію оперативної бази даних із багатовимірним сховищем даних, застосування *OLAP*-аналізу та методів *Data Mining*. Головна мета - забезпечення можливості виявлення закономірностей, прогнозування подій і надання аналітичних даних для прийняття обґрунтованих рішень.

Функціональний підхід є більш доречним для цієї системи, оскільки:

- Система базується на чітко визначених функціях, таких як формування гіпотез, їх перевірка за допомогою *KPI*, а також візуалізація аналітичних даних.

- Основний акцент робиться на обробці даних, моделюванні потоків інформації та функціональних зв'язках між компонентами.
- Функціональний підхід дозволяє зосередитися на аналізі процесів без складного структурування об'єктів.

Об'єктно-орієнтований підхід мав би бути корисним для побудови комплексних програмних систем із багатьма об'єктами, але у даній роботі це може ускладнити моделювання через акцент на функціональних зв'язках.

Застосування функціонального підходу дозволяє чітко окреслити основні процеси системи:

- Моделювання потоків даних між оперативною базою та сховищем даних.
- Визначення взаємодії компонентів для реалізації методів Data Mining.
- Побудова діаграм прецедентів і розгортання, які є зрозумілими й достатніми для подальшого етапу проєктування.

Отже, функціональний підхід є оптимальним для реалізації системи підтримки прийняття рішень у собаких притулках, бо він забезпечує простоту моделювання та дозволяє зосередитися на ключових процесах системи.

2.2 Загальні відомості про UML та моделювання

Основною перевагою *UML* є її здатність об'єднати розробників, аналітиків, менеджерів проєктів і замовників завдяки створенню зрозумілого уніфікованого візуального представлення системи. Це сприяє чіткій комунікації між сторонами проєкту, а також дозволяє уникнути помилок у вимогах і забезпечує ефективне планування етапів розробки[8].

У контексті створення системи підтримки прийняття рішень для собаких притулків *UML* є інструментом для моделювання основних компонентів системи, її функціональності та інфраструктури. Важливим завданням являється вибір тих типів діаграм, які найкраще відображають специфіку

проєкту. У цій роботі зосереджено увагу на діаграмі прецедентів і діаграмі розгортання, бо вони найбільш релевантні для аналізу та проєктування.

Діаграма прецедентів дозволяє відобразити функціональні можливості системи через її взаємодію з користувачами [9]. Дана система орієнтована на керівників притулків, тому що основними прецедентами є аналіз даних про собак за допомогою методів *Data Mining*, побудова прогнозів та оцінка ефективності управління за допомогою ключових показників ефективності. На діаграмі зазначено взаємодію користувачів із системою. Таким чином це дозволяє зрозуміти, які саме функції реалізовані та як вони використовуються в практичній роботі.

Діаграма розгортання демонструє фізичну архітектуру системи, показуючи взаємодію серверів, клієнтів і бази даних [9]. Ця робота відображає структуру, в якій оперативна база даних інтегрується зі сховищем даних, забезпечуючи синхронізацію та агрегацію інформації. Саме це дозволяє чітко уявити, як система працює на апаратному рівні, і забезпечує основи для визначення апаратних і програмних вимог.

В даній роботі використання *UML* є виправданим завдяки його здатності описувати складні системи. Діаграми прецедентів окреслює функціональні можливості системи, забезпечуючи розуміння ключових аспектів її роботи. Діаграма розгортання дає змогу візуалізувати технічну архітектуру.

2.3 Діаграма прецедентів

Діаграма використовується для моделювання функціональних вимог до системи. Діаграма демонструє взаємодію акторів із системою через певні прецеденти — функціональні можливості або процеси, котрі система повинна виконувати. Основна мета діаграми прецедентів полягає в тому, щоб відобразити всі основні функції системи з точки зору кінцевих користувачів, а також забезпечити цілісне розуміння того, як різні компоненти системи взаємодіють один з одним для досягнення певної мети [10].

Діаграма прецедентів складається з двох основних компонентів: акторів, які представляють різних користувачів або інші системи, що взаємодіють з

даною системою, та прецедентів, котрі представляють функціональні можливості системи. Кожен актор взаємодіє з одним або кількома прецедентами, які відповідають його ролі та завданням у рамках системи [11].

На рис. 2.1 представлена спроектована діаграма прецедентів для предметної області.

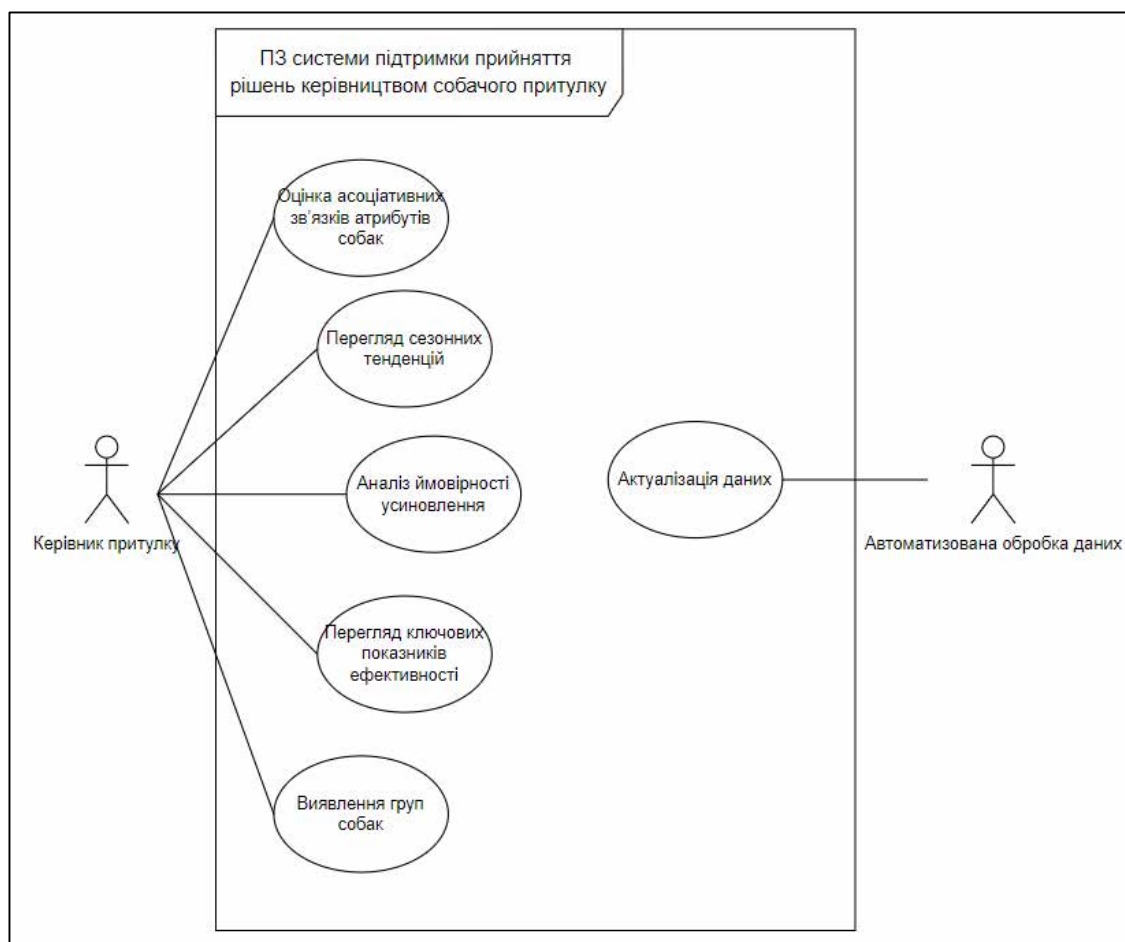


Рис. 2.1 Діаграма прецедентів

На даній діаграмі присутні два актори:

- Керівник притулку — основний користувач системи, що займається управлінням собачого притулку та приймає ключові рішення щодо ефективності роботи притулку, управління ресурсами та моніторингу стану тварин. Він має доступ до аналітичних інструментів, котрі дозволяють йому оцінювати різні аспекти роботи притулку, зокрема аналізувати асоціативні зв'язки, сезонні тенденції, ймовірність усиновлення та переглядати ключові показники ефективності.

- Автоматизована обробка даних — вторинний актор, що представляє автоматизовані процеси обробки даних у системі. Даний актор відповідає за оновлення та актуалізацію даних, що дозволяє підтримувати інформацію в системі актуальною для подальшого аналізу та прийняття рішень.

Прецеденти на діаграмі відображають основні функціональні можливості, які надає система керівнику притулку для прийняття обґрунтованих рішень. Кожен прецедент представляє певний процес або функцію, котра допомагає керівникові притулку краще розуміти та управляти ситуацією в притулку.

- Оцінка асоціативних зв'язків атрибутів собак — цей прецедент надає можливість аналізувати зв'язки між різними атрибутами собак, такими як порода, стать, вік та інші параметри, для виявлення певних закономірностей. Це дозволяє керівникові зрозуміти, які атрибути можуть впливати на ймовірність усиновлення тварин, і може допомогти в розробці спеціальних програм для збільшення шансів на усиновлення певних груп собак.
- Перегляд сезонних тенденцій — функція, яка дозволяє керівнику аналізувати сезонні коливання в кількості надходжень і усиновлень собак. Завдяки цьому керівник може планувати ресурси притулку більш ефективно, наприклад, підготувати додаткові ресурси або персонал у сезони, коли спостерігається збільшення надходжень собак, або навпаки, запроваджувати додаткові рекламні кампанії для стимулювання усиновлення в періоди, коли їх кількість знижується.
- Аналіз ймовірності усиновлення — цей прецедент дозволяє керівнику оцінювати ймовірність усиновлення тварин залежно від їхніх атрибутів. Наприклад, можна визначити, які породи або

вікові групи мають найвищі або найнижчі шанси на усиновлення. Це допомагає керівнику розробляти цільові стратегії для покращення процесу усиновлення, зосереджуючи увагу на певних групах тварин, які потребують більшої підтримки.

- Перегляд ключових показників ефективності — функціональність, яка надає доступ до основних показників ефективності притулку, таких як, середня тривалість перебування тварин, кількість усиновлень, а також інші важливі показники, що відображають ефективність управління притулком. Все це дозволяє керівнику оперативно реагувати на зміни та приймати обґрунтовані рішення, спрямовані на покращення умов утримання тварин і процесу усиновлення.
- Виявлення груп собак — прецедент, який дозволяє групувати собак за певними характеристиками, такими як вік або тривалість перебування. Це дозволяє керівнику притулку більш ефективно управляти ресурсами та ухвалювати рішення щодо пріоритетності тих чи інших груп тварин.
- Актуалізація даних — прецедент, який виконується при запиті до системи та відповідає за регулярне оновлення даних у системі. Завдяки актуалізації даних керівник притулку завжди має доступ до найсвіжішої інформації, яка допомагає йому в прийнятті рішень на основі поточного стану притулку.

2.4 Діаграма розгортання

Діаграма розгортання є інструментом у процесі моделювання [9]. Вона відображає фізичну структуру системи, її компоненти та зв'язки між ними. Діаграма розгортання наочно показує, як елементи системи взаємодіють між собою на рівні серверів, клієнтів та мережевих підключень. Дана діаграма є ключовою для розуміння того, як побудована система підтримки прийняття рішень для керівництва собачого притулку. Дане дозволяє бачити, як

розподіляються функції та як забезпечується обмін інформацією між основними компонентами системи.

У цьому випадку діаграма розгортання відображає кілька основних вузлів і компонентів, які працюють разом для забезпечення ефективної роботи системи.

Клієнтський вузол (*Client*) — цей елемент представляє кінцевого користувача, керівника притулку, який взаємодіє з системою через веб-браузер. За допомогою браузера користувач має доступ до інтерфейсу системи, де може здійснювати запити для перегляду аналітичних даних, що є необхідними для прийняття рішень щодо управління притулком.

Веб-сервер (*Web Server*) — центральний компонент, який обробляє запити від клієнта, виконує бізнес-логіку системи та забезпечує зв'язок між різними рівнями.

Сервер оперативної бази даних (*Database Server*) — тут розташована основна база системи. Вона містить актуальні дані про собак у притулку, їхні атрибути, інформацію про процеси усиновлення тощо. Дані з цієї бази регулярно передаються в сховище даних через *ETL*-процеси для подальшого аналітичного опрацювання.

Сервер сховища даних (*Data Warehouse Server*) — цей сервер містить сховище даних, де зберігаються агреговані дані для багатовимірного аналізу. Використання сховища даних забезпечує можливість зберігання та аналізу великих обсягів історичної інформації, що дозволяє керівництву притулку швидко отримувати детальні звіти, оцінювати ефективність управління та прогнозувати майбутні тенденції.

Веб-сервер включає наступні підсистеми:

Presentation Layer — забезпечує інтерфейс для користувача, відображаючи інформацію у зрозумілому та зручному форматі.

Database Interface — компонент, який здійснює з'єднання з оперативною базою даних та забезпечує доступ до даних.

ETL Service — підсистема, що відповідає за агрегацію даних із оперативної бази даних до сховища даних. Цей процес включає вилучення даних, їхню трансформацію відповідно до вимог багатовимірного аналізу та завантаження в сховище даних (*ETL* — *Extraction, Transformation, Loading*).

Діаграму розгортання продемонстровано на рисунку 2.2.

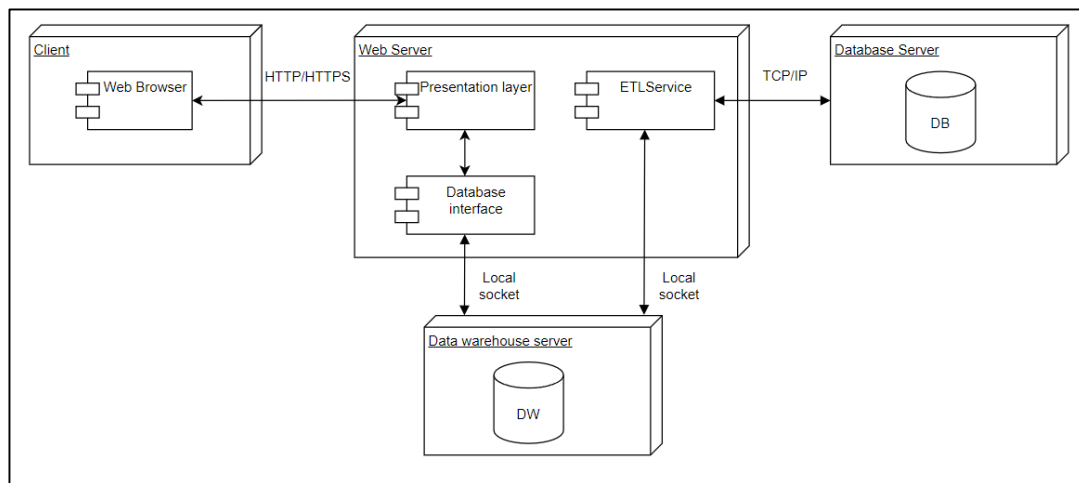


Рис. 2.2 Діаграма розгортання

Діаграма демонструє взаємодію між клієнтом та веб-сервером через *HTTP/HTTPS*-з'єднання, що забезпечує безпечний доступ до інтерфейсу системи. Веб-сервер використовує *TCP/IP*-протокол для зв'язку з сервером оперативної бази даних, де зберігаються основні дані про притулок. Компонент *ETL Service* виконує процеси агрегації, здійснюючи передачу вибірок із оперативної бази даних до сховища даних, що дозволяє регулярно оновлювати інформацію в аналітичному сховищі.

З'єднання між веб-сервером та сервером сховища даних налаштоване через локальний сокет, що дозволяє швидко отримувати доступ до багатовимірних даних для аналітичних запитів і підготовки звітів. Така взаємодія забезпечує інтегроване зберігання даних і дозволяє швидко обробляти запити користувачів щодо аналітичної інформації.

Запропонована архітектура системи є гнучкою та масштабованою. Чіткий розподіл обов'язків між компонентами сприяє підвищенню продуктивності системи та забезпечує надійний доступ до даних. Це дозволяє

керівництву притулку отримувати інформацію для прийняття рішень в режимі реального часу, що є надзвичайно важливим у процесі управління.

3 РОЗРОБКА СИСТЕМИ

3.1 Структура джерела інформації для проведення інтелектуального аналізу

Для забезпечення якісного та своєчасного аналізу даних у системі підтримки прийняття рішень (СППР) собачого притулку необхідна чітко організована структура джерела інформації. Основним джерелом даних є оперативна база, що містить первинну інформацію про кожного собаку, що надходить до притулку, його характеристики, місце знаходження та статус.

Оперативна база даних слугує для накопичення даних і здійснення операційного управління, тоді як її вміст періодично передається до сховища даних. У свою чергу, сховище даних організоване для проведення глибокого аналізу та побудови багатовимірних аналітичних звітів з використанням технологій *OLAP* та *Data Mining* [12].

Оперативна база даних і сховище даних функціонують у взаємодії, тому оперативна база забезпечує накопичення даних у реальному часі, а сховище даних виконує аналітичні задачі та зберігає агреговану інформацію для подальшого інтелектуального аналізу. Даний підхід забезпечує відокремлення поточних операційних процесів від аналітичних, що дозволяє досягти високої продуктивності системи та забезпечити коректність і узгодженість даних у процесі їх обробки.

3.1.1 Структура оперативної бази даних. Оперативна база даних є основним джерелом інформації для збору та зберігання поточних даних про собак у притулку. Вона реалізована на основі системи керування базами даних *Microsoft SQL Server* з використанням мови *T-SQL*. Ця структура дозволяє ефективно працювати з великими обсягами даних, забезпечуючи швидкий доступ та надійність збереження інформації [13].

Структура складається з декількох таблиць. Кожна таблиця має специфічну функцію і містить відповідні дані. Основні таблиці включають інформацію про породи собак, їхні місця знаходження, стать, а також деталі

кожної окремої тварини, такі як ім'я, вага, вік, колір, дата прибуття та дата усиновлення. Наведений код для створення структури бази даних представлений у ДОДАТКУ Б.

Основні таблиці структури оперативної бази даних включають:

- *Breeds* – зберігає інформацію про породи собак. Основне поле *BreedId* служить первинним ключем, а *BreedName* містить назву породи.
- *Locations* – зберігає інформацію про місцезнаходження притулку, включаючи місто та регіон. *LocationId* є первинним ключем, а поля *City* і *Region* вказують відповідно на місто та регіон.
- *Genders* – таблиця, що містить перелік можливих значень статі тварини. Основне поле *GenderId* служить первинним ключем, а *GenderName* зберігає значення статі.
- *Dogs* – основна таблиця для зберігання даних про кожного собаку в притулку.

Основні поля таблиці *Dogs*:

- *DogId* – унікальний ідентифікатор собаки.
- *DogName* – ім'я собаки.
- *BreedId* – зовнішній ключ, що вказує на таблицю *Breeds*.
- *LocationId* – зовнішній ключ, що вказує на таблицю *Locations*.
- *GenderId* – зовнішній ключ, що вказує на таблицю *Genders*.
- *DateArrived* – дата прибуття собаки в притулок.
- *DateAdopted* – дата усиновлення собаки, яка може бути *NULL*, якщо собака ще не була усиновлена.
- *DogWeight* – вага собаки.
- *Color* – колір собаки.
- *Age* – вік собаки на момент прибуття.
- *Sterilization* – поле, що відображає, чи стерилізована собака.

3.1.2 Загальні поняття за напрямку *OLAP*-технології. *Online Analytical Processing* — це технологія багатовимірного аналізу даних. Вона забезпечує зручне представлення, агрегацію та аналіз великих обсягів інформації. Основна перевага *OLAP* полягає у можливості швидкого виконання аналітичних запитів, які базуються на даних, агрегованих у спеціально створених структурах, званих *OLAP*-кубами. Це дозволяє отримувати необхідну інформацію з різних ракурсів і рівнів деталізації.

Технологія *OLAP* використовується для перевірки сформульованих гіпотез і аналізу даних. Вона дає змогу створювати багатовимірні моделі, що відображають ключові аспекти предметної області, такі як: вікові групи, породи, місця розташування, а також часові параметри прибуття і усиновлення. *OLAP* забезпечує зручний інструментарій для аналізу великих обсягів даних у різних зрізах, що сприяє ухваленню більш обґрунтованих рішень.

Однією з основних функцій *OLAP* є агрегація даних. Для забезпечення аналітики дані з оперативної бази переносяться до сховища, де вони агрегуються у формат, придатний для багатовимірного аналізу. Тому це дозволяє отримувати загальні показники, такі як середній час перебування собак у притулку чи кількість усиновлень у певному регіоні, а також виявляти залежності між різними параметрами.

Під час аналізу за допомогою *OLAP* створюються багатовимірні моделі, які дозволяють працювати з даними на різних рівнях деталізації. Наприклад, можна визначити загальні тенденції у тривалості перебування собак у притулку або заглиблюватися в деталі, такі як залежність тривалості перебування від віку чи породи. Завдяки цьому можна ефективно оцінювати стан процесів, виявляти проблемні зони та планувати ресурси.

Ще одним важливим аспектом є використання *OLAP* для розрахунку ключових показників ефективності. Це дозволяє перевіряти гіпотези, отримані за допомогою методів *Data Mining*. Наприклад, до *KPI* можуть належати середній час перебування собак певного віку, кількість тривалих перебувань,

сезонні коливання у кількості усиновлень та інші показники, які надають цінну інформацію для аналізу.

3.1.3 Структура сховища даних. Сховище даних є центральним елементом системи підтримки прийняття рішень, що забезпечує зберігання, організацію та обробку великих обсягів інформації, необхідних для виконання багатовимірного аналізу. Його основне призначення полягає у створенні умов для ефективного використання технологій *OLAP* та *Data Mining*, які реалізовані в системі [12].

У проєкті використовується багатовимірна модель сховища даних. Ця модель включає фактологічні та вимірювальні таблиці. Основна мета цієї структури — підтримка аналізу даних, таких як кількість прибуттів і усиновлень собак, середня тривалість їх перебування в притулку, а також сезонні коливання цих процесів. Усі ці дані агрегуються з оперативної бази, проходять етапи обробки та трансформації, після чого інтегруються у сховище.

Для зберігання даних у сховищі реалізовано такі таблиці вимірів:

- *BreedDimension* — містить інформацію про породи собак. Ця таблиця дозволяє аналізувати дані за породами, наприклад, визначати найпопулярніші породи для усиновлення.
- *GenderDimension* — описує стать тварин. Вона дає змогу оцінювати залежності між статтю та іншими параметрами, такими як тривалість перебування чи ймовірність усиновлення.
- *LocationDimension* — включає інформацію про міста та регіони, що дозволяє здійснювати регіональний аналіз, наприклад, виявляти притулки з найвищими показниками усиновлення.
- *ArrivalTimeDimension* та *AdoptionTimeDimension* — забезпечують багатовимірний аналіз за часовими параметрами. Вони дозволяють аналізувати сезонні тенденції, визначати пікові періоди надходжень чи усиновлень собак.

Фактологічні таблиці є основою для зберігання кількісних даних та виконання розрахунків:

- *ArrivalFacts* — включає інформацію про прибуття собак до притулку. У цій таблиці зберігається кількість прибуттів, параметри віку, ваги, а також тривалість перебування.
- *AdoptionFacts* — зберігає дані про усиновлення собак, включаючи їхні характеристики, такі як порода, стать, вік, а також кількість усиновлень за різними періодами часу.

Кожна з фактологічних таблиць містить посилання на відповідні таблиці вимірів, що забезпечує можливість багатовимірного аналізу. Наприклад, можна легко визначити кількість собак певної породи, які були усиновлені у конкретному регіоні в певний період часу.

Для створення сховища було використано *MS SQL Server*, а реалізація структури виконана за допомогою мови *T-SQL*. Весь код створення сховища даних наведено в ДОДАТКУ Б. Структура побудована таким чином, щоб забезпечити гнучкість і масштабованість системи, що дозволяє додавати нові виміри або показники за потреби.

3.1.4 Механізм вилучення, обробки і передачі даних.

Процес вилучення, обробки та передачі даних є ключовим елементом системи підтримки прийняття рішень. Цей механізм забезпечує переміщення інформації з оперативної бази даних до сховища даних, де вона використовується для багатовимірного аналізу та перевірки гіпотез. Для реалізації цього процесу використовується модуль *ETL* – «*Extract, Transform, Load*», що виконує три основні функції: вилучення даних, їх трансформацію та завантаження в структуру сховища даних.

При виконанні користувачем дії "оновити дані" запускається відповідний код, котрий здійснює повну агрегацію даних. Основною метою цього процесу є інтеграція даних, очищення таблиць сховища та наповнення їх актуальною інформацією з оперативної бази даних.

1. Вилучення даних. Дані з таблиць оперативної бази, такі як інформація про породи, статі, локації, а також про прибуття та усиновлення собак, вилучаються за допомогою *SQL*-запитів. Підключення до бази даних здійснюється за допомогою драйвера *msnodesqlv8*.
2. Очищення таблиць сховища даних. Перед початком перенесення даних виконується повне очищення таблиць сховища, щоб уникнути дублювання або конфліктів у даних. Лічильники *IDENTITY* також скидаються, щоб забезпечити послідовність записів.
3. Трансформація даних. У процесі перенесення даних виконується їх трансформація для відповідності структурам таблиць сховища. Наприклад, для таблиць часу *ArrivalTimeDimension* і *AdoptionTimeDimension* обчислюються додаткові параметри, такі як день, місяць, квартал, сезон і рік.
4. Наповнення вимірів. На першому етапі дані переносяться у таблиці вимірів, зокрема *BreedDimension*, *GenderDimension*, *LocationDimension*, *ArrivalTimeDimension* та *AdoptionTimeDimension*. Це гарантує, що всі зовнішні ключі в фактологічних таблицях матимуть посилання на відповідні виміри. Логіка перенесення вимірів забезпечує точність і відповідність кожного запису.
5. Наповнення фактологічних таблиць. Після завершення наповнення таблиць вимірів дані переносяться до фактологічних таблиць *ArrivalFacts* та *AdoptionFacts*. Цей етап передбачає розрахунок тривалості перебування собак у притулку. Для собак, які ще не були усиновлені, тривалість перебування обчислюється як різниця між максимальною встановленою датою і датою прибуття. Визначення статусу собак на основі заповнення поля *DateAdopted* в оперативній базі.

Код, що відповідає за механізм *ETL*, написано мовою *JavaScript* із використанням пакета *msnodesqlv8* для роботи з *SQL Server*. Усі етапи автоматизовано, включно з формуванням нових записів у таблицях часу, розрахунком тривалості перебування собак у притулку та визначенням їх статусу.

Логіка процесу забезпечує:

- Гнучкість у додаванні вимірів і фактів. Наповнення вимірів передусє створенню записів у фактологічних таблицях, що забезпечує зв'язність даних.
- Цілісність даних. Використання зовнішніх ключів у фактологічних таблицях гарантує відповідність кожного запису в вимірах і фактах.
- Автоматизацію обчислень. Наприклад, сезон і квартал обчислюються автоматично для кожного запису часу.

Механізм *ETL* є критично важливим для забезпечення актуальності даних у сховищі. Він дозволяє автоматизувати оновлення інформації, необхідної для виконання багатовимірного аналізу, перевірки гіпотез і розрахунку *KPI*. Це знижує ризик людських помилок і забезпечує оперативний доступ до точної та структурованої інформації.

Код реалізації цього процесу наведено в ДОДАТКУ Б.

3.2 Загальні поняття технології Data Mining

Інтелектуальний аналіз даних є важливою складовою сучасних інформаційних систем, що спрямовані на виявлення прихованих закономірностей, тенденцій і взаємозалежності у великих обсягах даних [14]. Цей процес базується на використанні математичних, статистичних і машинних методів для перетворення необроблених даних у цінну інформацію. Завдяки *Data Mining* можна отримувати нові знання, які не є очевидними при звичайному перегляді даних, що робить цю технологію критично важливою для багатьох галузей, включаючи управління ресурсами притулків для тварин [15].

Data Mining починається з підготовки даних, яка включає очищення, нормалізацію та агрегування інформації з різних джерел. Цей етап є фундаментальним, адже якість отриманих результатів залежить від того, наскільки правильно підготовлено початковий набір даних [15]. У розробленій системі ці дані агрегуються зі сховища, що отримує їх із оперативної бази. Підготовлені дані потім аналізуються за допомогою різних методів, таких як класифікація, кластеризація, асоціативні правила та часові ряди, що дозволяє виявляти ключові залежності між параметрами.

Методи класифікації використовуються для розподілу об'єктів за категоріями на основі певних характеристик [16]. У проєкті реалізовано алгоритми *One-Rule* і Наївного Байєса. Алгоритм *One-Rule* дозволяє визначити найбільш впливові атрибути, наприклад, вік, що впливає на тривалість перебування собак у притулку. Наївний Байєс дає змогу прогнозувати ймовірність усиновлення собак, враховуючи такі параметри, як порода та стать.

Кластеризація застосовується для поділу об'єктів на групи (кластери) за схожими характеристиками [17]. У системі цей метод використовується для аналізу тривалості перебування собак у притулку залежно від вікових категорій або інших характеристик. Наприклад, кластеризація може показати, що певна група собак старшого віку має більш довгий період перебування, що потребує додаткової уваги з боку притулків.

Асоціативні правила дозволяють встановлювати зв'язки між різними характеристиками, наприклад, породою, статтю та ймовірністю усиновлення. Це дозволяє притулкам краще розуміти, які комбінації факторів сприяють швидшому усиновленню собак і створювати відповідні програми для популяризації таких тварин.

Аналіз часових рядів є важливим для прогнозування сезонних тенденцій, таких як зростання кількості надходжень собак у притулки в певні місяці року або спад активності усиновлення в інші періоди. У нашому проєкті цей метод дозволяє заздалегідь планувати ресурси притулків, розподіл персоналу та організацію рекламних кампаній для стимулювання усиновлень.

3.3 Огляд методів технології Data Mining

Для реалізації задач *Data Mining* у межах проєкту використовуються сучасні інструменти, деяких з яких базуються на методах машинного навчання. Використані бібліотеки забезпечують високу продуктивність, зручність інтеграції з *Node.js* та широкий спектр аналітичних можливостей. Їх правильний вибір став ключовим фактором успіху проєкту.

Одним із ключових інструментів є пакет *ml-naivebayes*, який реалізує алгоритм Наївного Байєса. Цей пакет є потужним засобом для класифікації даних, особливо коли йдеться про ймовірнісні моделі. Завдяки оптимізації під *Node.js*, він швидко обробляє великі обсяги категорійних даних і забезпечує можливість обчислення ймовірностей для багатьох класів одночасно. *ml-naivebayes* включає методи для навчання моделей на основі вхідних даних і прогнозування вихідних значень із високою точністю [16].

Метод класифікації *One-Rule* реалізований на основі нативних можливостей *Node.js* для роботи з даними. Цей підхід базується на аналізі одного атрибута, що значно спрощує процес аналізу та виявлення основних закономірностей у даних. У проєкті він використовується для оцінки впливу віку, породи, статі та локації на тривалість перебування собак у притулку.

Для кластеризації використовувався пакет *ml-kmeans*, який забезпечує реалізацію алгоритму К-середніх. Цей пакет спеціалізується на поділі даних на групи, використовуючи метод найменших відстаней до центроїдів кластерів. Його інтеграція з *Node.js* дозволяє обробляти багатовимірні набори даних із різною кількістю кластерів. Ключовими перевагами *ml-kmeans* є простота використання, висока швидкість роботи та можливість конфігурації параметрів, таких як кількість ітерацій або початкові точки для центрів кластерів.

Для виконання асоціативного аналізу застосовується бібліотека *node-fpgrowth*, яка реалізує алгоритм *Frequent Pattern Growth*. Цей метод є одним із провідних у виявленні частих шаблонів у великих наборах даних. Бібліотека дозволяє ефективно знаходити взаємозв'язки між характеристиками собак, наприклад, порода, стать та ймовірність усиновлення. *FP-Growth*

оптимізований для роботи з великими обсягами даних, що робить його ідеальним інструментом для задач даного проєкту.

Для прогнозування сезонних коливань використовується бібліотека *ARIMA*, яка базується на моделі *ARIMA - Autoregressive Integrated Moving Average*. Цей інструмент є основним методом аналізу часових рядів і широко застосовується для прогнозування в аналітиці [17]. У рамках проєкту, пакет використовується для передбачення кількості надходжень собак у притулки залежно від сезонності. Завдяки використанню *ARIMA* вдалося визначити пікові періоди, що сприяє ефективному плануванню ресурсів притулків.

Крім основних бібліотек, у проєкті використовуються допоміжні інструменти, такі як *moment* і *math.js*. *Moment* використовується для роботи з датами задля розрахунку тривалості перебування собак у притулку. *Math.js* забезпечив виконання математичних операцій, включно зі статистичними обчисленнями, необхідними для аналізу даних.

Для візуалізації результатів аналізу інтегровані бібліотеки *Chart.js* та *D3.js*. *Chart.js* надає змогу створювати статичні графіки для демонстрації основних висновків. *D3.js*, своєю чергою, забезпечує створення інтерактивних елементів для візуалізації.

3.4 Реалізація задач Data Mining

Інтелектуальний аналіз даних, або *Data Mining*, є ключовим компонентом розробленої системи для підтримки прийняття рішень у притулках для собак. Він дозволяє як формувати гіпотези, так і перевіряти їх за допомогою спеціалізованих алгоритмів. У цьому контексті *Data Mining* забезпечує глибокий аналіз даних, виявлення закономірностей і прогнозування. У розробленій системі було реалізовано кілька задач, які охоплюють різні аспекти аналітики та управління даними.

Першою задачею є класифікація собак за віком і тривалістю перебування. Для цього використовується алгоритм *One-Rule*, який ідентифікує найзначніший фактор, що впливає на тривалість перебування. Аналізуючи атрибути, такі як вік, порода та стать, цей метод допомагає визначити, які групи

собак найчастіше залишаються в притулку довше, що дає змогу розробити більш цілеспрямовані стратегії усиновлення.

Друга задача стосується оцінки ймовірності усиновлення собак залежно від їхніх характеристик. Метод Наївного Байєса використовується для обчислення ймовірності усиновлення, враховуючи породу та стать. Це дозволяє робити висновки про те, які категорії собак мають вищі шанси бути усиновленими, а також сприяє прийняттю рішень щодо пріоритетів у просуванні певних груп.

Аналіз асоціацій є ще однією важливою задачею, реалізованою в системі. За допомогою алгоритму *FP-Growth* досліджуються взаємозв'язки між породою, статтю та фактом усиновлення. Цей підхід дає змогу знаходити часті шаблони в даних.

Прогнозування сезонних тенденцій є наступною реалізованою задачею. Модель *ARIMA* використовується для аналізу часових рядів, що дозволяє передбачати кількість надходжень і усиновлень собак у різні періоди. Цей підхід допомагає керівництву притулку готуватися до пікових навантажень і ефективніше розподіляти ресурси.

Останньою задачею є кластеризація собак залежно від тривалості їх перебування. Алгоритм *K-Means* дозволяє поділити дані на групи, що відображають довготривалі та короткотривалі перебування. Такий підхід допомагає сегментувати собак за їх характеристиками та визначити, які категорії потребують додаткової уваги.

3.4.1 Реалізація алгоритму *One-Rule*. Алгоритм *One-Rule* належить до категорії простих алгоритмів класифікації. Його мета — знайти атрибут, який найкраще передбачає цільову змінну. Цей алгоритм підраховує кількість помилок для кожного можливого значення атрибуту, щоб обрати той, який мінімізує похибки класифікації [18].

Алгоритм *One-Rule* працює шляхом аналізу кожного атрибуту в наборі даних для обчислення метрики, наприклад, частоти чи середнього значення, створення правил для кожного можливого значення атрибуту, оцінки точності

цих правил через обчислення частки правильних передбачень і, зрештою, вибору атрибуту з найменшою кількістю помилок для класифікації [18].

У реалізованій системі даний алгоритм використовується для аналізу впливу різних атрибутів на тривалість перебування собак у притулку. Основними досліджуваними атрибутами є вік собаки, порода, стать та локація притулку. Для цих атрибутів визначаються ключові залежності, де вихідні дані включають такі поля: "Тривале перебування" чи "Коротке перебування". Класифікація залежить від порогового значення — 125 днів. Якщо тривалість перебування перевищує поріг, собака відноситься до категорії "Тривале перебування", а якщо ні чи дорівнює — "Коротке перебування".

SQL-запит для аналізу впливу віку на тривалість перебування збирає дані за віком, підраховуючи загальну кількість собак для кожної вікової групи, кількість собак із тривалим та коротким перебуванням, а також обчислює помилкову класифікацію. У запиті використовується умовна логіка для класифікації перебування - "тривале" чи "коротке" на основі порівняння кількості записів, що належать до цих категорій. А також розраховується частка неправильних класифікацій для кожного віку, яка представляє співвідношення невірно класифікованих випадків до загальної кількості собак у відповідній віковій групі. Приклад *SQL* запиту можна побачити на рисунку 3.1. Повний код реалізації всіх атрибутів наведений в ДОДАТКУ Б.

Реалізація аналізу включає кілька ключових етапів. Спочатку здійснюється агрегація даних: за допомогою *SQL*-запиту дані групуються за віковими групами собак, розраховується кількість випадків тривалого та короткого перебування для кожної групи.

Потім проводиться визначення класифікації, де для кожної вікової категорії перевіряється, яке значення — тривале чи коротке перебування — переважає, і на основі цього встановлюється класифікація.

Наступним етапом є розрахунок коефіцієнта помилок, яке обчислюється як відношення неправильно класифікованих випадків до загальної кількості собак у відповідній групі.

Зрештою, отримані результати передаються на клієнтську частину для візуалізації в таблиці, що дозволяє здійснювати подальший аналіз та приймати рішення.

```

SELECT
  af.DogAge AS DogAge,
  COUNT(*) AS TotalCount,
  SUM(CASE WHEN af.LengthOfStay > 90 THEN 1 ELSE 0 END) AS LongStayCount,
  SUM(CASE WHEN af.LengthOfStay <= 90 THEN 1 ELSE 0 END) AS ShortStayCount,
  CASE
    WHEN SUM(CASE WHEN af.LengthOfStay > 90 THEN 1 ELSE 0 END) >
         SUM(CASE WHEN af.LengthOfStay <= 90 THEN 1 ELSE 0 END)
    THEN N'Тривале перебування'
    ELSE N'Коротке перебування'
  END AS Classification,
  CASE
    WHEN SUM(CASE WHEN af.LengthOfStay > 90 THEN 1 ELSE 0 END) >
         SUM(CASE WHEN af.LengthOfStay <= 90 THEN 1 ELSE 0 END)
    THEN CAST(SUM(CASE WHEN af.LengthOfStay <= 90 THEN 1 ELSE 0 END) AS FLOAT) / COUNT(*)
    ELSE CAST(SUM(CASE WHEN af.LengthOfStay > 90 THEN 1 ELSE 0 END) AS FLOAT) / COUNT(*)
  END AS ErrorRatio
FROM
  AdoptionFacts af
WHERE
  af.DogAge IS NOT NULL
GROUP BY
  af.DogAge;

```

Рис. 3.1 Приклад *SQL*-запиту для реалізації *One Rule* за атрибутом вік для класифікації перебування собаки

Результати аналізу надаються у вигляді таблиці, де для кожної вікової категорії собак відображаються ключові дані: загальна кількість собак у цій групі, кількість випадків тривалого перебування, короткого перебування, а також вказується класифікація для кожної категорії, яка визначає домінуючий тип перебування. Додатково розраховується відсоток неправильно класифікованих випадків. Цей відсоток дозволяє оцінити точність правила.

3.4.2 Реалізація методу Наївного Байєсу. У системі для притулку собак цей метод використовується для аналізу впливу породи та статі на ймовірність усиновлення собак. Завдяки цьому методу вдається виявити категорії собак, які мають вищі шанси бути усиновленими, а також створити відповідні програми для усиновлення.

Метод Наївного Байєса базується на теоремі Байєса, яка формулює ймовірність події C , наприклад, "усиновлений" за умови наявності характеристик X , наприклад, породи та статі [19]. Метод називається

"Наївним", оскільки він припускає незалежність усіх характеристик X , що спрощує обчислення [20].

Інформація про собак отримується з таблиці *ArrivalFacts*, яка включає поля, що містять ідентифікатори породи, статі та статусу усиновлення. *SQL*-запити об'єднують ці дані з таблицями вимірів *BreedDimension* і *GenderDimension*, щоб надати описові назви порід і статей. Фрагмент коду можна побачити на рисунку 3.2.

```
SELECT
    af.BreedId,
    af.GenderId,
    af.IsAdopted AS Adopted
FROM ArrivalFacts af
JOIN BreedDimension bd ON af.BreedId = bd.BreedId
JOIN GenderDimension gd ON af.GenderId = gd.GenderId
```

Рис. 3.2 Запит для отримання необхідних даних

Зібрані дані діляться на тренувальну вибірку 70% і тестову вибірку 30%. У разі значного дисбалансу між позитивними (усиновленими) та негативними (не усиновленими) прикладами виконується балансування позитивних зразків до досягнення приблизної рівності між класами. Це запобігає переважанню одного з класів.

Для побудови моделі використовується пакет *ml-naivebayes*, який реалізує алгоритм Наївного Байеса. Метод обчислює умовні ймовірності для кожної категорії (порода та стать) у класах "усиновлено" та "не усиновлено". Модель тренується на збалансованій тренувальній вибірці. Код для навчання зображено на рис. 3.3.

```
const nb = new MultinomialNB();
nb.train(balancedTrainFeatures, balancedTrainLabels);
```

Рис. 3.3 Фрагмент коду для навчання

В коді зображено *balancedTrainFeatures*, який містить характеристики собак (порода та стать), а *balancedTrainLabels* — мітки класів (усиновлено чи ні). Після навчання модель застосовується для прогнозування ймовірностей

$P(C|X)$, де C — це клас, усиновлено чи ні, а X — характеристики собаки, а саме порода та стать. Де клас із максимальною ймовірністю обирається як прогноз.

Для оцінки точності моделі в системі застосовуються кілька метрик. Точність (*Accuracy*) визначає частку правильних прогнозів серед усіх результатів. Точність позитивних прогнозів (*Precision*) відображає частку правильних передбачень для класу. Повнота (*Recall*) показує частку фактично усиновлених собак, які модель правильно ідентифікувала. *F1-score* є гармонійним середнім між точністю та повнотою, забезпечуючи збалансовану оцінку. Окрім цього, формується матриця невідповідностей (*Confusion Matrix*), яка деталізує кількість правильних і помилкових прогнозів для кожного класу.

Результати аналізу зображуються у вигляді графіків, які відображають ймовірності усиновлення або залишення собак залежно від їх породи та статі. На інтерфейсі також представлено ключові метрики моделі, включаючи точність, повноту, *F1-score* та інші показники, що дозволяє оцінити ефективність алгоритму. Для побудови графіків використовується бібліотека *Chart.js*, яка забезпечує наочність і зручність інтерпретації результатів. Реалізацію методу Наївного Байєса можна побачити в ДОДАТКУ Б.

3.4.3. Реалізація методу асоціативних правил. Асоціативні правила дозволяють знаходити приховані взаємозв'язки між характеристиками собак і їхнім статусом. У системі реалізовано алгоритм *FP-Growth* (*Frequent Pattern Growth*). Він вважається одним із найефективніших методів для пошуку частих шаблонів у великих наборах даних. *FP-Growth* був обраний через його здатність працювати з великими обсягами даних, а також через ефективне скорочення простору пошуку за допомогою *FP*-дерева. Для реалізації *FP-Growth* в данній системі використовується пакет *node-fpgrowth*.

FP-Growth має значні переваги над традиційними методами, такими як алгоритм *Apriori*. Алгоритм *Apriori* потребує багаторазового сканування сховища даних для генерації можливих шаблонів [21]. *FP-Growth*, у свою чергу, працює шляхом створення компактної деревовидної структури, яка дозволяє групувати транзакції за спільними елементами.

Алгоритм впорядковує елементи у транзакціях за частотою їхньої появи. Це дозволяє оптимізувати процес аналізу, виділяючи найбільш значущі шаблони. Алгоритм групує шаблони за спільними послідовностями, що суттєво скорочує обсяг даних. Такий підхід дозволяє алгоритму ефективно працювати, навіть, з великими наборами даних, мінімізуючи дублювання і забезпечуючи високу швидкість аналізу.

Транзакції можуть включати елементи, такі як "лабрадор", "самець", "усиновлено", які обробляються як єдина послідовність. Наприклад, у наборі транзакцій спільні елементи групуються. Це дозволяє швидше знайти часті шаблони. Приклад оброблених елементів транзакцій представлений на рисунку 3.4.

```
[ 'Метис', 'самець', 'Не усиновлено' ],
[ 'Метис', 'самець', 'Не усиновлено' ],
[ 'Лабрадор Ретривер', 'самець', 'Усиновлено' ],
[ 'Метис', 'самець', 'Не усиновлено' ],
[ 'Метис', 'самець', 'Усиновлено' ],
[ 'Метис', 'самець', 'Не усиновлено' ],
[ 'Метис', 'самець', 'Не усиновлено' ],
[ 'Метис', 'самець', 'Не усиновлено' ],
[ 'Метис', 'самка', 'Не усиновлено' ],
[ 'Вівчарка Німецька', 'самець', 'Не усиновлено' ],
[ 'Метис', 'самка', 'Усиновлено' ],
[ 'Метис', 'самка', 'Усиновлено' ],
[ 'Метис', 'самка', 'Не усиновлено' ],
[ 'Метис', 'самець', 'Не усиновлено' ],
[ 'Метис', 'самка', 'Не усиновлено' ],
[ 'Метис', 'самець', 'Не усиновлено' ],
[ 'Метис', 'самка', 'Не усиновлено' ],
[ 'Метис', 'самець', 'Не усиновлено' ],
```

Рис. 3.4 Елементи транзакцій для *FP-Growth*

У системі дані для алгоритму отримуються з таблиці *ArrivalFacts*, котрі містить основну інформацію про собак, їхні характеристики і статус усиновлення. Ця таблиця об'єднується з таблицями *BreedDimension* і *GenderDimension*, та надають детальні описи порід і статей собак. *SQL*-запит для отримання потрібних даних можна побачити на рис. 3.5.

```

SELECT
    bd.BreedName,
    gd.GenderDescription AS GenderName,
    CASE
        WHEN af.IsAdopted = 1 THEN N'Усиновлено'
        ELSE N'Не усиновлено'
    END AS AdoptionStatus
FROM ArrivalFacts af
JOIN BreedDimension bd ON af.BreedId = bd.BreedId
JOIN GenderDimension gd ON af.GenderId = gd.GenderId;

```

Рис. 3.5 *SQL*-запит для отримання потрібних даних

Результат цього запиту являє собою набір транзакцій, кожна з яких включає породу, стать і статус усиновлення собаки. Транзакція може виглядати наступним чином [Вівчарка Німецька, самець, не усиновлено]. Ці транзакції передаються до алгоритму *FP-Growth*, який шукає часті набори елементів, а потім формує на їх основі асоціативні правила.

Для аналізу знайдених правил важливо розуміти, як саме оцінюється їхня значущість і корисність. Мінімальна підтримка *support* визначає наскільки часто певний набір елементів має з'являтися у транзакціях, щоб вважатися значущим. В даній системі мінімальна підтримка встановлена на рівні 0.001, що дозволяє знаходити навіть рідкісні шаблони. Фрагмент коду реалізації методу можна побачити в ДОДАТКУ Б. Впевненість (*confidence*) показує, наскільки часто наслідок правила трапляється, коли виконуються його передумови. Наприклад, якщо правило "лабрадор - усиновлено" має впевненість 0.8, це означає, що в 80% випадків лабрадори, які зустрічаються в транзакціях, були усиновлені. Підйом (*lift*) дозволяє оцінити, наскільки знайдене правило є статистично значущим, порівнюючи впевненість правила з ймовірністю наслідку. Значення підйому більше 1 свідчить про позитивну кореляцію між передумовою і наслідком, тоді як значення менше 1 вказує на слабку або відсутню кореляцію. Важливість (*importance*), обчислюється як добуток впевненості та підйому. Це дозволяє розподіляти правила за їхньою практичною корисністю. Приміром є правило "лабрадор, самець - усиновлено"

може мати високу важливість, якщо ця комбінація є поширеною серед усиновлених собак.

Результати обчислень передаються на клієнтську частину у вигляді таблиці, яка включає підмножини, наслідки та обчислені метрики. Як приклад одне з правил можна побачити на рисунку 3.6.

```
{
  subset: [ 'Бультер`єр', 'Усиновлено' ],
  consequent: [ 'самець' ],
  probability: 0.0012106537530266344,
  confidence: 1,
  lift: 1.8183819482663732,
  importance: 1.8183819482663732
},
```

Рис. 3.6. Приклад правила для клієнтської частини.

3.4.4 Реалізація методу Кластеризації. Для методу кластеризації використовується алгоритм K-means для вивчення впливу віку собак на тривалість перебування в притулку. Кластеризація допомагає виявити закономірності та схожі групи собак.

Робота *K-Means* починається з вибору кількості кластерів K , значення K задається користувачем. Алгоритм випадковим чином обирає початкові центри кластерів, після чого кожен елемент даних призначається найближчому центру кластера. Відстань між точками і центрами обчислюється за допомогою евклідової метрики [22].

В цій системі він обраний через простоту реалізації та можливість обробки двовимірних даних — віку та тривалості перебування собак. Завдяки цьому алгоритм дозволяє виділяти групи собак із подібними характеристиками, наприклад, молоді собаки з коротким перебуванням або старші собаки з тривалим перебуванням.

Перед виконанням кластеризації всі дані нормалізуються, щоб зменшити вплив різниці в масштабах між віком (в роках) і тривалістю перебування (у днях). Для нормалізації використовується стандартне відхилення і середнє значення, що дозволяє привести всі показники до одного масштабу, тому що

алгоритм залежить від величин відстаней між точками. Приклад, нормалізації за допомогою пакету *ml-kmeans* приводиться на рисунку 3.7.

```
const meanAge = math.mean(ages);
const stdAge = math.std(ages);

const meanStay = math.mean(stays);
const stdStay = math.std(stays);

const vectors = data.map(item => [
  (item.DogAge - meanAge) / stdAge,
  (item.StayDuration - meanStay) / stdStay
]);
```

Рис. 3.7 Нормалізація даних

Цей фрагмент демонструє, як дані нормалізуються, щоб забезпечити однакову шкалу між віком і тривалістю перебування.

Після початкового призначення алгоритм обчислює нові центри кластерів як середнє значення всіх точок у кожному кластері. Цей процес повторюється до досягнення стабільності, тобто коли центри перестають змінюватися. Для виконання кластеризації з обраною кількістю кластерів використовується наступний код представлений на рисунку 3.8.

```
const kmeansResult = kmeans(vectors, k);

const clusteredData = data.map((item, index) => ({
  ...item,
  cluster: kmeansResult.clusters[index] + 1
}));
```

Рис 3.8 Виконання кластеризації з обраною кількістю кластерів

Тут значення *clusteredData* містить інформацію про кожну собаку, включаючи кластер, до якого вона належить.

Для визначення оптимальної кількості кластерів використовується метод "Ліктя". У ході цього обчислюється *WCSS* (*Within-Cluster Sum of Squares*) — сума квадратів відстаней від точок до центрів кластерів. Графік *WCSS* показує,

як змінюється якість кластеризації залежно від кількості кластерів K . Оптимальне значення K відповідає точці "згину" графіка, де подальше збільшення K перестає суттєво зменшувати $WCSS$. Приклад циклу для виконання $WCSS$ з певною K кількості кластерів зображено на рисунку 3.9.

```
const wcss = [];
const maxK = 10;
for (let k = 1; k <= maxK; k++) {
  const kmeansResult = kmeans(vectors, k);
  let sum = 0;
  for (let i = 0; i < vectors.length; i++) {
    const centroid = kmeansResult.centroids[kmeansResult.clusters[i]];
    const distance = math.distance(vectors[i], centroid);
    sum += Math.pow(distance, 2);
  }
  wcss.push(sum);
}
```

Рис. 3.9 Цикл виконання $WCSS$ з певною K

Результати кластеризації у системі представлені у вигляді інтерактивного графіка. На ньому точки представляють собак, розміщені у двовимірному просторі, де вісь X відповідає віку, а вісь Y — тривалості перебування. Кожна точка має свій колір, що відповідає кластеру.

3.4.5 Реалізація часових рядів. В системі для перевірки двох гіпотез використовувалися часові ряди та метод *ARIMA*. Метою є визначити, чи існують певні часові періоди, коли до притулку потрапляє більше собак у певному регіоні, та виявити сезонні коливання у кількості усиновлень в обраному регіоні.

Для першої гіпотези було сконцентровано акцент на аналізі даних про надходження собак до притулку. У файлі *seasonality.js* був створений маршрут */arima-analysis*, який отримує дані про кількість надходжень. Це було зібрано за сезонами та роками, з сховища даних за допомогою *SQL*-запиту. Даний *SQL*-запит продемонстровано на рис. 3.10.

```

SELECT
  atd.Season AS Season,
  atd.Year AS Year,
  SUM(af.ArrivalCount) AS ArrivalCount
FROM ArrivalFacts af
JOIN ArrivalTimeDimension atd ON af.ArrivalTimeId = atd.ArrivalTimeId
GROUP BY atd.Year, atd.Season
ORDER BY atd.Year, atd.Season;

```

Рис. 3.10 Запит для кількості надходжень

Результати цього запиту обробляються для створення масиву *timeSeriesData*, який містить послідовність кількостей надходжень за кожен сезон.

Після підготовки даних застосовано метод *ARIMA* для аналізу та прогнозування. Реалізація цього коду виглядає наступним чином на рисунку 3.11.

```

// Створюємо модель ARIMA
const arima = new ARIMA({ p: 4, d: 3, q: 4, seasonal: true, period: 4, verbose: false }).train(timeSeriesData);

// Прогнозуємо на наступні 8 сезонів
const [forecast] = arima.predict(8);

```

Рис 3.11 Застосування методу *ARIMA* для аналізу та прогнозування надходжень

Параметри для моделі встановлені наступним чином:

- $p = 4$ – число попередніх значень;
- $d = 3$ – обчислення різниць для стабілізації;
- $q = 4$ – порядок замінного середнього.

Параметр *seasonal* встановлено в *true*, а *period* дорівнює 4, що відповідає сезонності в році.

Модель *ARIMA* тренується на історичних даних, після чого виконується прогноз на наступні 8 сезонів (два роки). Результат прогнозу зберігається у змінній *forecast*.

Для другої гіпотези, що стосується усиновлень, процес реалізовано аналогічно. Маршрут */arima-analysis* у цьому файлі отримує дані про кількість усиновлень, агрегованих за сезонами та роками. Приклад коду на рисунку 3.12.

```

SELECT atd.Season AS Season,
       atd.Year AS Year,
       SUM(af.AdoptionCount) AS AdoptionCount
FROM AdoptionFacts af
JOIN AdoptionTimeDimension atd ON af.AdoptionTimeId = atd.AdoptionTimeId
GROUP BY atd.Year, atd.Season
ORDER BY atd.Year, atd.Season;

```

Рис. 3.12 Запит для кількості усиновлень

Після отримання даних створюється масив *timeSeriesData*, який містить кількість усиновлень за кожен сезон. Далі застосовується метод *ARIMA* з іншими параметрами, оскільки дані про усиновлення мають іншу структуру та характеристики. Це зображено на рисунку 3.13.

```

// Створюємо модель ARIMA
const arima = new ARIMA({ p: 2, d: 1, q: 2, seasonal: true, period: 4, verbose: false }).train(timeSeriesData);

// Прогнозуємо на наступні 8 сезонів
const forecast = arima.predict(8)[0];

```

Рис. 3.13 Застосування методу *ARIMA* для аналізу та прогнозування усиновлень

Тут параметри моделі встановлені наступним чином:

- $p = 2$;
- $d = 1$;
- $q = 2$.

Параметр *seasonal* залишаються *true*, а *period* – 4.

У обох випадках, після прогнозування, результати разом з історичними даними передаються на клієнтську частину для візуалізації за допомогою бібліотеки *Chart.js*. Це дозволяє побудувати графіки які відображають як історичні дані, так і прогноз на майбутні сезони.

Основна увага приділяється обробці часових міток для графіка. Приклад цього наведено на основі фрагменту коду з прогнозу надходжень на рисунку 3.14.

```

let nextYear = data.lastYear;
let currentSeasonIndex = labels.indexOf(data.lastSeason);

const forecastLabels = [];
for (let i = 0; i < data.forecast.length; i++) {
  currentSeasonIndex = (currentSeasonIndex + 1) % 4;
  if (currentSeasonIndex === 0) nextYear++;
  forecastLabels.push(`Pik ${nextYear} ${labels[currentSeasonIndex]}`);
}

```

Рис. 3.14 Обробка часових міток для графіка

Це правильно відображає прогнозовані дані на графіку з відповідними сезонами та роками.

Додатково, реалізовано перевірку на достатню кількість даних для коректної роботи моделі *ARIMA*. Якщо кількість точок у часовому ряді менша за необхідну, модель не будується, і користувач отримує відповідне повідомлення. Логіку можна побачити на рис. 3.15.

```

if (timeSeriesData.length < 8) {
  return res.status(400).json({ error: 'Недостатньо даних для ARIMA аналізу' });
}

```

Рис. 3.15 Логіка повідомлення

Це важливо, оскільки модель *ARIMA* потребує достатньої кількості даних для навчання та адекватного прогнозування.

3.5 Реалізація ключових показників ефективності

У процесі розробки системи підтримки прийняття рішень для керівництва собачого притулку було реалізовано декілька ключових показників ефективності (*KPI*), які дозволяють оцінювати різні аспекти діяльності притулку. Ці *KPI* допомагають не лише відстежувати поточний стан, але й аналізувати тенденції та ефективність управлінських рішень. Реалізація *KPI* базується на даних зі сховища, що забезпечує доступ до історичної та актуальної інформації для глибокого аналізу.

Кожен *KPI* у системі побудований на основі чотирьох компонентів:

- Поточне значення - відображає реальний показник за обраний період;
- Ціль - визначає бажаний рівень, якого прагне досягти притулок;

- Статус - показує, чи досягнуто цільове значення з урахуванням допустимої толерантності;
- Тенденція - демонструє динаміку зміни показника в порівнянні з попереднім періодом.

Середній час перебування собак віком до 5 років оцінює середню тривалість перебування молодих собак у притулку до моменту усиновлення. Поточне значення розраховується шляхом обчислення середнього значення поля *LengthOfStay* з таблиці *AdoptionFacts* для собак віком до 5 років за поточний рік. Ціль встановлюється як 95% від середнього часу перебування за попередній рік, стимулюючи зменшення цього показника на 5%. Статус визначається порівнянням поточного значення з ціллю: якщо воно менше або дорівнює цільовому, статус "Досягнуто"; якщо більше — "Не досягнуто". Тенденція визначається шляхом порівняння з минулорічним показником: зменшення середнього часу вказує на "Покращення", збільшення — на "Погіршення".

Середній час перебування собак віком від 6 років аналогічно оцінює середню тривалість перебування старших собак. Поточне значення обчислюється для собак віком від 6 років; ціль, статус та тенденція визначаються за тією ж логікою.

Кількість перебування серед собак віком до 5 років більше ніж 125 днів вказує на кількість молодих собак, які перебувають у притулку понад 125 днів. Поточне значення обчислюється шляхом підрахунку записів у таблиці *ArrivalFacts* для собак віком до 5 років, де *LengthOfStay* перевищує 125 днів. Ціль встановлюється як 95% від аналогічного показника за попередній рік. Статус та тенденція визначаються порівнянням з ціллю та попереднім значенням.

Кількість перебування серед собак віком від 6 років більше ніж 125 днів оцінює кількість старших собак, які перебувають у притулку понад 125 днів.

Логіка розрахунку поточного значення, цілі, статусу та тенденції аналогічна попередньому показнику.

Середній час усиновлення за породою оцінює середню тривалість перебування собак конкретної породи до моменту усиновлення. Поточне значення обчислюється шляхом вибірки з таблиці *AdoptionFacts* усіх записів за поточний рік для обраної породи та обчислення середнього *LengthOfStay*. Ціль встановлюється як 95% від середнього часу перебування за попередній рік для цієї породи. Статус та тенденція визначаються аналогічно.

Середній час усиновлення за статтю оцінює середній час перебування собак певної статі. Поточне значення обчислюється для обраної статі; ціль, статус та тенденція визначаються за тією ж логікою.

Кількість собак, що перебувають довго за породою, вказує на кількість собак певної породи, які перебувають у притулку понад 180 днів. Поточне значення обчислюється шляхом підрахунку записів у таблиці *ArrivalFacts* для обраної породи з *LengthOfStay*, що перевищує поріг. Ціль встановлюється як 95% від аналогічного показника за попередній рік. Статус та тенденція визначаються відповідно.

Кількість усиновлених собак за статтю відображає кількість собак обраної статі, які були усиновлені за поточний рік. Поточне значення отримується шляхом підсумовування *AdoptionCount* з таблиці *AdoptionFacts* для обраної статі. Ціль встановлюється як 105% від кількості усиновлень за попередній рік для цієї статі, прагнучи збільшити цей показник на 5%. Статус та тенденція визначаються аналогічно.

Кількість надходжень у сезоні аналізує кількість собак, що надходять до притулку в конкретний сезон. Поточне значення обчислюється шляхом сумування *ArrivalCount* з таблиці *ArrivalFacts*, поєднаної з *ArrivalTimeDimension* для визначення сезону. Ціль встановлюється як 95% від кількості надходжень за відповідний сезон минулого року. Статус та тенденція визначаються за тією ж логікою.

Максимальне надходження з усіх сезонів визначає сезон з найбільшою кількістю надходжень за поточний рік. Поточне значення — максимальне *ArrivalCount* серед усіх сезонів. Ціль встановлюється як 95% від максимального значення минулого року. Статус та тенденція визначаються аналогічно.

Мінімальне надходження з усіх сезонів визначає сезон з найменшою кількістю надходжень. Ціль встановлюється як 95% від мінімального значення минулого року. Статус та тенденція визначаються відповідно.

Зміна сезонної кількості надходжень, тобто середнє значення по всіх сезонах, відображає загальну тенденцію зміни кількості надходжень собак до притулку. Поточне значення обчислюється як середнє *ArrivalCount* по всіх сезонах за поточний рік. Ціль встановлюється як 95% від середнього значення минулого року. Статус та тенденція визначаються за стандартною логікою.

Кількість усиновлень у сезоні аналізує кількість усиновлень у конкретному сезоні. Поточне значення обчислюється шляхом сумування *AdoptionCount* з таблиці *AdoptionFacts*, поєднаної з *AdoptionTimeDimension* для визначення сезону. Ціль встановлюється як 105% від кількості усиновлень за відповідний сезон минулого року. Статус та тенденція визначаються аналогічно.

Максимальне усиновлення за сезон визначає сезон з найбільшою кількістю усиновлень за поточний рік. Поточне значення — максимальне *AdoptionCount* серед усіх сезонів. Ціль встановлюється як 105% від максимального значення минулого року. Статус та тенденція визначаються відповідно.

Мінімальне усиновлення за сезон визначає сезон з найменшою кількістю усиновлень. Ціль встановлюється як 105% від мінімального значення минулого року. Статус та тенденція визначаються аналогічно.

Кількість усиновлень у поточному році відображає загальну кількість усиновлень за поточний рік. Поточне значення отримується шляхом сумування *AdoptionCount* з таблиці *AdoptionFacts* за поточний рік. Ціль встановлюється як

105% від загальної кількості усиновлень за попередній рік. Статус та тенденція визначаються відповідно.

На клієнтській стороні *KPI* відображаються за допомогою веб-інтерфейсу, реалізованого з використанням *HTML*, *CSS* та *JavaScript*. Інтерфейс дає зрозуміле та інтуїтивне представлення кожного *KPI*, включаючи назву показника, поточне значення, цільове значення, статус та тенденцію. Візуальні елементи, такі як кольорові індикатори та символи, допомагають швидко оцінити стан показника. Наприклад, зелений колір та символ «Пташка» вказують на досягнення цілі, червоний колір та символ «X» — на її недосягнення, а жовтий колір та символ крапки — на незначні відхилення в межах допустимої толерантності. Ці візуальні підказки реалізуються через динамічне додавання *CSS*-класів залежно від статусу та тенденції.

Для уникнення частих змін статусу показника через незначні коливання було впроваджено поняття толерантності. Якщо відхилення поточного значення від цільового не перевищує 2%, статус встановлюється як "Без змін". Це дозволяє зосередитися на суттєвих змінах та уникнути зайвих тривог через незначні флуктуації даних. Наприклад, якщо ціль становить 100 усиновлень, то статус залишатиметься "Без змін" при значеннях у межах 98–102.

Реалізація всіх зазначених *KPI* у системі дозволяє керівництву притулку отримувати повну картину про діяльність організації. Використання реальних даних зі сховища та їх інтуїтивне відображення в інтерфейсі сприяють прийняттю обґрунтованих управлінських рішень. Система надає можливість вчасно виявляти проблемні області, оцінювати ефективність впроваджених заходів та планувати подальший розвиток притулку на основі об'єктивних даних.

3.6 Дані для аналізу

Дані для аналізу були зібрані з відкритих джерел, доступ до яких забезпечено через державні портали, зокрема портал «ДІЯ». Однак процес збору даних був ускладнений кількома факторами. Перш за все, не всі притулки для тварин публікують відповідну інформацію, що суттєво звужує коло

доступних джерел. Але навіть ті заклади, які оприлюднюють звіти, часто не дотримуються чіткої та стандартизованої структури даних, і це створює труднощі у їхньому використанні для систематичного аналізу. Попри ці виклики, вдалося зібрати та впорядкувати кілька наборів даних, які стали основою для дослідження.

Зокрема, було використано звіти, отримані з притулків для тварин у різних регіонах України. Ці звіти містять дані про вилов, стерилізацію, вакцинацію та ідентифікацію безпритульних тварин, а також відомості про їхню кількість та стан. Найбільш повні й доступні дані було отримано від таких установ:

- Прилуки (Чернігівська область): Реєстр обліку собак районного центру Прилуцької міської ради.
- Покров (Дніпропетровська область): Звіт по безпритульним тваринам, підготовлений виконавчим комітетом міської ради.
- Кам'янське (Дніпропетровська область): Дані про вилов, стерилізацію та ідентифікацію безпритульних тварин, що збираються комунальною установою притулку для тварин.
- Івано-Франківськ (Івано-Франківська область): Дані про вилов, стерилізацію та ідентифікацію безпритульних тварин, зібрані Івано-Франківською міською територіальною громадою.
- Тернопіль (Тернопільська область): Звіт Тернопільської міської ради, що включає дані про вилов та ідентифікацію тварин.
- Вараш (Рівненська область): Дані про вилов, стерилізацію, вакцинацію та ідентифікацію тварин у місті Вараш.
- Мукачево (Закарпатська область): Звіт виконавчого комітету Мукачівської міської ради щодо діяльності місцевого притулку.
- Коростень (Житомирська область): Дані про вилов та ідентифікацію безпритульних тварин, що були отримані від Коростенської міської ради.

- Ватутіне (Черкаська область): Інформація про вилов, стерилізацію та ідентифікацію тварин, надана Ватутінською міською територіальною громадою.
- Львів (Львівська область): Набір даних Львівської міської територіальної громади, що містить детальну інформацію про безпритульних тварин у регіоні.

Саме ці дані дозволили охопити географічно різні регіони України, що надало можливість врахувати регіональні особливості в аналізі. Наприклад, відмінності у кількості безпритульних тварин, їхній обробці та рівні усиновлення могли залежати від місцевих умов. Зібрані набори даних мають ряд специфічних характеристик. Вони містять інформацію про дати вилову та надходження тварин до притулків, їхню стерилізацію, вакцинацію, колір, вік, вагу, а також статус усиновлення. Такий склад даних надає змогу формувати багатовимірні аналітичні звіти та проводити глибокий аналіз із застосуванням методів *Data Mining*, що є ключовим для реалізації завдань даної магістерської роботи. Приклад звіту можна побачити на рис. 3.16.

Numb	kind	Color	residenceAddress	catchDate	dischargeDate	chipNumber
1	собака	рижа	Василя Спмоненка, 1Б	2023-05-03	2023-05-30	к. 00125 ч.900215003891413
2	собака	чорна	Фірма "Авакс Проф", вул. Деніса Лукіновича, 8	2023-05-05	2023-05-30	к. 00126 ч.900215003891415
3	собака	сіра	психоневрологічна лікарня, вул. Тролейбусна, 14	2023-05-11	2023-05-30	к.00127 ч.900215003891417
4	собака	рижа	вул. Дністрянського, 1	2023-05-11	2023-05-30	к.00128 ч.900215003891420
5	собака	чорно-рижа	вул. Лозовська, 26	2023-05-12	2023-05-30	к.00129 ч.900215005298399
6	собака	біло-рижа	вул. Карпенка, 11	2023-05-12	2023-05-30	к.00130 ч.900215005298383
7	собака	чорна	вул. Білогірська, 15	2023-05-12	2023-05-30	к.00131 ч.900215005298394
8	собака	сіра	с. Глядки	2023-05-16	2023-05-30	к.00133 ч.900215005298395
9	собака	чорна	с. Куршики	2023-05-16	2023-05-30	к.00136 ч.900215005298322
10	собака	чорний	вул. Транспортна, 4	2023-05-19	2023-05-30	к.00137 ч.900215005298311
11	собака	чорна	вул. Київська, 10	2023-05-19	2023-05-30	к.00138 ч.900215005298305
12	собака	чорно-рижа	вул. Будного, 13	2023-05-23	2023-05-30	к.00139 ч.900215005298306
13	собака	чорна	вул. Будного, 13	2023-05-23	2023-05-30	к.00140 ч.900215005298301
14	собака	біло-рижа	ТОВ "Кленовий гай", вул	2023-05-25	2023-05-30	к.00221 ч.900215005298304

Рис. 3.16 Приклад звіту з міста Тернопіль

Слід зазначити, що наявні звіти часто мають проблеми з якістю: неповні або неточні дані, відсутність уніфікованих форматів подання, а також розбіжності у структурі між різними джерелами. Тому попередня підготовка цих даних, їхня нормалізація та інтеграція в єдину базу були важливим етапом перед початком аналізу. Для забезпечення коректності та надійності результатів було проведено перевірку даних на помилки, видалено дублікатні записи та узгоджено різні формати подання.

4 РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

4.1 Дослідження використання задач класифікації

У системі підтримки прийняття рішень для управління собачим притулком використовуються два основних методи класифікації: *One-Rule* та Наївний Байєс.

4.1.1 Використання *One-Rule* для класифікації. Алгоритм *One-Rule* застосовується для оцінки впливу окремих атрибутів на результуючий показник. Метод базується на аналізі одного атрибуту, який найкраще прогнозує цільову змінну. Його простота та зрозумілість роблять цей алгоритм ефективним інструментом для попереднього аналізу великих даних.

У контексті даного дослідження, *One-Rule* було використано для класифікації тривалості перебування собак у притулку на основі таких атрибутів:

- Вік собаки
- Порода
- Стать
- Локація (місто притулку)

Метою дослідження є визначення впливу кожного з атрибутів — вік, порода, стать та локація — на тривалість перебування собак у притулку. Завдання дослідження полягає в ідентифікації найбільш вагомих для класифікації атрибутів, перевірці гіпотези щодо їх суттєвого впливу, а також розробці рекомендацій для оптимізації процесу усиновлення собак.

Результати класифікації за атрибутом вік. Аналіз даних за віком собак показав, що молодші собаки (до 5 років) частіше класифікуються як такі, що мають коротке перебування, тоді як старші (від 6 років) мають більшу ймовірність тривалого перебування. На рисунку 4.1 можна побачити графічну реалізацію для атрибута вік.

Вік	Кількість собак	Тривале перебування	Коротке перебування	Класифікація	Помилки
0	73	25	48	Коротке перебування	34.25%
1	74	29	45	Коротке перебування	39.19%
2	136	51	85	Коротке перебування	37.50%
3	136	57	79	Коротке перебування	41.91%
4	123	54	69	Коротке перебування	43.90%
5	80	27	53	Коротке перебування	33.75%
6	86	45	41	Тривале перебування	47.67%
7	85	29	56	Коротке перебування	34.12%
8	71	31	40	Коротке перебування	43.66%
9	69	26	43	Коротке перебування	37.68%
10	44	24	20	Тривале перебування	45.45%
11	52	35	17	Тривале перебування	32.69%
12	59	30	29	Тривале перебування	49.15%

Рис. 4.1 Графічна реалізація для атрибуту вік

Результати класифікації за атрибутом порода. Дані свідчать про те, що породи з більшою популярністю серед населення Йоркширський тер'єр, Лабрадор частіше класифікуються як короткотермінові. Натомість рідкісні породи або метиси мають вищий ризик тривалого перебування. На рисунку 4.2 зображена графічна реалізація для порода.

Оберіть атрибут для аналізу:
Порода

Показати записів

Значення атрибуту	Загальна кількість	Тривале перебування	Коротке перебування	Класифікація	Відсоток помилок (%)
Боксер	5	4	1	Тривале	20.00%
Бультер'єр	4	2	2	Коротке	50.00%
Бігль	4	1	3	Коротке	25.00%
Вельш Коргі Пемброк	4	1	3	Коротке	25.00%
Вівчарка Німецька	32	13	19	Коротке	40.63%
Голден Ретривер	28	11	17	Коротке	39.29%
Далматинець	1	0	1	Коротке	0.00%
Доберман	1	0	1	Коротке	0.00%
Йоркширський Тер'єр	5	0	5	Коротке	0.00%
Кане Корсо	1	0	1	Коротке	0.00%
Кокер-спанієль	3	2	1	Тривале	33.33%
Лабрадор Ретривер	35	17	18	Коротке	48.57%
Лайка	2	0	2	Коротке	0.00%
Мальтійська Болонка	1	0	1	Коротке	0.00%
Метис	933	396	537	Коротке	42.44%
Мопс	9	7	2	Тривале	22.22%

Рис. 4.2 Графічна реалізація за атрибутом порода

Результати аналізу за статтю показали, що розподіл між самцями та самками є досить рівномірним, однак самки мають трохи вищий відсоток

помилки у класифікації. На рисунку 4.3 можна побачити реалізацію для атрибуту стать.

Оберіть атрибут для аналізу:
Стать

Показати записів

Значення атрибуту	Загальна кількість	Тривале перебування	Коротке перебування	Класифікація	Відсоток помилок (%)
самець	565	234	331	Коротке	41.42%
самка	523	229	294	Коротке	43.79%

Рис. 4.3 Реалізація для атрибуту стать

Результати класифікації за атрибутом локація. Локація притулку суттєво впливає на тривалість перебування. Наприклад, у великих містах як Львів, Івано-Франківськ більше собак класифікується як такі, що мають короткий термін перебування, ніж у менших населених пунктах. На рисунку 4.4. демонструється графічна реалізація за атрибутом локація.

Оберіть атрибут для аналізу:
Локація

Показати записів

Значення атрибуту	Загальна кількість	Тривале перебування	Коротке перебування	Класифікація	Відсоток помилок (%)
м. Вараш, UA	56	22	34	Коротке	39.29%
м. Ватутіне, UA	45	21	24	Коротке	46.67%
м. Кам'янське, UA	157	62	95	Коротке	39.49%
м. Коростень, UA	80	40	40	Коротке	50.00%
м. Львів, UA	185	72	113	Коротке	38.92%
м. Мукачево, UA	118	61	57	Тривале	48.31%
м. Покров, UA	84	31	53	Коротке	36.90%
м. Прилуки, UA	73	34	39	Коротке	46.58%
м. Тернопіль, UA	134	60	74	Коротке	44.78%
м. Івано-Франківськ, UA	156	60	96	Коротке	38.46%

Рис. 4.4. Реалізація за атрибутом локація

Для перевірки гіпотези про те, що вік собак суттєво впливає на тривалість їхнього перебування у притулку, було використано аналіз ключових показників ефективності (*KPI*) за допомогою *OLAP*-технологій. Цей підхід дозволяє детально розглянути різні аспекти даних та отримати багатовимірний аналіз взаємозв'язків між показниками. Графічне представлення показників ефективності можна побачити на рис. 4.5.

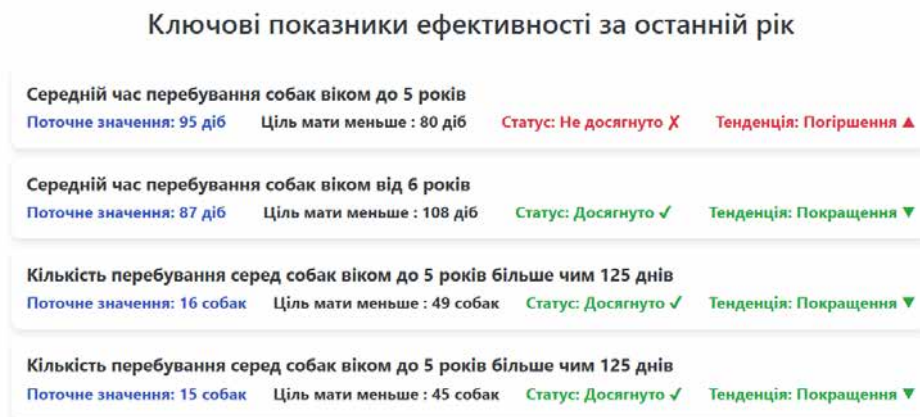


Рис. 4.5 Показник ефективності для перевірки *One-Rule*

Аналіз показує, що середній час перебування молодих собак у притулку перевищує цільове значення на 15 діб. Більше того, спостерігається тенденція до погіршення, що свідчить про збільшення тривалості перебування молодих собак порівняно з попереднім періодом. Це може вказувати на те, що молоді собаки не знаходять нових власників так швидко, як очікувалося.

Кількість молодих собак, які перебувають у притулку понад 125 днів, значно менша за цільове значення. Тенденція до покращення вказує на те, що все менше молодих собак затримуються у притулку на довгий термін, що підтверджує ефективність заходів щодо їхнього швидкого усиновлення.

Ситуація зі старшими собаками також позитивна: кількість тварин, що перебувають понад 125 днів, менша за цільове значення. Тенденція до покращення свідчить про успішне зменшення кількості довготривалих перебувань серед старших собак.

OLAP-аналіз *KPI* демонструє, що, незважаючи на збільшення середнього часу перебування молодих собак, кількість молодих собак, що перебувають у притулку понад 125 днів, суттєво зменшилася. Це може вказувати на те, що хоча загальний час перебування молодих собак збільшився, довготривале перебування серед них стало рідкістю.

Для старших собак спостерігається як зменшення середнього часу перебування, так і зниження кількості довготривалих перебувань. Це свідчить

про успішність заходів, спрямованих на прискорення усиновлення старших тварин.

Результати аналізу частково підтверджують гіпотезу про те, що вік впливає на тривалість перебування собак у притулку. Старші собаки, завдяки цільовим зусиллям, стали перебувати у притулку менше часу, що суперечить загальноприйнятій думці про складність усиновлення старших тварин.

Збільшення середнього часу перебування молодих собак може свідчити про зміну поведінки потенційних власників або про інші фактори, що впливають на процес усиновлення. Це вказує на необхідність подальшого дослідження та, можливо, коригування стратегій щодо молодих собак.

Гіпотеза про вплив атрибутів на тривалість перебування собак у притулку підтверджена. Вік є найбільш значущим фактором, тоді як стать має найменший вплив. Локація також відіграє важливу роль у процесі усиновлення. Результати дослідження можуть бути використані для розробки рекомендацій щодо оптимізації процесу усиновлення.

4.1.2 Використання метода Наївного Байєса для класифікації. Метод Наївного Байєса базується на застосуванні теореми Байєса із припущенням про незалежність характеристик (факторів). Реалізацію Наївного Байєса використано для класифікації даних щодо ймовірності усиновлення собак у притулках залежно від таких характеристик, як порода та стать.

Для цього було взято дані про собак із таблиці фактів *ArrivalFacts*. Дані були використані для тренування моделі Наївного Байєса, а також для перевірки її точності на тестовій вибірці. Характеристики породи *BreedId* та статі *GenderId* були обрані як вхідні змінні (фактори), а залежна змінна — статус усиновлення *Adopted*.

Для забезпечення коректного тренування моделі дані були розподілені на дві вибірки: навчальну - 70% для тренування моделі та тестову - 30% для її перевірки.

Аналіз показав, що клас "усиновлено" представлений значно меншою кількістю зразків порівняно з класом "не усиновлено". Щоб вирішити проблему

дисбалансу даних, було застосовано метод балансування вибірки менш представленого класу. Це дозволило більш ефективно моделі розпізнавати особливості обох класів.

Для тренування використовувався алгоритм *MultinomialNB* із пакету *ml-naivebayes*. Характеристики породи та статі собаки були використані як вхідні змінні, на основі яких модель навчалася передбачати статус усиновлення.

На тестовій вибірці було виконано оцінку моделі за основними метриками: точність, точність позитивних прогнозів, повнота та *F1*-міра. Крім того, була побудована матриця невідповідностей, яка наочно демонструє кількість правильних і помилкових прогнозів для кожного класу. Це дозволило краще зрозуміти ефективність моделі та її слабкі місця. Графічне зображення результатів зображено на рис. 4.6.

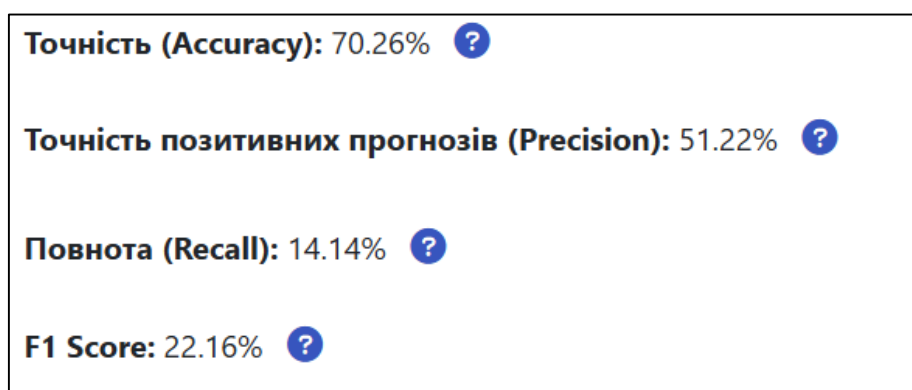


Рис. 4.6 Графічне представлення результатів тестування даних

Результати виконання алгоритму на тестових даних показали наступне:

- Точність (ассурасу): 70.26%, демонструє загальну частку правильних прогнозів.
- Точність позитивних прогнозів (precision): 51.22%, вказує на точність передбачень усиновлення.
- Повнота (recall): 14.14%, вказує на здатність моделі виявляти фактичні випадки усиновлення.
- *F1*-міра: 22.16%, є гармонійним середнім між точністю та повнотою.

Низьке значення повноти свідчить про складність моделі у виявленні випадків "усиновлено", що може бути зумовлено дисбалансом класів або недостатньою кількістю значущих факторів. Проте висока точність вказує на те, що модель працює добре для домінуючого класу.

Матриця невідповідностей показала, що модель добре передбачає випадки "не усиновлено" ($TN = 655$), але менш ефективна у прогнозуванні "усиновлено" ($TP = 42$). Матрицю невідповідностей зображено на рис. 4.7.

Матриця невідповідностей		
	Прогнозовано: Не усиновлено	Прогнозовано: Усиновлено
Фактично: Не усиновлено	655	40
Фактично: Усиновлено	255	42

Правильно передбачено не усиновлено (TN): 655 випадків, коли модель **правильно** передбачила, що собака **не буде усиновлена**.

Неправильно передбачено усиновлено (FP): 40 випадків, коли модель **помилково** передбачила, що собака **буде усиновлена**, хоча насправді її **не усиновили**.

Неправильно передбачено не усиновлено (FN): 255 випадків, коли модель **помилково** передбачила, що собака **не буде усиновлена**, хоча насправді її **усиновили**.

Правильно передбачено усиновлено (TP): 42 випадків, коли модель **правильно** передбачила, що собака **буде усиновлена**.

Рис. 4.7 Матриця невідповідностей

Для полегшення інтерпретації результатів були побудовані графіки ймовірностей усиновлення залежно від породи та статі. Графіки демонструють, які породи та статі собак мають вищі шанси на усиновлення. На рисунку 4.8 зображено графік ймовірності усиновлення за породою.

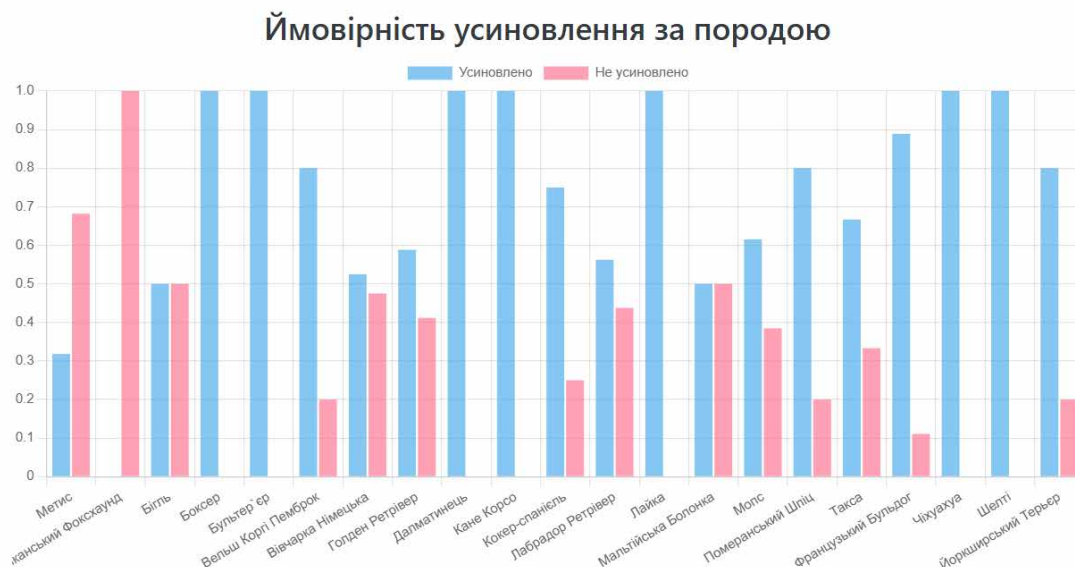


Рис. 4.8 Графік ймовірності усиновлення за породою

На рисунку 4.9 зображено графік ймовірності усиновлення за статтю.

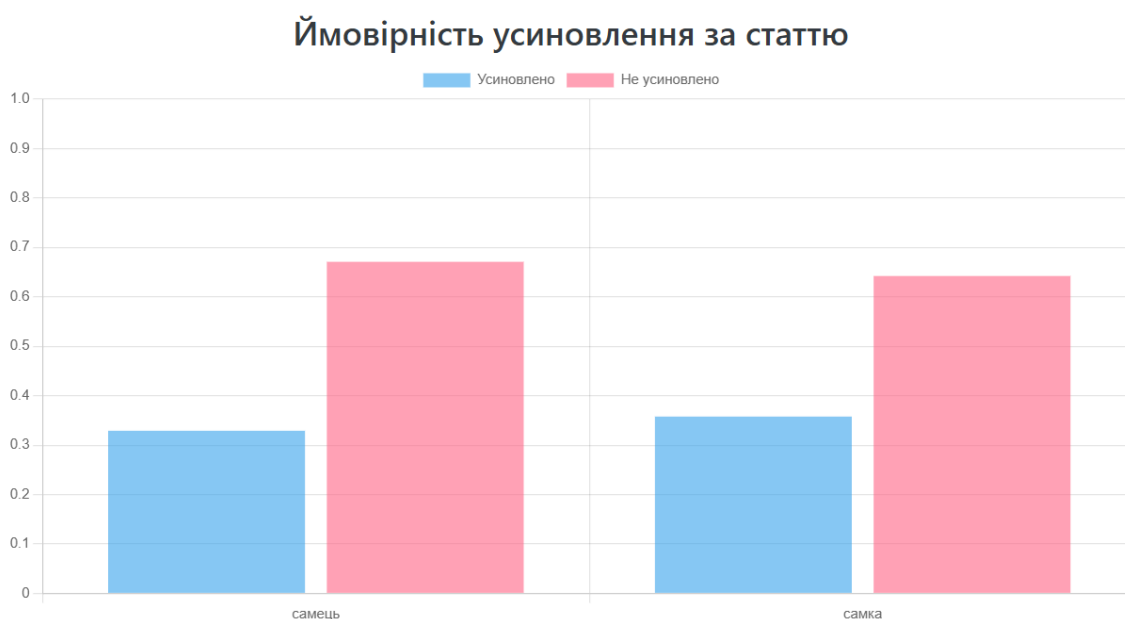


Рис. 4.9 Графік ймовірності усиновлення за породою

На графіку ймовірності усиновлення за породою видно, що собаки певних порід мають значно вищу ймовірність бути усиновленими. А саме Боксер, Йоркширський, Далматинець Тер'єр, Чіхуахуа і Кокер-спанієль демонструють майже 100% успішного усиновлення. Водночас породи, такі як Американський фоксхаунд, метис, мають нижчі шанси на усиновлення, що може свідчити про триваліше перебування цих собак у притулку.

Графік ймовірності усиновлення за статтю свідчить про майже однакові показники для самців і самок, хоча самки мають незначну перевагу. Це вказує на те, що стать собаки не є вирішальним фактором для усиновлення. Водночас значна частина собак, незалежно від статі, залишаються не усиновленими, що потребує додаткових заходів для покращення цієї ситуації.

Гіпотеза про вплив породи та статі собаки на ймовірність усиновлення була перевірена за допомогою ключових показників ефективності. Аналіз середнього часу усиновлення за породою показав, що найкращий середній час демонструє Йоркширський Тер'єр (10 днів), а найгірший — метис (94 днів). Популярні породи, такі як Йоркширський Тер'єр, Лайка, мають значно нижчий середній час усиновлення, що підтверджує гіпотезу про вплив породи.

Аналіз кількості собак, які довго перебувають у притулку (більше 180 днів), показав, що Метиси мають найвищу кількість тривалих перебувань (243 собаки). Це свідчить про їхню низьку популярність або недостатній інтерес до таких собак. Водночас Вівчарки Німецькі та Лабрадори, попри високу ймовірність усиновлення, також мають кілька випадків довготривалого перебування, що може бути пов'язано з їхніми специфічними вимогами щодо утримання.

Щодо кількості усиновлених собак за статтю, дані *KPI* свідчать, що за останній рік більше самців було усиновлено, ніж самок. Обидва показники перевищили встановлені цілі, проте різниця між ними є незначною. На рисунку 4.10 зображено реалізацію показників ефективності для перевірки методу Наївного Байєса.

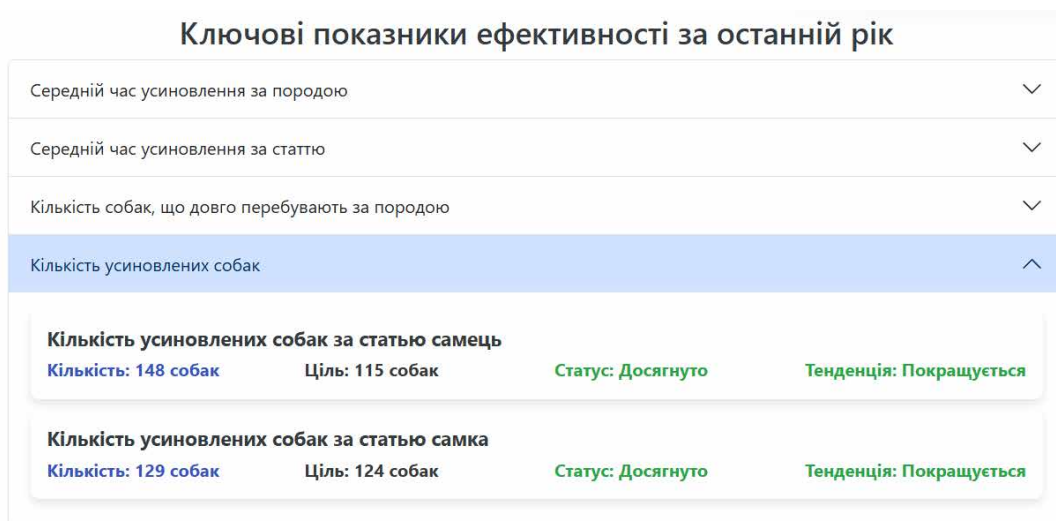


Рис. 4.10 Показники ефективності для перевірки

Графіки та КРІ підтверджують, що порода суттєво впливає на ймовірність усиновлення та середній час перебування у притулку. Популярні породи швидше знаходять нових власників, тоді як менш популярні, наприклад, Метиси, залишаються в притулку довше. Стать, своєю чергою, не має суттєвого впливу на ймовірність усиновлення. Незначні відмінності між самцями та самками можуть бути помітними, але вони не є визначальними.

4.2 Дослідження використання методу асоціативних правил

Аналіз правил дозволив виявити кілька цікавих тенденцій. Породи собак, такі як французький бульдог, померанський шпіц та йоркширський тер'єр, мають найвищу ймовірність усиновлення, особливо серед самок. З іншого боку, менш популярними для усиновлення виявилися метиси, лабрадори та голден ретрівери, причому ймовірність усиновлення знижується для самців цих порід. Гендерні особливості відіграють важливу роль, оскільки в деяких породах, як-от французький бульдог, самки демонструють вищу ймовірність усиновлення порівняно із самцями. На рисунку 4.11 можна побачити графічне представлення асоціативних правил.

Підмножина	Наслідок	Вірогідність	Впевненість	Підйом	Важливість
Померанський Шпіц, самка	Усиновлено	0.0012	1.0000	3.0368	3.0368
Французький Бульдог, самка	Усиновлено	0.0018	1.0000	3.0368	3.0368
Померанський Шпіц, Усиновлено	самка	0.0012	1.0000	2.2219	2.2219
Бультер`єр, Усиновлено	самець	0.0012	1.0000	1.8184	1.8184
Бультер`єр	самець	0.0015	1.0000	1.8184	1.8184
самець, Не усиновлено	Метис	0.3617	0.9545	1.0370	0.9898
Не усиновлено	Метис	0.6380	0.9513	1.0335	0.9832
самка, Не усиновлено	Метис	0.2763	0.9471	1.0290	0.9746
самець	Метис	0.5070	0.9218	1.0016	0.9233
самка	Метис	0.4134	0.9186	0.9981	0.9169
Французький Бульдог	Усиновлено	0.0027	0.9000	2.7331	2.4598

Рис. 4.11 Графічне представлення асоціативних правил

Отримані результати дозволяють адаптувати стратегії пошуку домівок для собак. Для порід із низькою ймовірністю усиновлення можна розробити окремі маркетингові кампанії або підвищити їхню видимість серед потенційних власників. Крім того, результати підкреслюють важливість врахування гендерних особливостей при розробці політик щодо усиновлення.

4.3 Дослідження використання алгоритму кластеризації

Для визначення оптимальної кількості кластерів використовувався метод "Лікоть". Згідно з отриманими результатами, оптимальною кількістю кластерів було обрано $K = 3$, що відповідає балансу між деталізацією та простотою аналізу.

Результати кластеризації були представлені у вигляді інтерактивного графіка, на якому кожна точка відповідає одній собаці, а її колір вказує на приналежність до певного кластера. На графіку осі відображають вік собаки (по горизонталі) та тривалість перебування (по вертикалі). Це можна побачити на рис. 4.12. Аналіз кластерів показав, що у другий кластер увійшли молоді собаки з коротким терміном перебування. Перший кластер об'єднав собак з тривалим

часом перебування. Третій кластер складається зі старших собак, які мають тенденцію до короткого перебування в притулку.

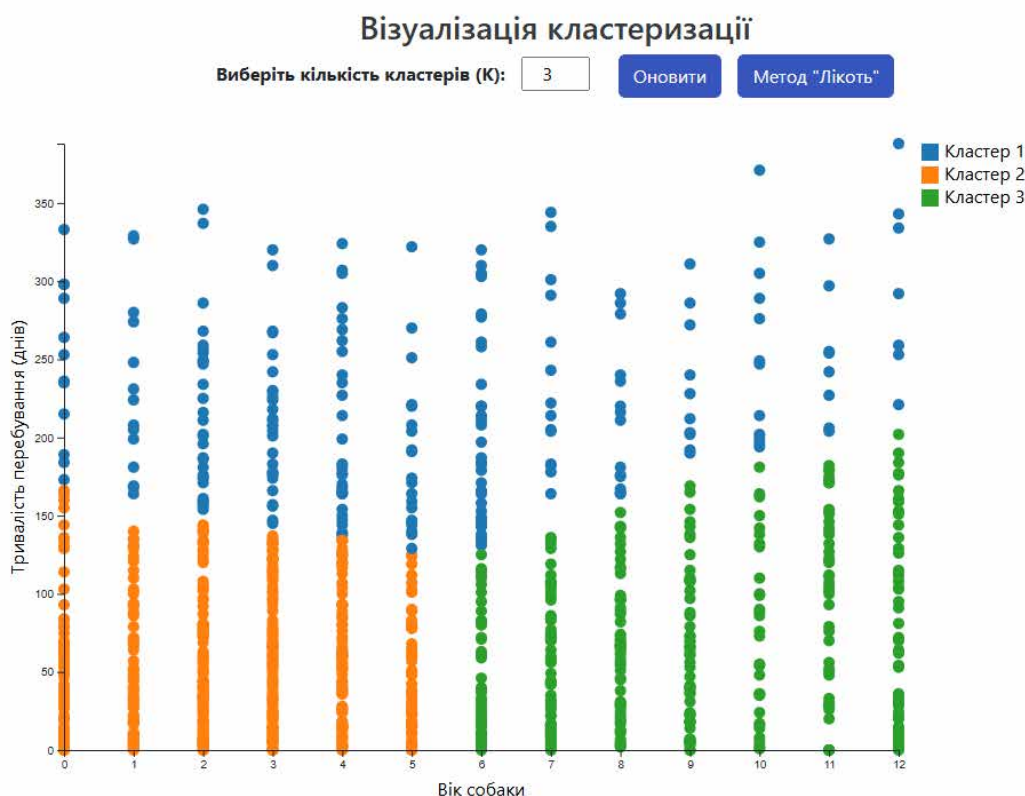


Рис. 4.12 Графічне представлення кластиризації

Отримані кластери дозволили виявити, що молоді собаки мають значно вищі шанси на швидке усиновлення, тоді як старші собаки зазвичай перебувають у притулку довше.

Також були виявлені виняткові випадки, коли молоді собаки мали тривалий термін перебування, а старші собаки – короткий. Це може свідчити про інші фактори, які впливають на усиновлення (наприклад, порода, стан здоров'я або поведінкові характеристики).

Для перевірки гіпотези було проаналізовано ключові показники ефективності за останній рік. Аналіз дозволив оцінити вплив віку собаки на тривалість її перебування у притулку. Результати показують наступне:

Середній час перебування собак віком до 5 років становить 95 діб при цільовому значенні 80 діб. Статус цього *KPI* є "Не досягнуто", а тенденція свідчить про "Погіршення". Це може свідчити про зростання тривалості перебування молодих собак у притулку, що може бути пов'язано із змінами у

попиті на молодих собак або іншими факторами, які вимагають подальшого аналізу. Середній час перебування собак віком від 6 років становить 87 діб, що є меншим за цільове значення 108 діб. Статус КРІ "Досягнуто", а тенденція показує "Покращення". Це вказує на позитивну динаміку скорочення часу перебування старших собак у притулку, що, ймовірно, є наслідком спеціальних заходів, спрямованих на підвищення їх шансів на усиновлення.

Кількість собак віком до 5 років із тривалістю перебування понад 125 днів становить 16 собак при цільовому значенні 49 собак. Цей КРІ має статус "Досягнуто", а тенденція демонструє "Покращення". Це свідчить про значне зниження кількості молодих собак із надмірно тривалим перебуванням у притулку, що є позитивним результатом. Кількість собак віком від 6 років із тривалістю перебування понад 125 днів становить 15 собак при цільовому значенні 45 собак. Статус цього КРІ також "Досягнуто", а тенденція показує "Покращення". Це підтверджує позитивну динаміку для старших собак, які раніше залишалися у притулку на довший час.

Результати КРІ не підтверджують, що вік собаки має значний вплив на тривалість її перебування в притулку. Молоді собаки мають тенденцію до тривалішого перебування, ніж очікувалось, що свідчить про потенційні проблеми в їхньому усиновленні. Натомість старші собаки демонструють позитивну динаміку із зменшенням як середнього часу перебування, так і кількості тривалих перебувань. Це свідчить про ефективність заходів, спрямованих на підвищення шансів на усиновлення старших собак. Однак, для молодих собак необхідно розробити додаткові стратегії для покращення їхніх показників.

4.4 Дослідження використання методу часових рядів для прогнозування

Метод часових рядів застосовується для виявлення сезонних тенденцій та прогнозування змін кількості надходжень тварин до притулку. Основною метою є аналіз історичних даних для визначення закономірностей і побудови прогнозів на майбутні періоди.

Для дослідження використано історичні дані про надходження тварин до притулку за період з 2019 до 2023 року. За допомогою алгоритму ARIMA побудовано прогноз на два наступні роки. На графіку представлено динаміку історичних даних та прогнозні значення. Цей графік зображено на рис. 4.13.



Рис. 4.13 Графік прогнозу кількості надходжень

Аналіз сезонних змін кількості надходжень підтверджує наявність значних коливань між сезонами, що є важливим фактором для ефективного управління притулками. Ключові показники ефективності дозволили оцінити поточну ситуацію та виявити тенденції. У сезоні весна середня кількість надходжень становить 201 собаку, що значно перевищує цільове значення 166, демонструючи тенденцію до погіршення. У зимовий сезон середня кількість надходжень складає 184 собаки, що нижче цілі 185, свідчить про досягнення цільових показників із тенденцією до покращення. У літній сезон середня кількість надходжень дорівнює 197 собак, що перевищує ціль 160, а осінній сезон демонструє 196 собак, що також перевищує ціль 173. Максимальне надходження за сезон складає 201 собаку, що перевищує ціль 185, а мінімальне надходження за сезон становить 184 собаки, що перевищує ціль 160. Середня кількість надходжень за всі сезони становить 194 собаки, що перевищує цільове значення 171, підтверджуючи загальну тенденцію до перевищення цільових показників. Крім того, аналіз по окремих локаціях, таких як Вараш, Львів,

Мукачево та інші, демонструє, що навесні та влітку більшість притулків стикаються з перевищенням середніх значень, тоді як взимку результати є більш стабільними. Ці дані вказують на необхідність адаптації підходів до управління ресурсами, враховуючи сезонність, це дозволить забезпечити своєчасне реагування на пікові навантаження, оптимізувати розподіл ресурсів та покращити ефективність роботи притулків.

Для підтвердження гіпотези "Чи є певні часові періоди, коли в притулок потрапляє більше тварин в певному регіоні?" було проведено аналіз сезонних змін кількості надходжень у різних локаціях за допомогою *KPI* та прогнозування на основі часових рядів. Ключові показники ефективності по кожній локації дозволили виявити значні сезонні коливання. В більшості локацій навесні та влітку кількість надходжень перевищує цільові значення, що свідчить про пікові періоди, зумовлені, ймовірно, сезонними факторами, такими як активний період розмноження тварин чи зміна утримання людьми. Осінь і зима демонструють відносно стабільніші показники, з тенденцією до зменшення кількості надходжень. Прогнозування на наступні два роки підтверджує подібну сезонність, котрий дозволяє притулкам краще планувати свої ресурси та оптимізувати роботу залежно від очікуваних змін. Ключові показники для перевірки зображено на рис. 4.14.



Рис. 4.14 Ключові показники ефективності для перевірки гіпотези

Графіки демонструють сезонні коливання кількості усиновлень у різних регіонах, враховуючи історичні дані та прогноз на два наступні роки. У місті Ватутіне кількість усиновлень залишається стабільною, без значних піків. Історичні дані показують невеликі коливання, але прогноз вказує на стабільну тенденцію у майбутніх сезонах, таким чином це свідчить про залежність рівня усиновлень у цьому регіоні від постійних факторів без значного впливу сезонності. У місті Івано-Франківськ спостерігаються значні сезонні піки, особливо восени, коли кількість усиновлень досягає максимуму. Прогноз на наступні два роки вказує на подальший ріст восени, це підтверджує осінь як піковий сезон для усиновлень. Це може бути пов'язано із завершенням літнього періоду. У місті Кам'янське дані показують помірні коливання з піками взимку, а прогноз вказує на стабілізацію кількості усиновлень, що робить зиму відносно активним періодом.

У Коростені історичні дані показують нерівномірні зміни, а прогноз вказує на стабільний рівень усиновлень без значних піків. У Львові спостерігається стабільний ріст протягом останніх років із чіткими піками восени. Прогноз показує значне зростання восени 2024-2025 років, це підтверджує виражену сезонність із осінню як піковим сезоном. У Мукачеві дані вказують на стабільний ріст із піками взимку та восени, а прогноз підтверджує цю тенденцію, що свідчить про дві основні активні пори року для усиновлень. У Покрові значні коливання з високим рівнем усиновлень восени, а прогноз на наступні два роки вказує на збереження цієї тенденції. Весна демонструє стабільне зростання, яке додає сезонності цьому регіону. У Прилуках спостерігається стабільне зростання восени, тоді як літо та зима залишаються менш активними. Основний період для усиновлень тут – осінь, тоді як інші сезони мають стабільні показники. У Тернополі спостерігається високий рівень сезонності з піками восени. Прогноз підтверджує подальший ріст у цей сезон, тому це робить осінь головним періодом для усиновлень із поступовим зростанням активності весною. У Вараші історичні дані демонструють нерівномірні показники, але прогноз вказує на стабільний рівень

усиновлень, хоча літо демонструє зростання. На зображенні 4.15 продемонстровано результати прогнозу усиновлень.

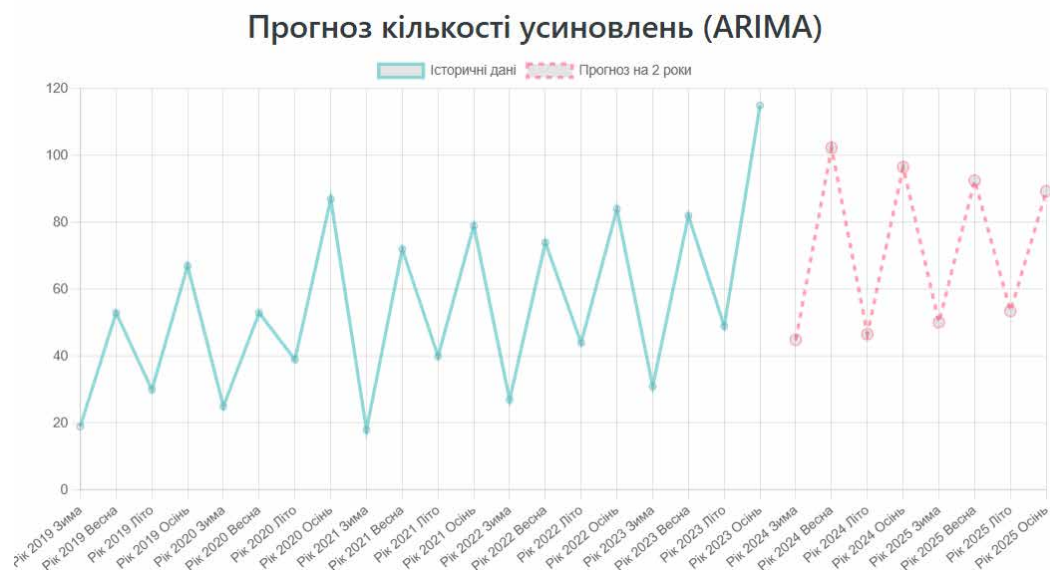


Рис. 4.15 Графік прогнозу усиновлення

Аналіз *KPI* усиновлень показує, що у багатьох регіонах, таких як Мукачево, Львів, Тернопіль, спостерігається покращення та перевищення цільових показників. У регіонах із меншою активністю, як Ватутіне та Коростень, спостерігається стабільність або незначне покращення, проте не у всіх сезонах досягаються цілі. Регіони з вираженою сезонністю, такі як Львів і Тернопіль, мають чітку тенденцію до зростання восени. Інші регіони, зокрема Коростень і Ватутіне, демонструють мінімальні зміни. Водночас, у таких регіонах, як Вараш та Івано-Франківськ, є кілька сезонів із зниженням усиновлень, що може вказувати на необхідність додаткових заходів для стимулювання.

Гіпотеза "Які сезонні коливання спостерігаються у кількості усиновлень в притулку в обраному регіоні?" підтверджується аналізом даних. Осінь є піковим сезоном для більшості регіонів, таких як Львів, Тернопіль і Мукачево. Це може бути пов'язано із завершенням літнього періоду або зміною погодних умов. У регіонах із меншою кількістю усиновлень, як Ватутіне та Коростень, спостерігається стабільна кількість усиновлень без виражених піків. Регіони з низькою активністю в окремих сезонах, у тому числі Вараш та Івано-

Франківськ, потребують додаткових маркетингових кампаній та стимулювання усиновлень. Пікові сезони, а саме осінь у Львові та Тернополі, вимагають збільшення ресурсів і підготовки персоналу. Використання ARIMA для прогнозування підтверджує тенденцію до зростання усиновлень у більшості регіонів, особливо восени. Це дозволяє притулкам краще планувати ресурси та розробляти стратегії для підтримання високої активності. Отже, гіпотеза підтверджується, оскільки сезонні коливання мають значний вплив на кількість усиновлень, особливо в регіонах із вираженою сезонністю, насамперед, Львів, Тернопіль та Мукачево.

4.5 Апаратні та програмні вимоги

Система підтримки прийняття рішень для управління притулками для тварин базується на сучасних технологіях веб-розробки з використанням *Node.js*. Система орієнтована на аналітику, візуалізацію даних та прогнозування, а її архітектура складається з веб-сервера, бази даних та клієнтської частини. Основні компоненти системи включають інтерфейс користувача для перегляду прогнозів і ключових показників ефективності, серверний *API* для обробки запитів та взаємодії з базою даних, а також аналітичні функції, реалізовані за допомогою бібліотек *Node.js*.

Система не є вимогливою до ресурсів, оскільки її архітектура оптимізована для роботи з обмеженим навантаженням (наприклад, кілька користувачів одночасно). Використання бібліотек *Express.js* для побудови серверної частини, *D3.js* для візуалізації даних і баз даних *SQL* забезпечує простоту реалізації та підтримки. Основна обчислювальна потужність потрібна для роботи з базами даних та виконання ETL-процесів, а клієнтська частина потребує лише сучасного веб-браузера.

Вимоги до серверної частини:

- Процесор: *Intel Core i3*.
- Оперативна пам'ять: від 4 ГБ RAM.
- Диск: від 5 ГБ *SSD* для зберігання системних файлів.

- Операційна система: *Linux (Ubuntu 20.04 LTS або вище)*
- Інтернет-з'єднання: Стабільне підключення зі швидкістю 10 Мбіт/с. або більше

Сервер сховища даних:

- Процесор: Intel Core i3 або аналогічний.
- Оперативна пам'ять: від 4 ГБ RAM
- Диск: від 10 ГБ SSD для зберігання даних.
- Операційна система: *Linux (Ubuntu)*.
- База даних: *MS SQL Server 2019*.

Клієнтські пристрої:

- Процесор: Intel Core i3 або аналогічний.
- Операційна система: *Windows, macOS, Linux*.
- Веб-браузер: Остання версія *Google Chrome, Safari, Mozilla Firefox Safari* або *Microsoft Edge*.
- Оперативна пам'ять: від 4 ГБ RAM.

ВИСНОВКИ

Для магістерської кваліфікаційної роботи було розроблено програмне забезпечення системи підтримки прийняття рішень керівництвом собачого притулку, яке відповідає сформульованій меті дослідження. У результаті побудовано оперативна база даних та сховище даних, тим самим забезпечило організацію зберігання та обробку даних. Також розроблено веб-додаток із використанням технологій, таких як *Node.js*, *Express*, *D3.js*, який забезпечує інтерактивний аналіз даних та їх візуалізацію. Зокрема, у системі застосовано методи *Data Mining*, а саме: часові ряди, асоціативні правила, метод Наївного Байєса, *One-Rule*, через які реалізуються гіпотези. Крім того створено та візуалізовано ключові показники ефективності, які дозволяють оцінювати діяльність притулків у динаміці.

Також у ході досліджень встановлено, що сезонні коливання у кількості надходжень і усиновлень собак найбільш виражені у осінньо-зимовий період. Це підтверджено завдяки використанню показників ефективності. Метод асоціативних правил виявив закономірності у впливі характеристик, таких як порода та стать собак, на ймовірність їх усиновлення, де найвищі шанси мають молоді собаки популярних порід, що було перевірено за допомогою показників ефективності. Використання Наївного Байєса показало, що найбільшою ймовірність є серед популярних порід собак, що було підтверджено через *KPI*. Аналіз методом *One-Rule* показав, що вік є основним фактором, який впливає на тривалість перебування собак у притулку, що дозволяє точніше планувати рекламні компанії.

Розроблена система демонструє значні переваги, зокрема швидке прийняття рішення керівництвом через інтерактивну візуалізацію та гнучкість у застосуванні різних методів аналізу для дій притулку. Надалі можливе розширення функціональності системи шляхом інтеграції з іншими джерелами даних, такими як економічні або медичні показники роботи притулку.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Безпритульні тварини у прифронтових регіонах: які проблеми та як держава може допомогти [Електронний ресурс] – Режим доступу до ресурсу: <https://acmc.ua/bezpritulni-tvaryny-u-pryfrontovyh-regionah-yaki-problemy-ta-yak-derzhava-mozhe-dopomogty/>
2. Навіщо потрібен притулок? [Електронний ресурс] – Режим доступу до ресурсу: <https://goodwin.crimea.ua/navishho-potriben-pritulok/>
3. Безкоштовна вакцинація та адопція: як у Вінницькому муніципальному притулку допомагають тваринам [Електронний ресурс] – Режим доступу до ресурсу: <https://vmr.gov.ua/bezkoshtovna-vaktsynatsiia-ta-adoptsii-iak-u-vinnytskomu-munitsypalnomu-pritulku-dopomahaiut-tvarynam>
4. Animal Shelter Facility Use and Design [Електронний ресурс] – Режим доступу до ресурсу: <https://www.aspcapro.org/topics-shelter-medicine/animal-shelter-facility-use-and-design>
5. Standardized Shelter Intake Data Saves More Lives [Електронний ресурс] – Режим доступу до ресурсу: <https://www.humananimalsupportservices.org/blog/standardized-shelter-intake-data-saves-more-lives/>
6. System Development Life Cycle [Електронний ресурс] – Режим доступу до ресурсу: <https://www.geeksforgeeks.org/system-development-life-cycle/>
7. An Introduction to Model-Based Systems Engineering [Електронний ресурс] – Режим доступу до ресурсу: <https://insights.sei.cmu.edu/blog/introduction-model-based-systems-engineering-mbse/>
8. UML: від теорії до практики [Електронний ресурс] – Режим доступу до ресурсу: <https://javarush.com/ua/groups/posts/uk.1972.uml-vd-teor-do-praktiki>

9. Діаграма варіантів використання UML: підручник із ПРИКЛАДОМ [Електронний ресурс] – Режим доступу до ресурсу: <https://www.guru99.com/uk/use-case-diagrams-example.html>

10. Як будувати UML-діаграми. Розбираємо три найпопулярніші варіанти [Електронний ресурс] – Режим доступу до ресурсу: <https://dou.ua/forums/topic/40575/>

11. UML для бізнес-моделювання: для чого потрібні діаграми процесів [Електронний ресурс] – Режим доступу до ресурсу: <https://evergreens.com.ua/ua/articles/uml-diagrams.html>

12. Створення сховищ даних. Технології OLAP та Data Mining [Електронний ресурс] – Режим доступу до ресурсу: https://pidru4niki.com/16120414/informatika/stvorenniya_shovisch_danih_tehnologiyi_olap_data_mining#google_vignette

13. Оптимізація бази даних: шлях до високої продуктивності [Електронний ресурс] – Режим доступу до ресурсу: <https://whileweb.com/uk/blog/optimizaciya-bazi-danih-shlyah-do-visokoyi-produktivnosti/>

14. Що таке сховище даних? Типи, визначення та приклад [Електронний ресурс] – Режим доступу до ресурсу: <https://www.guru99.com/uk/data-warehousing.html>

15. Підручник з інтелектуального аналізу даних: що таке інтелектуальний аналіз даних? Техніка, процес [Електронний ресурс] – Режим доступу до ресурсу: <https://www.guru99.com/uk/data-mining-tutorial.html>

16. Метод класифікації (типології) [Електронний ресурс] – Режим доступу до ресурсу: https://stud.com.ua/49717/ekonomika/metod_klasifikatsiyi_tipologiyi

17. Метод кластеризації: основні принципи та застосування [Електронний ресурс] – Режим доступу до ресурсу: <https://atku.org.ua/metod-klasterizacii-osnovni-principi-ta-zastosuvannya.html>

18. Learn-One-Rule Algorithm [Электронный ресурс] – Режим доступа до ресурсу: <https://www.geeksforgeeks.org/learn-one-rule-algorithm/>
19. Наївний алгоритм класифікатора Байеса [Электронный ресурс] – Режим доступа до ресурсу: <https://www.guru99.com/uk/naive-bayes-classifiers.html>
20. Naive Bayes Models [Электронный ресурс] – Режим доступа до ресурсу: <https://itwiki.dev/data-science/ml-reference/ml-glossary/naive-bayes-models>
21. Apriori Algorithm [Электронный ресурс] – Режим доступа до ресурсу: <https://www.geeksforgeeks.org/apriori-algorithm/>
22. An Introduction to K-Means Clustering [Электронный ресурс] – Режим доступа до ресурсу: <https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/>